

Induktive Kategorienbildung in der Inhaltsanalyse: Kombination automatischer und manueller Verfahren

*Annie Waldherr, Lars-Ole Wehden, Daniela Stoltenberg, Peter Miltner,
Sophia Ostner & Barbara Pfetsch*

Keywords:

Inhaltsanalyse;
Text Mining;
Kategorienbildung;
Bürgerdialog;
Codebuch

Zusammenfassung: Kernstück jeder Inhaltsanalyse ist ein Kategoriensystem, das häufig induktiv-qualitativ an einer kleinen Stichprobe von Texten entwickelt wird. Methoden des *Text Mining* ermöglichen es heute, eine nahezu unbegrenzte Anzahl an Texten effizient, schnell und nachvollziehbar zu explorieren. In diesem Beitrag wird ein Verfahren vorgeschlagen, bei dem solche Methoden eingesetzt werden, um induktiv aus einem umfangreichen Textkorpus Kategorien für eine Inhaltsanalyse zu bilden. Diese Methoden werden mit einer qualitativen, manuellen Inhaltsanalyse kombiniert. Die Kombination verschiedener Verfahren besteht darin, dass zunächst mittels *Text Mining* thematische Oberkategorien aus einem vorliegenden Textkorpus extrahiert, anschließend manuell validiert und in einer qualitativen Inhaltsanalyse um Unterkategorien erweitert wurden. Das Vorgehen wird beispielhaft an einem Codebuch erläutert, welches im Rahmen der Auswertung des "Bürgerdialogs" der Bundesregierung "Gut leben in Deutschland" zum Thema Lebensqualität entwickelt und angewendet wurde.

Inhaltsverzeichnis

- [1. Einleitung und Problemstellung](#)
 - [2. Automatisierung in der Inhaltsanalyse](#)
 - [3. Kombination von *Text Mining* und manueller, qualitativer Inhaltsanalyse zur induktiven Kategorienbildung](#)
 - [3.1 Problemstellung und Datenmaterial](#)
 - [3.2 Bildung von thematischen Oberkategorien mit *Text Mining*](#)
 - [3.3 Ausdifferenzierung von thematischen Unterkategorien mittels qualitativer Inhaltsanalyse](#)
 - [4. Mögliche Anwendungsgebiete des Verfahrens](#)
 - [5. Fazit und Ausblick](#)
- [Anhang 1: Legende zu Tabelle 1](#)
- [Anhang 2: Legende zu Tabelle 2](#)
- [Literatur](#)
- [Zu den Autorinnen und Autoren](#)
- [Zitation](#)

1. Einleitung und Problemstellung

In Zeiten der voranschreitenden Digitalisierung, in denen immer größere Textmengen zu sozialwissenschaftlichen Analyse Zwecken zur Verfügung stehen, hat die automatische Inhaltsanalyse sowohl in der Wissenschaft als auch in der Marktforschung einen regelrechten Boom erlebt (SCHARKOW 2013a). Insbesondere Verfahren des *Text Mining* werden hierfür zunehmend eingesetzt, also Ansätze der algorithmischen Extraktion von Informationen aus großen Datenmengen. Solche Verfahren lassen die Grenzen zwischen qualitativer und quantitativer Textanalyse zunehmend verschwimmen, da sie wie die meisten Ansätze qualitativer Inhaltsanalyse induktiv nach Mustern in unstrukturierten Daten suchen, hierzu aber gleichzeitig große Textmengen heranziehen, was quantitative Aussagen über das Material möglich macht (WIEDEMANN 2013). [1]

In der Forschungspraxis hat sich gezeigt, dass Ansätze des *Text Mining* weder alleine stehen noch vollständig ausgereifte Lösungen anbieten können. Vielmehr gilt es, diese Methoden sinnvoll in verschiedene Arbeitsprozesse zu integrieren und mit anderen, bewährten Verfahren der manuellen Inhaltsanalyse zu kombinieren (SCHARKOW 2013a; WIEDEMANN & LEMKE 2016). Mit überlegener Rechenleistung ausgestattete Computer und zur kognitiven Analyse und Interpretation fähige, menschliche Codierende verfügen jeweils über sich wechselseitig ergänzende Kompetenzen. Die Herausforderung ist, diese Kompetenzen und Fähigkeiten in einer sozialwissenschaftlichen Untersuchung möglichst gewinnbringend zu vereinen, sodass sie möglichst effizient, reliabel und valide durchgeführt werden kann (STULPE & LEMKE 2016; WETTSTEIN 2015). [2]

In diesem Beitrag zeigen wir, wie sich Methoden des *Text Mining* und der qualitativen Inhaltsanalyse sinnvoll kombinieren lassen, um ein inhaltsanalytisches Kategoriensystem induktiv und explorativ zu entwickeln. Der Schritt der Kategorienbildung ist sowohl in qualitativen als auch in quantitativen Verfahren der Inhaltsanalyse zentral. Ist das zu untersuchende Textmaterial zu umfangreich für eine rein qualitative Inhaltsanalyse, so werden häufig Vorstudien durchgeführt, in denen die Kategorien für eine spätere, quantitative Inhaltsanalyse an einer kleinen Stichprobe des Materials qualitativ-induktiv entwickelt werden (FRÜH 2007; MAYRING 2001). Dieser Arbeitsschritt lässt sich mit *Text-Mining*-Methoden gewinnbringend unterstützen. Wir stellen hierzu ein Verfahren vor, das automatische und manuelle Inhaltsanalyse kombiniert und zur Entwicklung eines Codebuchs führt, mit dem in einem Textkorpus vorhandene Themen identifiziert, und quantitativ messbar gemacht werden können. Als *Themen* verstehen wir in Anlehnung an KEPPLINGER (2001, S.119-120) Sinnkomplexe, denen Ereignisse und Stellungnahmen zugeordnet werden können. [3]

Wir konzentrieren uns dabei auf den Prozess der Kategorienbildung und Codebuchentwicklung für eine *Themefrequenzanalyse*. Hier geht es darum, ein hierarchisches Kategoriensystem zu entwickeln, in dem Themen und Unterthemen in *Ober- und Unterkategorien* geordnet, beschrieben und damit

messbar gemacht werden. Die daran potenziell anschließende quantitative Inhaltsanalyse ist nicht Gegenstand dieses Beitrags. Hier sei stattdessen auf die einschlägige Fachliteratur zur quantitativen Inhaltsanalyse (z.B. FRÜH 2007; RÖSSLER 2010) verwiesen. [4]

Das beschriebene Vorgehen fand Anwendung bei der inhaltsanalytischen Auswertung von Freitext-Antworten, die Bürgerinnen und Bürger im Rahmen des von der Deutschen Bundesregierung initiierten "Bürgerdialogs" "Gut leben in Deutschland – Was uns wichtig ist" (im Folgenden kurz: "Bürgerdialog"; BUNDESREGIERUNG DER BUNDESREPUBLIK DEUTSCHLAND 2015) einreichten. Ziel der Analyse war es, die aus Sicht der Bürgerinnen und Bürger wichtigsten Themen im Bereich Lebensqualität in einem Kategoriensystem zu erfassen und auf verschiedenen Ebenen zu ordnen. [5]

Neben der großen Menge an Texten war hierbei insbesondere deren heterogene Struktur eine zentrale Herausforderung. Die Antworten der Bürgerinnen und Bürger variierten in formaler Hinsicht stark in Länge und Aufbau, in inhaltlicher Hinsicht in Bezug auf den Detailgrad und die Anzahl der im Zusammenhang mit Lebensqualität thematisierten Aspekte. Um dieser Heterogenität gerecht zu werden und dennoch ein handhabbares, vollständiges Kategoriensystem zu entwickeln, setzten wir eine Kombination aus Methoden des *Text Mining* und qualitativer Inhaltsanalyse ein. Zunächst wurden mithilfe von *Text-Mining*-Verfahren automatisiert allgemeine, thematische Oberkategorien gebildet, die im Anschluss manuell validiert und mittels qualitativer Inhaltsanalyse in Unterkategorien ausdifferenziert wurden. Durch dieses Verfahren wurden auf der einen Seite die Vorteile des *Text Mining* genutzt, welches erlaubt, mit geringem Aufwand schnell und effizient große Textkorpora so zu explorieren, dass sie relativ intuitiv interpretierbar sind. Gleichzeitig profitierte die Analyse vom Einsatz menschlicher Codierender mit Kontextwissen, die relevante Sinnkomplexe in diesen Texten erfassen und interpretieren können – ein wesentlicher Vorteil der manuellen Inhaltsanalyse. [6]

Im Folgenden wird in Abschnitt 2 zunächst ein kurzer Überblick über gängige Verfahren der automatisierten Inhaltsanalyse gegeben, bevor wir in Abschnitt 3 detailliert die methodische Vorgehensweise unseres Ansatzes zur Entwicklung eines Kategoriensystems vorstellen. In Abschnitt 4 zeigen wir weitere mögliche Anwendungsgebiete dieser Vorgehensweise auf, bevor wir in Abschnitt 5 ein abschließendes Fazit ziehen. [7]

2. Automatisierung in der Inhaltsanalyse

Zur automatisierten Erfassung von textlichen Inhalten steht heute eine Vielzahl verschiedener Ansätze des *Text Mining* zur Verfügung. Der Begriff des *Text Mining* stammt aus der Informatik und bezeichnet hier eine Gruppe von Verfahren, mit denen sich Texte automatisch strukturieren und relevante Informationen aus ihnen extrahieren lassen (HEYER, QUASTHOFF & WITTIG 2006). Dabei werden Texte als unstrukturierte "Datenmenge verstanden, welche sich nach bestimmten Informationen wie dem Auftreten von Schlagwörtern oder der Anzahl der Begriffe durchsuchen lässt" (WETTSTEIN 2014, S.20-21). [8]

In den Sozialwissenschaften werden *Text-Mining*-Verfahren zunehmend zur Inhaltsanalyse von Texten eingesetzt. Sie werden auch für Nicht-Informatiker/innen immer handhabbarer, bieten allerdings keine feststehenden und fertigen Lösungsansätze für die angewandte Forschung (SCHARKOW 2013a). Es gibt keine Verfahren, die "auf Knopfdruck" Texte auswerten und Antworten auf komplexe Fragestellungen liefern. Vielmehr müssen Forschende grundlegende Kenntnisse über Verfahrensgrundlagen, Prozessparameter und Ergebnisvisualisierung mitbringen, um Analysen gezielt zu steuern und Ergebnisse sinnvoll interpretieren zu können (WIEDEMANN & LEMKE 2016). Daher müssen selbst bei stark automatisierten Verfahren stets auch menschliche Entscheidungen hinzugezogen werden. [9]

Bei *überwachten Verfahren* wird die Codierung manuell durch die Forschenden vorbereitet, aber dann komplett maschinell durchgeführt. So wird bei den *diktionärbasierten* Ansätzen vor der maschinellen Codierung ein Kategoriensystem aufgestellt (SCHARKOW 2013a). In diesem Verfahren, das erstmals mittels der Software [General Inquirer](#) umgesetzt wurde (STONE, DUNPHY, SMITH & OGILVIE 1966), werden jeder Kategorie Wörter oder Wortgruppen zugewiesen, die als Indikatoren für das aufzufindende Konstrukt (bspw. das Hauptthema eines Nachrichtenbeitrages) dienen. Das Computerprogramm wendet diese dann auf ein Textkorpus an und versieht einzelne Texte auf dieser Basis mit einem bestimmten Kategorienlabel. Dieser Ansatz weist jedoch einige Schwächen auf (WIEDEMANN 2013, §31 und Abschnitt 3.2.1). Insbesondere bleibt häufig unklar, wie genau die Kategorien und das zugehörige Diktionär gebildet werden, die dann als Grundlage für die Codierung herangezogen werden. [10]

Auch überwachte Verfahren des *maschinellen Lernens* folgen einer deduktiven Logik. Hier wird ein lernender Algorithmus mit Trainingsmaterial, welches zuvor von menschlichen Codierenden manuell codiert wurde, auf das Zuordnen von Texten zu Kategorien geschult (SCHARKOW 2012; SEBASTIANI 2002). Auch um einen solchen Klassifikator trainieren zu können, muss jedoch zuvor ein fertiges Kategoriensystem bereitstehen und man muss über ein Trainingskorpus verfügen, das möglichst alle aufgestellten Kategorien abbilden sowie kohärent, trennscharf und vielfältig sein sollte (WALDHERR, HEYER, JÄHNICHEN, NIEKLER & WIEDEMANN 2016). [11]

Das Training des Klassifikators und seine Anwendung können in integrierten Softwarelösungen auch miteinander verzahnt werden (z.B. WETTSTEIN 2014), sodass automatische und manuelle Inhaltsanalyse parallel ablaufen und sich wechselseitig ergänzen. Dabei nehmen menschliche Codierende über eine Eingabemaske Codierentscheidungen vor. Ein damit verknüpfter Algorithmus lernt auf dieser Basis und nimmt eigene Codierungen vor bzw. macht den Codierenden Vorschläge zur Vergabe von Variablenausprägungen. Auch bei diesem Verfahren müssen jedoch zunächst Kategorien für die Inhaltsanalyse gebildet werden. [12]

Zur explorativen, induktiven Identifikation von Kategorien eignen sich *unüberwachte Verfahren* (WIEDEMANN & LEMKE 2016). Einfache lexikometrische Verfahren wie die *Frequenz-* und die *Differenzanalyse* (nähere Informationen hierzu in Abschnitt 3) beruhen auf Wortstatistiken von automatisch aus Texten extrahierten Schlagwörtern, während mit komplexeren Verfahren wie *Ko-Okkurrenzanalysen* das gemeinsame Auftreten von Schlagwörtern und die daraus entstehenden Wortnetzwerke bzw. Wortcluster untersucht werden. Die Information über Ko-Okkurrenzen wird auch genutzt, um ganze Dokumente automatisch zu thematischen Subgruppen zu clustern (*document clustering*, GRIMMER & KING 2011, S.2644). [13]

Auch Verfahren des *Topic Modeling* basieren auf Wort-Ko-Okkurrenzen, sie modellieren Themen aber als latente Faktoren in einem Bayesschen Wahrscheinlichkeitsmodell (z.B. "Latent Dirichlet Allocation", BLEI, NG & JORDAN 2003). Dies hat zum einen den Vorteil, dass einem Dokument mehrere Themen zugeordnet werden können (SCHARKOW 2013a). Zum anderen können latente semantische Beziehungen zwischen Wörtern aufgedeckt werden, d.h. Wörter, die einem Thema mit hoher Wahrscheinlichkeit zugeordnet werden, müssen nicht unbedingt tatsächlich gemeinsam in einem Dokument auftreten (WALDHERR, HEYER et al. 2016). [14]

Während die deduktiven, überwachten Verfahren einen erheblichen Aufwand zur Vorbereitung der eigentlichen Codierung erfordern, fällt dieser bei den induktiven, unüberwachten Verfahren des *Text Mining* weg. Dafür erfordern letztere Verfahren im Nachhinein sehr viel Aufwand für die Interpretation und Validierung der vollautomatisch extrahierten Schlagwörter und Themenkategorien (GÜNTHER & QUANDT 2016). So fordern etwa STULPE und LEMKE (2016) einen Ansatz des *blended reading* (vgl. Abschnitt 3), bei dem die Vorteile des *distant reading* eng mit den Vorteilen des *close reading* relevanter Einzeltexte verknüpft werden. *Distant reading* ist dabei die automatisierte und computerbasierte Analyse der betreffenden Texte durch unüberwachte Verfahren des *Text Mining*, während *close reading* "konventionelle [...], durch einen lesenden Interpreten durchgeführte[...] Analysen" bezeichnet (S.39-40 in Anlehnung an MORETTI 2013). [15]

Manuelle und automatische Methoden der Inhaltsanalyse sind also nicht als miteinander im Wettbewerb stehend, sondern als einander ergänzend zu begreifen (SCHEU, VOGELGESANG & SCHARKOW 2018; WIEDEMANN 2013).

Die Vorteile automatischer Inhaltsanalysen (insbesondere Effizienz und Gründlichkeit) und manueller Inhaltsanalysen (insbesondere die Interpretationsfähigkeit und das Weltwissen menschlicher Codierender sowie eine hohe Validität; WETTSTEIN 2014) lassen sich in vielfältiger Hinsicht kombinieren. [16]

Meist werden einzelne Arbeitsschritte der Analyse durch Automatisierung effizienter gestaltet (STUCKARDT 2000; WETTSTEIN 2014). So ist es beispielsweise gängige Praxis, relevante Texte für Inhaltsanalysen anhand bestimmter Schlagwörter in Datenbanken zu identifizieren (SCHARKOW 2012). Um relevante Texte in Bezug auf diffuse Themen zu identifizieren, für die keine griffigen Schlagwörter zur Verfügung stehen, eignen sich lernende Klassifikationsalgorithmen (WALDHERR, MAIER, MILTNER & GÜNTHER 2016). Auch der Prozess der Codierung selbst kann durch Automatisierung effizienter und reliabler gestaltet werden. Dies gelingt bereits recht zuverlässig bei der Codierung von Themen (z.B. SCHARKOW 2013b) oder Stimmungen (*sentiments*; z.B. CERON, CURINI, IACUS & PORRO 2014). [17]

Noch selten werden automatisierte Verfahren für die Entwicklung von Kategoriensystemen für die Inhaltsanalyse – also die Codebuchentwicklung – eingesetzt. Mit dem "Term-Mapping" schlägt WETTSTEIN (2012) ein teilautomatisiertes, ko-okkurrenzbasiertes Verfahren vor, um sich einen Überblick über wesentliche Dimensionen einer öffentlichen Debatte zu einem Thema zu verschaffen. Zunächst wird auf der Basis ein bis drei manuell identifizierter Kernbegriffe, die für eine thematische Debatte als charakteristisch gelten, eine Stichprobe von Artikeln gezogen. Diese Kernbegriffe dienen dabei als Ausgangspunkt für die Stichprobenziehung. In diesen werden dann automatisiert Schlagwörter identifiziert und mittels multidimensionaler Skalierung ("Smallest Space Analysis", WETTSTEIN 2012) geclustert. Das gemeinsame Auftreten von Wörtern sowie ihre Zentralität für eine Debatte (gemessen über das gemeinsame Auftreten mit den zuvor manuell identifizierten Kernbegriffen) bestimmen dabei ihre Anordnung in der Abbildung der *Smallest Space Analysis*. Auf Grundlage der entstehenden Begriffscluster können schließlich die für die Debatte wichtigen Unterthemen (Kategorien) qualitativ-interpretativ erschlossen werden. [18]

WETTSTEIN wendet das Verfahren erfolgreich auf ein journalistisches Textkorpus an und findet je eine einstellige Zahl an Unterthemen in einer thematisch relativ eng eingegrenzten Auswahl journalistischer Texte (z.B. Unterthemen zum Thema Arbeitslosigkeit). Er warnt jedoch selbst davor, sich ganz auf die automatisch generierten Themenzuordnungen zu verlassen:

"Das Term-Mapping ersetzt dabei jedoch keine manuelle Inhaltsanalyse, da die Verwendung der einzelnen Begriffe und der Kontext, in dem sie genannt wurden, vom Verfahren nicht berücksichtigt werden. [...] [E]s hängt vom Wissen des Forschers ab, die einzelnen Begriffe in einen sinnvollen Kontext zu setzen. Die Interpretation ist dadurch unter Umständen nicht intersubjektiv nachvollziehbar" (S.153-155) [19]

Ein weiteres teilautomatisiertes Verfahren, das sich zur thematischen Exploration eines Textkorpus eignet, implementiert LEJEUNE (2011) in der Software [Cassandra](#). Diese Software ist in erster Linie dazu gedacht, die qualitative Kategorienentwicklung aus dem Textmaterial zu unterstützen. Auch hier müssen jedoch zunächst Kernbegriffe als *Marker* definiert werden. Die Automatisierung beschränkt sich darauf, alle Textpassagen nebeneinander anzuzeigen, die den jeweiligen Kernbegriff enthalten. So können die Forschenden einfach durch das Korpus browsen, um relevante Textpassagen zu finden und den Kontext zu beurteilen. Auch die schrittweise Bildung und Zuordnung abstrakter Kategorien wird durch die Software unterstützt. [20]

In diesem Beitrag schlagen wir ebenfalls ein Verfahren vor, das genutzt werden kann, um ein Textkorpus thematisch zu dimensionieren und daraus induktiv Themenkategorien abzuleiten. Anders als WETTSTEINs (2012) und LEJEUNEs (2011) Verfahren ist unser Ansatz jedoch vollständig explorativ. Es müssen nicht vorab zentrale Kernbegriffe definiert werden, da die Identifikation der zentralen Oberkategorien und sie bezeichnender Schlagwörter automatisch durch unüberwachte *Text-Mining*-Verfahren erfolgt. [21]

Zudem eignet sich unser Verfahren für besonders allgemeine und verschachtelte Themen mit Unterthemen auf mehreren Hierarchieebenen sowie nutzer/innengenerierte Inhalte, die thematisch sehr breit gefasst sind und sehr unterschiedliche Themen in einem Textdokument vereinen. Für das Textkorpus des "Bürgerdialogs", dessen Antworten zum Thema Lebensqualität genau diese Charakteristika zeigten, erwiesen sich ko-okkurrenzbasierte Verfahren des Themenclustering als nicht praktikabel für die Kategorienbildung (WALDHERR et al. 2015). [22]

3. Kombination von *Text Mining* und manueller, qualitativer Inhaltsanalyse zur induktiven Kategorienbildung

Das in diesem Beitrag vorgeschlagene Verfahren greift die Forderungen nach einer engeren Verzahnung von automatischen und manuellen Verfahren der Inhaltsanalyse auf und begegnet den oben formulierten Herausforderungen. Gemäß der Idee des *blended reading* (STULPE & LEMKE 2016) verknüpfen wir unüberwachte *Text-Mining*-Verfahren (*distant reading*), mittels derer wir Themenkategorien identifizieren, mit dem Lesen von Texten durch menschliche Codierende (*close reading*), durch das die Themenkategorien validiert werden. [23]

Zur weiteren Ausdifferenzierung breiter Oberthemen in Unterkategorien systematisieren wir die Strategien des *close reading* zudem noch weiter und folgen dabei bewährten Grundregeln der qualitativen Inhaltsanalyse (MAYRING 2010). Da in allen Arbeitsschritten mehrere menschliche Codierende beteiligt sind, die nach definierten Regeln arbeiten und sich nach dem Konsensprinzip abstimmen, begegnen wir außerdem den Forderungen nach einer höheren Intersubjektivität der qualitativen Codebuchentwicklung. [24]

Mit dem auf diese Weise entwickelten Kategoriensystem lassen sich dann verschiedene Forschungsvorhaben weiterverfolgen. Dazu zählen die quantitative manuelle Inhaltsanalyse (wie in dem diesem Artikel zugrunde liegenden Forschungsprojekt), die Definition von Wörterbüchern zur Nutzung der diktionsbasierten Codierung oder das Training eines Klassifikators (vgl. Abschnitt 4). [25]

Eine überblicksartige Darstellung des Vorgehens kann dem Ablaufdiagramm in Abbildung 1 entnommen werden. Im Folgenden werden die einzelnen, aufeinander aufbauenden Schritte der Kategorienentwicklung detailliert erläutert.

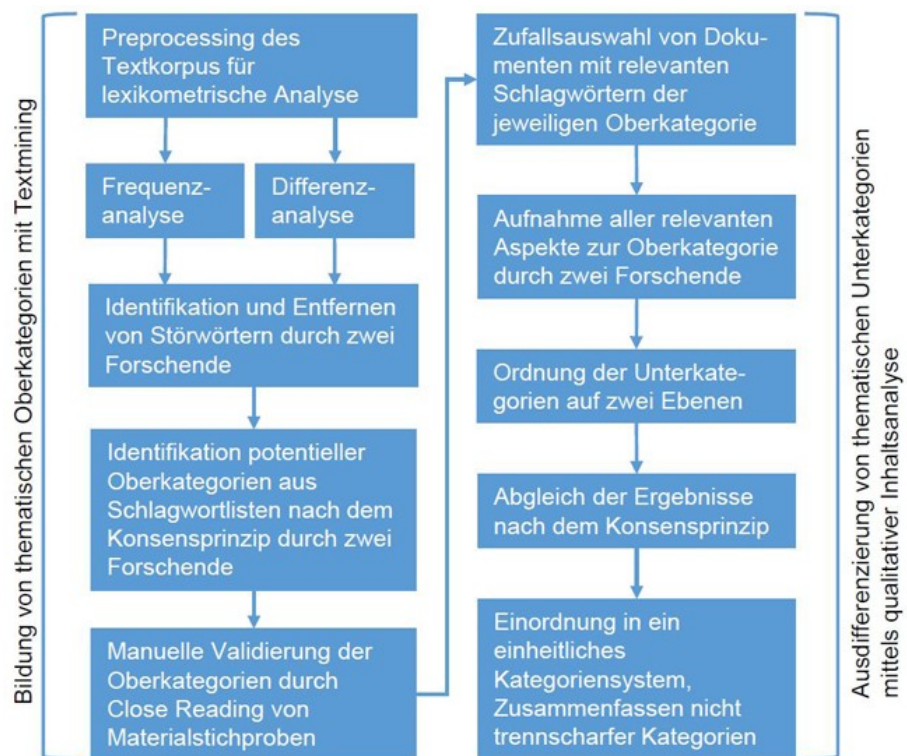


Abbildung 1: Ablaufdiagramm zum methodischen Vorgehen bei der Kategorienbildung [26]

3.1 Problemstellung und Datenmaterial

Die hier beschriebene Methodik wurde innerhalb eines Forschungsprojekts entwickelt, welches das Ziel hatte, Textbeiträge von Bürgerinnen und Bürgern auszuwerten, die innerhalb des "Bürgerdialogs" der Deutschen Bundesregierung zum Thema "Gut leben in Deutschland" eingingen. Es wurde ein Codebuch entwickelt, um später die Antworten auf zwei Fragen zur Lebensqualität, die im Rahmen des Bürgerdialogs gestellt wurden, quantitativ zu codieren (Frage 1: "Was ist Ihnen persönlich wichtig im Leben?", Frage 2: "Was macht Ihrer Meinung nach gutes Leben in Deutschland aus?"). Das Kategoriensystem sollte Themen und Unterthemen auf mehreren Ebenen hierarchisch abbilden. [27]

Die Antworten der Bürgerinnen und Bürger, die in Form von 2.522 Online-Beiträgen eingingen, waren als Freitext mit einer maximalen Länge von 1.000 Zeichen gestaltet. Dabei unterschieden sich die Antworten sowohl inhaltlich als auch formal sehr stark voneinander. Inhaltlich wurde eine ganze Bandbreite von Themen mit Bezug zu persönlicher Lebensqualität angesprochen. Dabei enthielten einzelne Beiträge oftmals viele verschiedene Themen, und teilweise wurden diese verschiedenen Themen auch innerhalb eines Satzes aufgezählt. Somit wiesen die Texte andere Charakteristika auf als z.B. Nachrichtenbeiträge, die sich häufig in einem thematisch eng eingegrenzten Feld bewegen. [28]

Im Gegensatz zu nachrichtlichen Texten war eine inhaltliche Zuordnung von jeweils einem Thema oder Themenkomplex zu einem Textabschnitt, also z.B. zu einem Satz oder Absatz, über alle Texte hinweg nicht möglich, da die Bürgerinnen und Bürger ihre Antworten auch formal sehr unterschiedlich aufbauten. Die Bandbreite reichte von Beiträgen, die nur ein einzelnes Stichwort enthielten, über durch Kommata getrennte Aufzählungen von Stichwörtern, Stichwortlisten mit jeweils einem neuen Absatz pro Thema bis hin zu aufwendig ausformulierten Fließtexten. Diese Fließtexte enthielten in einigen Fällen wiederum nur ein einziges inhaltliches Thema zu Lebensqualität, das dann in verschiedenen Facetten detailliert erläutert wurde. In anderen Fließtexten wurden hingegen viele verschiedene Themen ohne weitere Untergliederung angesprochen. Teilweise besprachen Antwortende ein einzelnes für sie wichtiges Thema sehr breit und schlossen daran zusätzliche Themen in Form reiner Aufzählungen an. [29]

Diese Bandbreite der Texteingaben erschwerte eine automatische Codierung der im "Bürgerdialog" aufgebrachten Themen (etwa durch Themenclustering) erheblich, denn "während menschliche Codierende sich zumeist inhaltlicher, d.h. syntaktischer oder semantischer Kriterien bedienen, kann das Unitizing [Zerlegung des Untersuchungsmaterials in Analyseeinheiten] bei automatischen Verfahren häufig nur auf formale Kriterien, d.h. Satzzeichen oder Absatzmarken, zurückgreifen" (SCHARKOW 2013a, S.294; vgl. auch WIEDEMANN 2013, §20). [30]

Ziel des Forschungsprojekts war es jedoch, zunächst alle angesprochenen allgemeinen Themen zu erfassen und diese dann in einem Kategoriensystem auf mehreren Ebenen sinnvoll zu ordnen und auszudifferenzieren. Automatischen Verfahren gelingt es bereits gut, die Hauptthemen monothematisch angelegter Nachrichtenbeiträge zu identifizieren (SCHARKOW 2012, 2013b). Die Erfassung und Ordnung komplexerer semantischer Einheiten und ihrer Zusammenhänge wie eine Kategorisierung in Themen und Unterthemen lässt sich hingegen noch nicht zuverlässig und vollständig automatisieren (SCHARKOW 2013a; WIEDEMANN & LEMKE 2016). Allerdings ist die Materialfülle häufig, so auch in diesem Projekt, so umfangreich, dass eine vollständige händische Exploration und Kategorienbildung nicht oder nur unter erheblichem Personal- und Zeitaufwand zu bewältigen wäre (STULPE & LEMKE 2016). In diesen Fällen ist es hilfreich, einen Mittelweg anzustreben und die Vorteile der automatischen Inhaltsanalyse mit denen der manuellen zu verbinden, wie wir es nachfolgend skizzieren. [31]

3.2 Bildung von thematischen Oberkategorien mit *Text Mining*

Unser Verfahren ist vollständig explorativ und kombiniert daher induktive Verfahren sowohl der manuellen als auch der automatischen Inhaltsanalyse. Ausschlaggebend hierfür war die Überlegung, dass die Kategorien explorativ aus den Antworten der Bürgerinnen und Bürger entwickelt werden sollten, um ggf. neue Kategorien von Lebensqualität zu entdecken, die in bestehenden wissenschaftlichen Systematiken von Lebensqualität nicht bereits enthalten sind. Dies schloss eine deduktive Vorgehensweise aus, die z.B. die *a priori* Definition von Kernbegriffen, wie bei WETTSTEIN (2012), verlangt. [32]

Wie oben erläutert, eignen sich unüberwachte Verfahren des *Text Mining* besonders dazu, große Datenmengen systematisch induktiv zu erschließen und schnell zu explorieren. Diesem Vorteil steht der Nachteil gegenüber, dass sich die Texte nicht in so großer Tiefe analysieren lassen wie mit einer manuellen, qualitativen Textanalyse (WALDHERR, HEYER et al. 2016). Dafür läuft die Analyse zunächst völlig datengetrieben und autonom ab und ist damit äußerst effizient. Die Ergebnisse dieser ersten, automatischen Analyse müssen von den Forschenden dann in einem zweiten Schritt weiter interpretiert werden (DZUDZEK, GLASZE, MATTISSEK & SCHIRMEL 2009; WIEDEMANN 2013). [33]

Für die systematische Entwicklung des Kategoriensystems erwiesen sich die lexikometrischen Verfahren der Frequenzanalyse und der Differenzanalyse (DZUDZEK et al. 2009) als besonders geeignet, um die qualitative, manuelle Identifikation von Themenkategorien vorzubereiten und zu unterstützen. Ziel beider Methoden ist die Extraktion von besonders häufig genannten bzw. charakteristischen Substantiven (Schlagwörtern) aus einem Korpus. [34]

Für die Auswertung des "Bürgerdialogs" wurde hierzu die kostenpflichtige, browserbasierte Software [Topic Analyst](#) (CID GMBH 2017) verwendet, innerhalb derer die genannten Text-Mining-Verfahren implementiert sind. Eine Freeware-Alternative, mit der sich ebenfalls browserbasiert Frequenz- und Differenzanalysen durchführen lassen, bieten SINCLAIR und ROCKWELL (2016) mit [Voyant Tools](#). Für Forschende mit Programmierkenntnissen gibt es für solche Analysen zudem eine stetig wachsende Zahl an Softwarepaketen innerhalb der Programmiersprachen *R* und *Python*. Eine aktuelle Einführung in Text Mining mit *R* geben z.B. SILGE und ROBINSON (2017). [35]

3.2.1 *Preprocessing*

Die Texteingaben des "Bürgerdialogs" wurden im Rahmen der automatischen Sprachverarbeitung in *Topic Analyst* zunächst für die eigentliche Analyse vorbereitet (*Preprocessing*). Hierzu wurden die Texte interpretiert, d.h., verschiedene Satzbestandteile bzw. Wortgruppen (bspw. Verben, Substantive) wurden identifiziert. Substantive wurden lemmatisiert (auf die Grundform gebracht) und als Schlagwörter (Keywords) an den Text annotiert. Stoppwörter (Wörter, die keinen Nutzen für die Analyse bringen, wie "z.T." und "z.B.") wurden entfernt. Bei den hier betrachteten Schlagwörtern handelte es sich vornehmlich

um Substantive, die auf ihre Grundform reduziert wurden (also z.B. *Freunde* zu *Freund*). Darüber hinaus erkannte die Software einige Mehrwortterme (z.B. feststehende Floskeln wie "Mit freundlichen Grüßen"). Zu näheren Informationen über gängige *Preprocessing*-Verfahren sei auch auf SCHARKOW (2013a) sowie HOTHO, NÜRNBERGER und PAASS (2005) verwiesen. [36]

Die weiteren Analysen fokussierten sich auf die extrahierten und annotierten Substantive als Schlagwörter. Gegenüber anderen Wortarten haben Substantive den Vorteil, dass sie recht spezifisch sind und meist die wesentlichen Informationen innerhalb einer Aussage transportieren. Sie weisen häufig bereits auf Themen hin, während Verben, Adjektive und Adverbien eher unspezifisch sind und Handlungen und Stimmungen ausdrücken (HEYER et al. 2006). Um Themenkategorien abzuleiten, erwies sich daher die Schlagwortanalyse von Substantiven als besonders hilfreich, da so schnell ein breiter Überblick über häufig genannte und inhaltlich relevante Begriffe gewonnen werden konnte. [37]

3.2.2 Frequenzanalyse

In einer Frequenzanalyse wird automatisiert ausgezählt, wie häufig bestimmte Schlagwörter in einem Textkorpus vorkommen. Dabei gibt *Topic Analyst* die *Dokumentfrequenz* aus, d.h. die im Folgenden berichteten Werte beziehen sich stets auf die Anzahl der Antworten¹, in denen das jeweilige Schlagwort mindestens einmal genannt wurde. [38]

Frequenzanalysen bieten die Möglichkeit, mit relativ geringem Aufwand einen ersten Eindruck vom vorliegenden Textkorpus zu erhalten, und eignen sich daher gut als Startpunkt für eine Inhaltsanalyse (TRIBBLE & JONES 1997). Dieses *distant reading* erleichtert die Datenstrukturierung (STULPE & LEMKE 2016, S.40) und liefert erste Anhaltspunkte für nachfolgende Analyseschritte, die zudem relativ intuitiv zugänglich und interpretierbar sind (WIEDEMANN & LEMKE 2016). Eine Häufung von Wortfrequenzen während eines bestimmten Zeitabschnitts weist etwa darauf hin, dass die betreffenden Begriffe in dieser Zeit besonders stark diskutiert wurden bzw. besonders relevant oder kontrovers waren. Hier ergeben sich dann Einstiegspunkte für das *close reading* (STULPE & LEMKE 2016, S.40). [39]

Somit stellt die Frequenzanalyse ein gutes Instrument dar, um zunächst aus den wichtigsten Schlagwörtern in einem Textkorpus grobe Kategorien zu formen, die später weiter ausdifferenziert werden können. Allerdings ist zu bedenken, dass Schlagwortlisten einzelner Substantive allein nicht in jedem Fall als guter Indikator für einen ähnlichen Inhalt in Dokumenten mit ähnlichen Schlagwörtern dienen können (WETTSTEIN 2014; WIEDEMANN 2013). Deshalb ist eine zusätzliche, qualitative Validierung unerlässlich. [40]

1 Die Dokumentfrequenzen beziehen sich stets auf eine der beiden Fragen zur Lebensqualität (Frage 1: "Was ist Ihnen persönlich wichtig im Leben?", Frage 2: "Was macht Ihrer Meinung nach gutes Leben in Deutschland aus?").

Bei der Betrachtung reiner Wortfrequenzen in Texten ist zudem zu beachten, dass Wörter, die ohnehin häufig in einer Sprache vorkommen, mit hoher Wahrscheinlichkeit auch häufig im interessierenden Textkorpus verwendet werden, z.B. *Frau*, *Mann*, *Mensch* usw. in der deutschen Sprache. Dies sind aber häufig gerade nicht die Wörter, die das Korpus inhaltlich aussagekräftig beschreiben, da sie sehr allgemein sind und in verschiedenen Kontexten verwendet werden (BARON, RAYSON & ARCHER 2009). Viele Wörter, die ein Korpus inhaltlich aussagekräftig beschreiben, treten zwar nicht so häufig wie diese sehr allgemeinen Begriffe auf, sind aber im Korpus prominenter vertreten als in der Allgemeinsprache (DUNNING 1993). Daher wurde zusätzlich zur Frequenzanalyse das Instrument der Differenzanalyse eingesetzt. [41]

3.2.3 Differenzanalyse

Bei dieser Methode wird das Vorkommen von Wörtern im interessierenden Analysekorpus mit dem Vorkommen von Wörtern in einem Referenzkorpus verglichen (RAYSON & GARSIDE 2000). Dieses Referenzkorpus sollte üblicherweise deutlich umfangreicher und allgemeiner sein als das Analysekorpus und damit einen Ausschnitt aus dem allgemeinen Sprachgebrauch liefern, d.h., es sollte eine Vielzahl von Sätzen aus unterschiedlichen Themenbereichen umfassen, ohne inhaltlich in eine bestimmte thematische Richtung verzerrt zu sein (WALDHERR, HEYER et al. 2016). Die in Analyse- und Referenzkorpus enthaltenen Texte müssen jeweils in derselben Sprache verfasst sein. Die Differenzanalyse ist in der automatisierten Sprachverarbeitung und in der Computerlinguistik ein häufig eingesetztes Verfahren (BARON et al. 2009; HEYER et al. 2006). [42]

Als Referenzkorpus für den "Bürgerdialog" wurde in Anlehnung an BIEMANN, HEYER, QUASTHOFF und RICHTER (2007) ein allgemeinsprachliches Referenzkorpus aus nachrichtlichen Texten gebildet und in die Software *Topic Analyst* eingepflegt. Es umfasste 121.317 Nachrichtendokumente in deutscher Sprache, die durch Web-Crawler² aus den Online-Ausgaben der Zeitungen und Magazine *Frankfurter Allgemeine Zeitung*, *Frankfurter Neue Presse*, *Focus*, *taz* und *Spiegel* gesammelt wurden, und die im Zeitraum vom 27.12.2014 bis 30.07.2015 erschienen sind.³ Um eine größtmögliche thematische Vielfalt zu garantieren, wurden stets sämtliche Ressorts aus einer Quelle übernommen. [43]

Ein allgemeinsprachliches Korpus sollte, um repräsentativ für eine Sprache zu sein, möglichst Exemplare verschiedener Textgattungen enthalten (LEECH 1993). Die Häufigkeit, mit der verschiedene Textgattungen in das Korpus eingehen, sollte dabei möglichst proportional zu ihrer Verwendung in der Alltagssprache sein (CLEAR 1992). Allgemeinsprachliche Korpora enthalten

2 Ein Crawler ist eine Softwarekomponente, welche textliche Inhalte auf Webseiten extrahiert. Werbung, Metadaten oder nicht zum Text gehörende Inhalte werden automatisch herausgefiltert.

3 Das Korpus wurde mit insgesamt 48 Crawlern erhoben und setzt sich wie folgt zusammen: *Frankfurter Allgemeine Zeitung* (11 Crawler, 22.309 Dokumente), *Focus* (12 Crawler, 26.662 Dokumente), *Frankfurter Neue Presse* (3 Crawler, 42.181 Dokumente), *Spiegel* (17 Crawler, 21.342 Dokumente), *taz* (5 Crawler, 8.823 Dokumente).

neben Zeitungstexten daher in der Regel auch größere Anteile belletristischer Texte und Fachtexte. Beispiele dafür sind das Digitale Wörterbuch der Deutschen Sprache (GEYKEN 2007) oder das Deutsche Referenzkorpus (INSTITUT FÜR DEUTSCHE SPRACHE 2016). [44]

Alternativ werden in der automatisierten Sprachverarbeitung auch Korpora aus Zeitungstexten als geeignete Referenzkorpora angesehen und häufig eingesetzt, zum Beispiel der Leipziger Wortschatz (BIEMANN et al. 2007). Da dieser allerdings für die systembedingten Prozesse des *Topic Analyst* zu umfangreich war, wurde nach dessen Vorbild wie oben beschrieben ein eigenes Korpus aus damals aktuellen Zeitungstexten erstellt. Als Referenz für die Antworten des "Bürgerdialogs" erschienen diese als besonders geeignet, da sie nah an die gesprochene bzw. geschriebene Sprache der Bürgerinnen und Bürger heranreichen (im Gegensatz etwa zu Belletristik), mit vertretbarem Aufwand zusammengestellt werden konnten und zugleich über die Bandbreite an Ressorts eine große thematische Vielfalt abdecken. [45]

Für den Vergleich des Analysekorpus mit dem Referenzkorpus eignet sich DUNNING's (1993) *log-likelihood ratio test*. Teilweise werden auch Chi-Quadrat-Tests oder andere Signifikanztests durchgeführt, ihre Angemessenheit ist auf dem Gebiet der automatisierten Sprachverarbeitung allerdings umstritten (vgl. z.B. BESTGEN 2014; RAYSON & GARSIDE 2000). Insbesondere basieren die Standardtests wie Chi-Quadrat auf Normalverteilungsannahmen, die die Verteilung von Worthäufigkeiten erheblich verletzt. Sprache enthält deutlich mehr seltene Ereignisse, als bei einer Normalverteilung angenommen wird (DUNNING 1993). So setzt sich das gesamte Vokabular einer Sprache sowohl aus einigen sehr häufig wie auch vielen sehr selten gebrauchten Wörtern zusammen. Dabei transportieren gerade die seltenen Wörter den meisten Inhalt. [46]

Als Alternative schlägt DUNNING den *log-likelihood ratio test* vor, für den keine spezifischen Verteilungsannahmen getroffen werden müssen. Hierfür werden zunächst Frequenzlisten für die Wörter im Referenzkorpus sowie im Analysekorpus erstellt. Für jedes Wort im Referenzkorpus wird ein statistischer Erwartungswert für seine Häufigkeit berechnet. Die tatsächlich beobachteten Häufigkeiten der Wörter im Analysekorpus werden nun mit diesen Erwartungswerten verglichen. Dazu wird ein logarithmierter Wert berechnet, der angibt, wie stark ein Wort in seiner tatsächlichen Häufigkeit im Analysekorpus von seiner erwarteten Häufigkeit abweicht, also wie überraschend es ist, das betreffende Wort im Analysekorpus in der vorliegenden Häufigkeit vorzufinden. Basierend auf diesem Wert lässt sich nun eine Liste der Wörter des Analysekorpus ausgeben, die verglichen mit dem Referenzkorpus besonders häufig vorkommen und daher besonders charakteristisch für das Analysekorpus sind. Dieses Verfahren liefert in der Regel sehr gute Ergebnisse (BORDAG 2008; RAYSON & GARSIDE 2000; WALDHERR, HEYER et al. 2016). [47]

Dieser Logik folgt die Software *Topic Analyst* im Wesentlichen, der genaue Algorithmus ist allerdings nicht dokumentiert.⁴ Ausgegeben werden Listen von bis zu 70 Schlagwörtern mit den höchsten Werten der Differenzanalyse. Diese Werte werden in der Software als "Signifikanzwert" bezeichnet. Es handelt sich allerdings nicht um einen p-Wert, der ein bestimmtes Signifikanzniveau bezeichnet, sondern um einen Wert, der angibt, inwieweit die tatsächlich beobachtete Frequenz eines Schlagworts im Analysekorpus vom statistischen Erwartungswert auf Basis der Frequenz im Referenzkorpus abweicht. Deshalb verwenden wir im Folgenden den Begriff *Differenzscore*. [48]

Die Differenzanalyse ermöglicht es zusammenfassend, Schlagwörter im Analysekorpus zu identifizieren, die durch die reine Betrachtung der Frequenz verloren gingen, aber im Vergleich zum Referenzkorpus überdurchschnittlich häufig genannt werden. Die so zusätzlich identifizierten Schlagwörter sind daher besonders spezifisch und charakteristisch für das analysierte Textkorpus. Für weiterführende Informationen zur genauen mathematischen Funktionsweise von Differenzanalysen sei auf RAYSON und GARSIDE (2000) verwiesen. [49]

3.2.4 Ergebnisse der Frequenz- und Differenzanalysen

Ziel der Frequenz- und Differenzanalysen war es, möglichst viele aussagekräftige Schlagwörter im Analysekorpus zu identifizieren, die Anhaltspunkte für mögliche Oberthemen im Korpus geben konnten und später als Einstiegspunkte in die vertiefende, qualitative Inhaltsanalyse dienen sollten. Daher wurden die Instrumente der Frequenz- und Differenzanalyse gleichwertig in die Untersuchung einbezogen. Beide Methoden wurden jeweils separat auf die Antwortdokumente angewandt, die zu einer der beiden Fragen des "Bürgerdialogs" vorlagen. Die Frequenzanalyse lieferte für jede Frage eine Liste der 70 am häufigsten genannten Schlagwörter im Korpus, die absteigend nach ihrer absoluten Dokumentfrequenz sortiert wurden. Die Differenzanalyse ergab zusätzlich zwei Listen der 70 je Frage charakteristischsten Schlagwörter, absteigend geordnet nach ihrem *Differenzscore*. [50]

Diese Listen enthielten zunächst noch einige Störwörter, die nicht sinnvoll zur Analyse beitrugen, etwa Schlagwörter, die nur aus Großbuchstaben bestanden, Schreibfehler enthielten oder Eigennamen darstellten. Da diese Wörter zwar hohe Scores in der Differenzanalyse erzielten, im Sinne des Erkenntnisinteresses aber nicht valide sind (WIEDEMANN 2013), wurden sie manuell entfernt. Ebenso wurden Störwörter aussortiert, die so allgemein waren, dass sie nicht auf ein bestimmtes Thema hinwiesen (z.B. *Gesellschaft, Mensch, Ding*). Sie eigneten sich demnach nicht, spezifische Themen im Bereich Lebensqualität abzuleiten. Diese allgemeinen Störwörter traten besonders häufig in den durch die Frequenzanalyse generierten Wortlisten auf. [51]

4 Es sei an dieser Stelle darauf hingewiesen, dass es für Forschende, die sich der automatisierten Differenzanalyse mittels von Drittparteien geschriebener Software bedienen, häufig schwierig ist, die Ergebnisse dieser Analysen intersubjektiv nachvollziehbar und reproduzierbar zu kommunizieren. Der Veröffentlichung der Algorithmen, die kommerziellen Softwarelösungen zugrunde liegen, stehen beachtliche lizenz- und datenschutzrechtliche Probleme gegenüber (WIEDEMANN & LEMKE 2016).

Zur Entfernung der Störwörter gingen zwei Forschende die vier Listen zunächst unabhängig voneinander durch, um mögliche Störwörter zu identifizieren. Danach wurden die Ergebnisse dieser Durchläufe abgeglichen, und es ergab sich eine prozentuale Übereinstimmung zwischen 87 und 93 Prozent bei der Identifikation der Störwörter. Anschließend wurden nur solche Störwörter aus den Listen entfernt, die von beiden Forschenden übereinstimmend codiert wurden. [52]

In Tabelle 1 ist exemplarisch (für Frage 2 des "Bürgerdialogs") das Ergebnis dieser Vorgehensweise dargestellt. Bei einer ersten Betrachtung dieser Liste fällt auf, dass die Schlagwortliste der Frequenzanalyse deutlich kürzer ist als die der Differenzanalyse. Das liegt daran, dass erstere deutlich mehr Störwörter enthielt, die von beiden Forschenden übereinstimmend als solche identifiziert wurden. Dies ist ein Hinweis darauf, dass die Differenzanalyse tatsächlich besser geeignet ist, um aussagekräftige Schlagwörter zu identifizieren, die ein spezifisches Textkorpus inhaltlich charakterisieren.

Schlagwort	Rang Frequenz	Rang Differenz	Schlagwort	Rang Frequenz	Rang Differenz
Kind	1	7	Generation	35	53
Bürger	2	4	Job	36	-
Freiheit	3	1	Straße	37	-
Sicherheit	4	3	Miteinander	38	12
Arbeit	5	10	Krieg	38	-
Bildung	6	2	Arbeitsplatz	38	-
Politik	7	9	Lohn	38	48
Geld	8	41	Gleichberechtigung	-	20
Familie	9	25	Bildungssystem	-	23
Zeit	10	-	Hebamme	-	24
Staat	10	22	Grundgesetz	-	27
Politiker	12	26	Gesundheitssystem	-	28
Demokratie	13	6	Zufriedenheit	-	29
Regierung	14	-	Sozialsystem	-	30
Wohlstand	15	5	Nachhaltigkeit	-	31
Wirtschaft	16	38	Lebensstandard	-	33
Schule	17	49	Absicherung	-	34
Recht	18	-	Freizeit	-	35
Zukunft	18	37	Mitmensch	-	36
Friede	18	8	Vielfalt	-	39

Schlagwort	Rang Frequenz	Rang Differenz	Schlagwort	Rang Frequenz	Rang Differenz
Kultur	21	16	Gleichheit	-	40
Gesundheit	22	14	Entfaltung	-	42
Gesetz	23	-	Meinungsfreiheit	-	43
Beruf	24	21	Grundeinkommen	-	44
Natur	25	17	Respekt	-	45
Flüchtling	26	-	Einkommen	-	46
Rente	27	19	Bürokratie	-	47
Umwelt	28	12	Ehrlichkeit	-	50
Infrastruktur	29	18	Ungerechtigkeit	-	51
Volk	30	32	Menschlichkeit	-	52
Stadt	31	-	Gesundheitswesen	-	54
Elter	32	-	Wohnraum	-	55
Gerechtigkeit	32	11	Chancengleichheit	-	56
Toleranz	34	15			

Tabelle 1: Wortranking nach Frequenz- und Differenzanalyse zu Frage 2 des "Bürgerdialogs" der Bundesregierung (in Anlehnung an WALDHERR et al. 2015, S.63-64). Siehe Legende zu Tabelle 1 in [Anhang 1](#). [53]

3.2.5 Qualitative Ableitung möglicher Oberkategorien

Die Listen der Frequenz- und Differenzanalyse wurden in der Folge manuell mit dem Ziel weiterbearbeitet, Oberkategorien des Konstrukts Lebensqualität abzuleiten. Hierzu war es nötig, thematisch ähnliche Begriffe unter Oberbegriffen zusammenzufassen, sodass als Ergebnis schließlich eine kürzere, nicht redundante Liste von Oberbegriffen entstand. Dieser Prozess sollte intersubjektiv nachvollziehbar verlaufen. Daher gingen zwei Forschende die vier Schlagwortlisten wieder unabhängig voneinander und regelgeleitet durch. Folgende Regeln wurden vorab formuliert und strukturierten den Prozess:

1. Bedeutungsähnliche, verwandte Schlagwörter wurden zu einer Kategorie zusammengefasst. Als verwandte Begriffe wurden auch Gegensatzpaare wie *Frieden – Krieg*, *Gerechtigkeit – Ungerechtigkeit* verstanden.
2. Zur Benennung der Oberkategorie wurde dem allgemeineren Schlagwort der Vorrang vor dem spezifischeren gegeben.
3. Ebenso wurde das ranghöhere Schlagwort vor dem rangniederen bevorzugt. [54]

Jede/r Forschende entwickelte auf diese Weise eine Liste mit möglichen Oberkategorien, denen wiederum andere Schlagwörter zugeordnet waren. Beim

Abgleich dieser Listen zeigte sich, dass die Forschenden überwiegend die gleichen Schlagwörter als Oberkategorien wählten. Von allen benannten Oberkategorien fanden sich 78 Prozent auf beiden Listen. Auch hatten die Forschenden zwei Drittel der Schlagwörter übereinstimmend diesen Oberkategorien zugeordnet. Die Abweichungen wurden nach dem Abgleich der Ergebnisse diskutiert, bis ein Konsens über die Oberkategorien gefunden wurde. Schlagwörter, die sich auch nach längerer Diskussion nicht zweifelsfrei einer einzigen Oberkategorie zuordnen ließen, wurden mehreren Oberkategorien gleichzeitig zugeordnet. Diese Form der Zuordnung von Materialeinheiten zu mehr als einer Kategorie ist ein bei der qualitativen Kategorienbildung gängiges Verfahren (vgl. z.B. MAYRING & FENZL 2014, S.553), und die erzielten Übereinstimmungen liegen im akzeptablen Rahmen. [55]

Tabelle 2 enthält die auf diese Weise induktiv aus dem Textkorpus abgeleiteten möglichen Oberkategorien. Die Oberkategorien haben unterschiedliche Abstraktionsgrade, z.B. sind unter dem Begriff *Politik* mit den Schlagwörtern *Demokratie*, *Lobbyismus* und *Bürokratie* potenziell sehr unterschiedliche Aspekte versammelt. Hingegen ist der Oberkategorie-Begriff *Rente* vergleichsweise konkret und wird daher auch nur durch ein einziges weiteres Schlagwort (*Altersarmut*) ergänzt. [56]

Teilweise ließen sich die konkreteren Begriffe auch anderen Oberkategorien zuordnen. So könnte *Familie* als Unteraspekt der Kategorie *Miteinander* verstanden werden, als eine Form sozialer Beziehungen neben engen Bezugspersonen und anderen Mitmenschen. Da der Begriff der *Familie* aber so prominent in den Wortrangelisten auftauchte und auch eine hohe Frequenz im Textkorpus aufwies, wurde er als eigene Oberkategorie berücksichtigt.

Oberkategorien	Zugeordnete Schlagwörter
Arbeit	Beruf, Job, <i>Einkommen</i> , Arbeitsplatz, Lohn, <i>Work-Life-Balance</i> , Grundeinkommen
Bildung	Schule, Ausbildung, Bildungssystem
Familie	Kind, Elter, Mutter, Elternteil
Freiheit	Selbstverwirklichung, Selbstbestimmung, Meinungsfreiheit, Entfaltung
Gerechtigkeit	Recht, Gesetz, soziale Gerechtigkeit, Chancengleichheit, Gleichberechtigung, Grundgesetz, Gleichheit, Ungerechtigkeit, Sozialsystem
Gesundheit	Hebamme, Gesundheitssystem, Gesundheitswesen
Infrastruktur	Stadt, Straße, Wohnraum
Kultur	
Miteinander	Freund, Umfeld, <i>Mitmensch</i>
Nachhaltigkeit	Zukunft, Generation, Ressource

Oberkategorien	Zugeordnete Schlagwörter
Politik	Regierung, Staat, Politiker, Demokratie , Volk, Bürger, Bundesregierung, Lobbyismus, Bürokratie, Ehrlichkeit, <i>Teilhabe</i>
Rente	Altersarmut
Sicherheit	Friede, Krieg, Absicherung
Toleranz	Flüchtling, Menschlichkeit, <i>Mitmensch</i> , Vielfalt,
Umwelt	Natur, Umweltschutz
Wohlstand	Geld, Wirtschaft , <i>Einkommen</i> , Steuer, Euro, Lebensstandard, Auskommen, Marktwirtschaft, Armut
Zeit	Freizeit, <i>Work-Life-Balance</i>
Zufriedenheit	Glück, Respekt, <i>Teilhabe</i> , Wertschätzung

Tabelle 2: Oberkategorien mit zugeordneten Schlagwörtern (aus WALDHERR et al. 2015, S.66). Siehe Legende zu Tabelle 2 in [Anhang 2](#). [57]

3.2.6 Manuelle Validierung der thematischen Oberkategorien

Laut WIEDEMANN und LEMKE (2016, S.410) sollten "Interpretationen auf Basis algorithmisch identifizierter Strukturen in einem close reading von Stichproben semantischer Einheiten (z.B. Dokumente, Absätze oder Sätze), welche diese Strukturen prominent enthalten, validiert werden". Im Sinne dieser Idee des *blended reading* wurden für alle als mögliche Oberkategorien identifizierten Schlagwörter Stichproben von mindestens $n = 5$ Dokumenten, die den betreffenden Begriff enthielten, gelesen, um die vorgenommenen Zuordnungen zu validieren. Standen die Begriffe in den gelesenen Antworten thematisch für sich und wurden nicht im Kontext anderer Begriffe verwendet, wurden sie als Oberkategorien bestätigt (Beispiele: *Umwelt*, *Gesundheit*, *Bildung*). Kamen die Begriffe hingegen in unterschiedlichen Bedeutungskontexten anderer, übergeordneter Begriffe vor, wurden sie als potenzielle Unterkategorien notiert und schieden als eigenständige Oberkategorien aus (Beispiele: *Natur*, *Gesundheitssystem*, *Schule*). Konnten die Begriffe nicht sinnvoll zur Kategorienbildung beitragen, da sie kein inhaltliches Thema beschrieben, wurden sie als Störwörter vermerkt und in eine separate Liste aufgenommen (Beispiele: *Volk*, *Bürger*). [58]

Dieses Vorgehen war auch erforderlich, um sogenannte "polyvalente Schlagwörter" (STULPE & LEMKE 2016, S.46f.) identifizieren und handhaben zu können. Polyvalent sind Schlagwörter, sofern sie einerseits Bestandteil der Alltagssprache sind, andererseits aber auch auf ein gesuchtes Konzept referieren. Ein Beispiel ist das Wort *Alter*. Es könnte sich zum einen der Kategorie Rente zuordnen lassen, andererseits ist das Wort Teil des allgemeinen Sprachgebrauchs, etwa wenn über das Alter konkreter Personen oder die Nutzungsdauer von Gegenständen gesprochen wird. Zudem ist *Alter* ein umgangssprachlicher Ausdruck. In einem solchen Fall musste eine Stichprobe

aus der zu analysierenden Dokumentenkollektion daraufhin überprüft werden, ob das betreffende polyvalente Schlagwort überwiegend im Kontext der interessierenden Kategorie oder überwiegend alltagssprachlich verwendet wurde (STULPE & LEMKE 2016). [59]

Zusätzlich zu den bereits zuvor identifizierten 18 Oberkategorien (vgl. Tabelle 2) wurden durch diesen Prozess der Validierung zwei weitere Begriffe als mögliche Oberkategorien aufgenommen: *Demokratie* und *Wirtschaft*. So zeigte sich, dass das Stichwort *Demokratie* meist als Wert für sich genannt wurde, vergleichbar z.B. mit *Freiheit*, und sich weniger auf die Ausgestaltung des politischen Systems und politische Prozesse bezog; es wurde also meist nicht im Kontext von *Politik* genannt, sondern eigenständig. Auf ähnliche Weise kristallisierte sich *Wirtschaft* als eigene Oberkategorie heraus, da der Begriff meist eigenständig in Bezug auf das Wirtschaftssystem verwendet wurde und nicht im Kontext der Kategorie *Wohlstand*, die unmittelbaren Bezug zu Lebensqualität als Ergebnis wirtschaftlicher Prozesse hat. [60]

3.3 Ausdifferenzierung von thematischen Unterkategorien mittels qualitativer Inhaltsanalyse

Die durch die *Text-Mining*-Analyse identifizierten Oberkategorien hatten überwiegend ein recht hohes Abstraktionsniveau (z.B. *Arbeit*, *Wirtschaft*). Um herauszufinden, unter welchen Aspekten die Antwortenden diese allgemeinen Themen im Online-Dialog genauer ansprachen, wurden die eingegebenen Texte systematisch qualitativ analysiert. Dabei wurde auf die Methode der zusammenfassenden, qualitativen Inhaltsanalyse nach MAYRING (2010) zurückgegriffen, die sich sehr gut für die Auswertung offener Antworten in Fragebögen eignet und sich hier als Standard etabliert hat. [61]

Die qualitative Analyse wurde durchgeführt, da menschliche Codierende besser in der Lage sind, spezifische Aussagen, wie sie in den Antworten des "Bürgerdialogs" vorkamen, zu erfassen und den im Vorfeld mittels Text Mining identifizierten Oberkategorien als Unterthemen zuzuordnen. Da die Bürgerinnen und Bürger in ihren Antworten häufig eine Vielzahl unterschiedlicher Themen mit unterschiedlichem Abstraktionsgrad ansprachen, erwiesen sich die Möglichkeiten, spezifische Unterkategorien mit ko-okkurrenzbasierten Verfahren ausdifferenzieren, als sehr begrenzt. Hier zeigten sich menschliche Codierende als geeigneter. [62]

Zunächst wurde jeweils aus allen Dokumenten, die das betreffende Keyword einer Oberkategorie (z.B. *Familie*) beinhalteten, eine Zufallsstichprobe von $n = 100$ Dokumenten gezogen (je 50 Dokumente zu jeder der beiden Fragen). Für diejenigen Schlagwörter, die ein Konstrukt zwar abstrakt gut beschreiben, aber in der Alltagssprache nicht verwendet werden, wurden Wortlisten gebildet (STULPE & LEMKE 2016). Zum Beispiel wurde die Stichprobe für die auf dem Schlagwort *Miteinander* basierende Oberkategorie mit einer Wortliste (*Freund*, *Nachbar* etc.) gezogen. [63]

Um auch den Prozess der Ausdifferenzierung von Kategorien intersubjektiv nachvollziehbar zu gestalten, arbeiteten zu jeder Oberkategorie jeweils Paare von Forschenden unabhängig voneinander am Material. Jeweils eine Person bearbeitete alle ausgewählten Antworten zu Frage 1 mit dem betreffenden Schlagwort, die andere diejenigen zu Frage 2. Um möglichen materialabhängigen Fehlerquellen zu begegnen, wurde darüber hinaus zusätzlich darauf geachtet, dass alle Forschenden über alle Stichproben hinweg in etwa gleich viele Dokumente aus Frage 1 und Frage 2 sichteten. [64]

Um die Interpretationsleistung gering zu halten und den Einfluss subjektiver WahrnehmungsfILTER auf die Codierung zu minimieren, wurden die verschiedenen Themenaspekte sehr detailliert erhoben, d.h. auf einem sehr geringen Abstraktionsgrad, der bis hin zur Nennung konkreter Maßnahmen reichte. Dabei orientierten wir uns an den Regeln, die MAYRING (2010) formulierte, um eine qualitative, zusammenfassende Inhaltsanalyse systematisch und intersubjektiv nachvollziehbar durchzuführen. Diese beinhalten die Arbeitsschritte Paraphrasierung (Streichung nicht inhaltstragender Textbestandteile und Reduzierung auf die grammatikalische Kurzform), Generalisierung auf das Abstraktionsniveau, erste Reduktion (um bedeutungsgleiche Paraphrasen innerhalb der Auswertungseinheiten) und zweite Reduktion (Bündelung und Integration ähnlicher Aussagen). Abbildung 2 verdeutlicht das Vorgehen bei der qualitativen Ausdifferenzierung der Unterkategorien gemäß den von MAYRING formulierten Analyseschritten. [65]

Die Forschenden identifizierten unabhängig voneinander für jede Oberkategorie thematische Unterkategorien auf zwei Ebenen aus dem Material. So konnten allgemeinere von spezifischeren Aspekten unterschieden werden. Jede gefundene Kategorie wurde mit einem Zahlencode, einem Namen und einer inhaltlichen Kurzbeschreibung versehen. Ferner wurden jeweils prägnante Schlagwörter, die im Zusammenhang mit der Dimension aufkamen, sowie prägnante Zitate im Wortlaut aufgenommen. Um eine möglichst große Vollständigkeit zu gewährleisten, wurden in diesem Schritt alle genannten Aspekte unabhängig von ihrer Häufigkeit aufgenommen. Es zeigte sich, dass sich ca. ab dem 35. codierten Dokument eine Sättigung einstellte, d.h., die Codierung weiterer Dokumente ergab nur vereinzelt neue Aspekte. Somit erwies sich die gewählte Stichprobengröße als ausreichend für eine möglichst vollständige Inventur der wichtigsten im "Bürgerdialog" genannten Themenaspekte. [66]

Eine solche datengetriebene, vollständige Inventur der genannten Themenaspekte hat gegenüber der theoretischen oder deduktiven Erstellung von inhaltlichen Kategorien einen zusätzlichen Vorteil. Kategorien, die zwar theoretisch Sinn für den zu untersuchenden Sachverhalt ergeben, im Material aber nicht vorkommen, werden gar nicht erst ins Codebuch aufgenommen. Das erleichtert die Codierung in der anschließenden quantitativen Analyse, da keine Auswahlmöglichkeiten gegeben sind, die empirisch keine Rolle spielen (vgl. dazu WETTSTEIN 2014). [67]

Im Anschluss an die individuelle Identifikation der Themenaspekte glichen die Forschenden ihre Befunde zu Frage 1 und 2 ab. Sie führten ihre gefundenen Kategorien zusammen und ließen inhaltlich übereinstimmende Kategorien ineinander aufgehen, sodass am Ende *ein* ausdifferenziertes Kategoriensystem für jede Oberkategorie vorlag. [68]

Im Anschluss daran wurden alle ausdifferenzierten Oberkategorien in ein übergreifendes Kategoriensystem zusammengeführt, wobei jede Kategorie mit ihren Unterkategorien in mehreren Überarbeitungsrounden im Team diskutiert wurde. Dabei wurden Unterkategorien, die sich in verschiedenen Oberkategorien fanden, eindeutig einer Oberkategorie zugeordnet. Redundante Aspekte innerhalb einer Oberkategorie wurden zusammengeführt, zu spezifische Kategorien zu allgemeineren abstrahiert und nicht trennscharfe Kategorien besser voneinander abgegrenzt. In diesem Zuge wurden auch die inhaltlichen Beschreibungen der Kategorien überarbeitet und insbesondere um Hinweise zu Abgrenzungen ergänzt. [69]

Des Weiteren wurden einzelne Oberkategorien, die in weiten Teilen sehr ähnliche Aspekte enthielten, fusioniert und andere Oberkategorien, deren Aspekte sich auf viele verschiedene andere Oberkategorien verteilen, aufgelöst. Am Ende dieses Prozesses umfasste das vollständige Kategoriensystem schließlich 17 Oberkategorien auf oberster Ebene, 161 Unterkategorien auf Ebene zwei und 255 Unterkategorien auf Ebene drei.

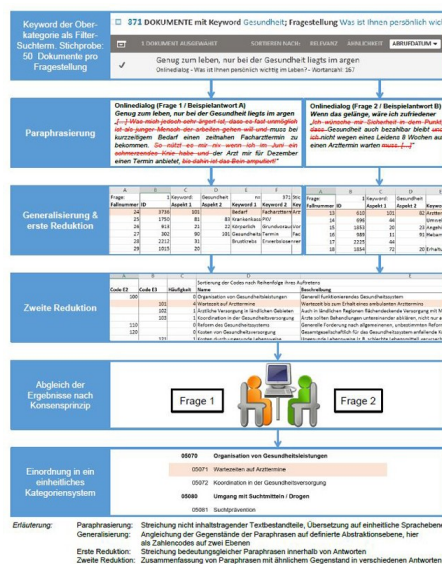


Abbildung 2: Von den Einzelantworten zur Unterkategorie in Anlehnung an die zusammenfassende Inhaltsanalyse nach MAYRING (2010, S.606). Bitte klicken Sie [hier](#) oder auf die Abbildung für eine Vergrößerung. [70]

Die *Oberkategorien* beinhalteten alle großen Themenbereiche, die aus Sicht der Bürgerinnen und Bürger zur Lebensqualität in Deutschland beitragen. Die meisten dieser Oberkategorien bezogen sich dabei auf einzelne Lebensbereiche oder Politikfelder sowie damit zusammenhängende materielle und immaterielle

Ressourcen: *Arbeit, Wirtschaft, Wohlstand, Bildung, Gesundheit, innere und äußere Sicherheit, Kultur, Infrastruktur und Verkehr sowie Umwelt und Nachhaltigkeit*. Darüber hinaus gingen aus dem "Bürgerdialog" auch Oberkategorien hervor, deren Fokus eher im Bereich des sozialen Miteinanders zu verorten ist: *Altersvorsorge und soziale Sicherheit, Familie und soziale Beziehungen*. Auch individuelle Werte und Normen kristallisierten sich als wesentliche Aspekte für das gesellschaftliche Zusammenleben und die Lebensqualität heraus: *Demokratie und politischer Prozess, Freiheit, Gerechtigkeit, Toleranz und Integration sowie subjektive Zufriedenheit*. [71]

Die *Unterkategorien* umfassten jeweils spezifische inhaltliche Bereiche zu den Oberthemen, zu denen die Bürgerinnen und Bürger Ansprüche bzw. Forderungen formulierten. Sie waren hierarchisch auf zwei Ebenen geordnet. Während auf Ebene 2 die jeweilige Oberkategorie auf einen inhaltlichen Teilbereich präzisiert wurde (z.B. *Vereinbarkeit von Familie und Beruf* in der Oberkategorie *Familie*), befanden sich auf Ebene 3 konkrete Forderungen und Ansprüche bezüglich dieser Themengebiete (z.B. *kostenlose Kinderbetreuung*). [72]

Da der "Bürgerdialog" während der Auswertungsphase noch weiterlief, gab es in jeder Oberkategorie sogenannte "Themenspeicher". Hier wurden Aspekte gesammelt, die sich zwar klar einer Kategorie auf Ebene 1 zuordnen ließen, aber in keine vorhandene Unterkategorie passten. Sonstige Aspekte zu Lebensqualität, die sich auch den inhaltlichen Bereichen der vorhandenen Oberkategorien entzogen, wurden zudem in einem allgemeinen Themenspeicher gesammelt. Hierdurch wurde sichergestellt, dass kein Aspekt verloren ging, der nicht zuvor in der qualitativen Inhaltsanalyse identifiziert worden war. Vor der Endauswertung wurden alle in den Themenspeichern vorhandenen Kategorien überprüft und an inhaltlich passender Stelle in das Kategoriensystem integriert. Ein umfassender Überblick über das Kategoriensystem und die Dokumentation des Codebuchs finden sich im Abschlussbericht des Projekts (WALDHERR et al. 2015). [73]

4. Mögliche Anwendungsgebiete des Verfahrens

Die beschriebene Methode wurde entwickelt, um auf Basis offener Antworten von Bürgerinnen und Bürgern in Kurztextform ein Kategoriensystem zum Thema Lebensqualität zu erstellen. Die Methode ist aber keineswegs auf diese Thematik oder eine bestimmte Textform begrenzt. Sie bietet sich unserer Ansicht nach immer dann an, wenn Texte analysiert werden müssen, die in Inhalt und Form sehr uneinheitlich ausfallen (bspw. in Bezug auf Satz- und Absatzgliederung, Art und Häufigkeit verwendeter Satzschlusszeichen, Fließtext vs. stichpunktartige Aufzählungen). Textgattungen, für die wir die vorliegende Methode als geeignet erachten, sind daher beispielsweise Nutzer/innenkommentare auf den Seiten von Online-Nachrichtenportalen. Auch andere nutzer/innengenerierte Textbeiträge im Internet eignen sich für eine derart gestaltete Exploration. [74]

Bezüglich zu untersuchender Themen sind nahezu keine Grenzen gesetzt. Zu beachten ist, dass es sich hier um ein Verfahren handelt, welches das

vorliegende Textmaterial zu einer bestimmten Thematik explorieren und mittels einer anschließenden, standardisierten Inhaltsanalyse quantifizierbar machen soll. Somit eignet sich die Methodik insbesondere, um in einem Textkorpus vorkommende Themen und Unterthemen auf mehreren Ebenen erfassbar zu machen. So könnten etwa die Haupt- und Unterthemen evaluiert werden, die von Nutzerinnen und Nutzern eines Internetforums oder in einer Diskussion auf Social Media angesprochen oder im Zusammenhang mit spezifischen Artikeln zu einem bestimmten Thema auf einer Online-Nachrichtenseite in Nutzer/innenkommentaren diskutiert werden. Dabei spielt es zunächst keine Rolle, ob die anschließende quantitative Inhaltsanalyse automatisch oder manuell erfolgen soll. [75]

Zur Auswertung des "Bürgerdialogs" wurde das entwickelte Kategoriensystem für eine quantitative Inhaltsanalyse eingesetzt, in der mehrere Codierende alle Texte manuell einer oder mehreren Kategorien von Lebensqualität zuordneten (für eine Beschreibung des Vorgehens und der Ergebnisse der quantitativen Inhaltsanalyse vgl. WALDHERR et al. 2015). Es ist aber ebenfalls denkbar, ein mit dieser Vorgehensweise entwickeltes Kategoriensystem als Grundlage für eine automatische Inhaltsanalyse einzusetzen. So könnten etwa die im Prozess der Kategorienbildung erfassten charakteristischen Schlagwörter genutzt werden, um Wortlisten für eine diktionsbasierte Analyse zu generieren. Ebenso ist es möglich, auf Basis des Kategoriensystems zunächst eine Stichprobe von Dokumenten manuell zu codieren und diese dann als Trainingsmaterial für einen lernenden Klassifikator einzusetzen. In einer halbautomatischen Inhaltsanalyse können diese Prozesse sogar parallel und miteinander verflochten ablaufen. [76]

Bei der Entscheidung, ob eine Inhaltsanalyse manuell oder automatisch durchgeführt werden soll, ist immer die Relation aus Forschungsaufwand und -nutzen zu bedenken. Der Aufwand bei automatischen Analysen ist im Regelfall zu Beginn höher als bei der manuellen Analyse. Jedoch können mit diesen Verfahren letztendlich ausgesprochen große Fallzahlen bewältigt werden, die mit menschlichen Codierenden in dieser Form nicht bearbeitbar sind (SCHARKOW 2013a). [77]

5. Fazit und Ausblick

Herzstück jeder Inhaltsanalyse ist ein Kategoriensystem, mit dem Texte (oder visuelle Elemente) systematisch beobachtet und beschrieben werden können. Bestrebungen der Automatisierung von Inhaltsanalysen konzentrieren sich hingegen meist darauf, die Auswahl von relevantem Untersuchungsmaterial oder den Prozess der Codierung an den Computer zu delegieren. Die Bildung der hier zum Einsatz kommenden Kategorien wird jedoch noch meist händisch vorgenommen. In unserem Beitrag zeigen wir, dass sich auch die Entwicklung eines Kategoriensystems durch *Text-Mining*-Verfahren gewinnbringend unterstützen und zumindest teilautomatisieren lässt. [78]

Moderne Verfahren des *Text Mining* ermöglichen es heute relativ ressourceneffizient, qualitative Ansätze der Exploration von Textmaterial in

quantitativ sehr großen Dimensionen durchzuführen. Waren und sind manuelle Verfahren stets nur mit sehr begrenztem Untersuchungsmaterial durchführbar, so kann durch automatisierte Verfahren sehr schnell ein grober thematischer Überblick über ein umfangreiches Textkorpus gewonnen werden. Ein Verfahren, das gewissermaßen "auf Knopfdruck" automatisch ein valides Kategoriensystem erstellt, ist allerdings noch in weiter Ferne: "There is no, and probably will never be a 'one button' solution to CATA [computer assisted text analysis], because of the simple fact, that generic approaches are not appropriate to satisfy specific and complex research needs" (WIEDEMANN 2013, §53). [79]

Unser Ansatz für die induktive Materialexploration und Kategorienbildung ist daher, induktive lexikometrische *Text-Mining*-Verfahren mit lange erprobten und etablierten Vorgehensweisen der qualitativen Inhaltsanalyse (MAYRING 2010; SCHREIER 2014) zu kombinieren und so den Prozess des *blended reading* (STULPE & LEMKE 2016) regelgeleitet und intersubjektiv nachvollziehbar zu gestalten: Zunächst werden mittels Frequenz- und Differenzanalyse Schlagwörter aus einem umfangreichen Textkorpus extrahiert (*distant reading*), die breite und häufig genannte Themenfelder als Oberkategorien beschreiben. Diese werden in einem qualitativen Verfahren verdichtet, hierarchisch sortiert und validiert (*close reading*). Zu allen Schlagwörtern, die als Oberkategorien identifiziert werden, werden dann themenspezifische Zufallsstichproben von Dokumenten gezogen und nach den Regeln der qualitativen, zusammenfassenden Inhaltsanalyse bearbeitet, um Unterasspekte auf spezifischeren Abstraktionsebenen zu erfassen. Am Ende dieses Prozesses steht ein ausdifferenziertes, themenspezifisches Kategoriensystem, das als inhaltsanalytisches Forschungsinstrument für die Bearbeitung vielfältiger Forschungsfragen einsetzbar ist. [80]

Durch die Kombination beider Ansätze ist es möglich, die Stärken beider Verfahren zu nutzen und durch eine wechselseitige Ergänzung Schwächen eines Ansatzes durch die Stärken des anderen Ansatzes auszugleichen. Während die automatische Inhaltsanalyse effizient und reliabel sehr große Textmengen bearbeiten kann, sind menschliche Codierende hinsichtlich ihres Kontextwissens sowie ihrer Interpretationsfähigkeiten überlegen. Die Kombination beider Verfahren vereint das beste beider Welten (LEJEUNE 2011; WETTSTEIN 2014). [81]

Grenzen der beschriebenen Vorgehensweise liegen darin, dass die hier eingesetzten *Text-Mining*-Verfahren der Frequenz- und Differenzanalyse auf relativ einfachen Algorithmen basieren und zudem nur einzelne Schlagwörter (im konkreten Fall zudem fast ausschließlich Substantive) berücksichtigen, aus denen nicht unbedingt in jedem Fall auf den Inhalt eines Textes geschlossen werden kann (WETTSTEIN 2014; WIEDEMANN 2013). Vorteil ist jedoch, dass diese Verfahren mittlerweile breit erprobt und in gängigen Softwarepaketen implementiert sind, sodass die Durchführung auch für computerwissenschaftliche Lai/innen machbar ist. Auch sind die resultierenden Wortlisten relativ klar und intuitiv interpretierbar. [82]

Das hier vorgestellte Verfahren eignet sich besonders gut für sehr heterogenes und wenig strukturiertes Textmaterial (wie im vorgestellten Fall des

"Bürgerdialogs") und zur Entwicklung eines komplexen, hierarchischen Kategoriensystems mit mehreren Abstraktionsebenen. Liegen homogenere Dokumente mit nur einem oder wenigen genannten Themen vor und wird ein weniger komplexes Kategoriensystem angestrebt, ist auch denkbar, den Prozess der Kategorienbildung mit ko-okkurrenzbasierten Verfahren des *Themenclustering* (wie von WETTSTEIN 2012 vorgeschlagen) oder Verfahren des *Topic Modeling* (GUO, VARGO, PAN, DING & ISHWAR 2016; MAIER et al. 2018) noch weiter zu automatisieren. [83]

Anhang 1: Legende zu Tabelle 1

Sortierung:	Rangfolge der häufigsten Schlagwörter ohne Störwörter, absteigend sortiert nach Frequenz. Rechte Spalte: Rangfolge der Schlagwörter nach der Differenzanalyse, ohne Störwörter.
Frequenz:	Anzahl der Dokumente, in denen das entsprechende Schlagwort mindestens einmal auftritt.
Differenz:	Logarithmierter Wert, der angibt, inwieweit die tatsächlich beobachtete Frequenz des Schlagworts im Analysekorpus vom statistischen Erwartungswert auf Basis des Referenzkorpus abweicht.
Datengrundlage:	N = 1.955 Dokumente mit im Zeitraum vom 13.04.-05.08.2015 gegebenen Antworten auf Frage 2: "Was macht Ihrer Meinung nach Lebensqualität in Deutschland aus?"
Übereinstimmungsquote beim Entfernen der Störwörter:	Die Übereinstimmungsquote von zwei unabhängig voneinander arbeitenden Forschenden bei der Einteilung von Schlagwörtern in für die Analyse relevante Wörter und Störwörter betrug 91,3 Prozent (Frequenz) bzw. 92,9 Prozent (Differenz).
Entfernte Störwörter (Frequenz):	Mensch, Land, Leben, Jahr, Gesellschaft, Möglichkeit, Welt, Angst, Frau, Problem, Ding, Frage, Chance, Thema, Wert, Bereich, Leute, Qualität, Teil, Versorgung, Tag, Bevölkerung, System, Gefühl, Interesse, Beispiel, Kosten, Weg
Entfernte Störwörter (Differenz):	Leben, Mensch, Gesellschaft, Land, Möglichkeit, Versorgung, leben, Lebenswert, Angst, Qualität, Ding, Förderung, gut

Anhang 2: Legende zu Tabelle 2

Sortierung:	alphabetisch
Kursive Schlagwörter:	Schlagwörter, die sich auch nach Absprache nicht zweifelsfrei einer Oberkategorie zuordnen ließen, wurden in bis zu zwei Oberkategorien gleichzeitig eingeordnet.
Kursive und gefettete Schlagwörter:	Schlagwörter, die sich im weiteren Prozess der Validierung anhand von Stichproben des Materials als eigene Kandidaten für Oberkategorien herausstellten.
Übereinstimmung bei Festlegung der Oberkategorien:	Die Übereinstimmung bei der Festlegung, welche und wie viele Schlagwörter als Oberkategorien definiert werden sollten, lag bei zwei unabhängig voneinander arbeitenden Forschenden bei 78,3 Prozent.
Übereinstimmung bei Zuordnung der Schlagwörter zu den Oberkategorien:	66,2 Prozent der Schlagwörter wurden von zwei unabhängig voneinander arbeitenden Forschenden übereinstimmend der identischen Oberkategorie zugeordnet.

Literatur

- Baron, Alistair; Rayson, Paul & Archer, Dawn (2009). Word frequency and keyword statistics in historical corpus linguistics. *Anglistik: International Journal of English Studies*, 20(1), 41-67, [http://www.research.lancs.ac.uk/portal/en/publications/-\(bf0d0d9d-10e1-4a1c-ad22-fc2635f18776\).html](http://www.research.lancs.ac.uk/portal/en/publications/-(bf0d0d9d-10e1-4a1c-ad22-fc2635f18776).html) [Zugriff: 9. Januar 2019].
- Bestgen, Yves (2014). Inadequacy of the chi-squared test to examine vocabulary differences between corpora. *Lit Linguist Computing*, 29(2), 164-170.
- Biemann, Chris; Heyer, Gerhard; Quasthoff, Uwe & Richter, Matthias (2007). The Leipzig corpora collection: Monolingual corpora of standard size. *Proceedings of Corpus Linguistic 2007*, Centre for Corpus Research, Birmingham, UK, <https://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2007/190Paper.pdf> [Zugriff: 9. Januar 2019].
- Blei, David M.; Ng, Andrew Y. & Jordan, Michael I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 933-1022.
- Bordag, Stefan (2008). A comparison of co-occurrence and similarity measures as simulations of context. In Alexander Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing* (pp.52-63). Berlin: Springer.
- Bundesregierung der Bundesrepublik Deutschland. (2015). *Deutschland im Dialog*, https://buergerdialog.gut-leben-in-deutschland.de/DE/Home/home_node [Zugriff: 30. Januar 2017].
- Ceron, Andrea; Curini, Luigi; Iacus, Stefano M. & Porro, Giuseppe (2014). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media and Society*, 16(2), 340-358.
- CID GmbH (2017). *The topic analyst ecosystem*, <https://cid.com/en/products-services/topic-analyst.html> [Zugriff: 30. Januar 2017].
- Clear, Jeremy (1992). Corpus sampling. In Gerhard Leitner (Hrsg.), *New directions in English language corpora* (S.21-31). Berlin: Mouton-de-Gruyter.
- Dunning, Ted (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19, 61-74, <http://dl.acm.org/citation.cfm?id=972454> [Zugriff: 9. Januar 2019].
- Dzudzek, Iris; Glasze, Georg; Mattissek, Annika & Schirmel, Henning (2009). Verfahren der lexikometrischen Analyse von Textkorpora. In Georg Glasze & Annika Mattissek (Hrsg.), *Handbuch*

Diskurs und Raum. Theorien und Methoden für die Humangeographie sowie die sozial- und kulturwissenschaftliche Raumforschung (S.233-260). Bielefeld: transcript.

Früh, Werner (2007). *Inhaltsanalyse: Theorie und Praxis*. Konstanz: UVK.

Geyken, Alexander (2007). The DWDS corpus: A reference corpus for the German language of the 20th century. In Christiane Fellbaum (Hrsg.), *Collocations and idioms: Corpus-based linguistics and lexicographic studies* (S.23-41). Birmingham: Continuum Press, <https://www.dwds.de/r> [Zugriff: 9. Januar 2019].

Grimmer, Justin & King, Gary (2011). General purpose computer-assisted clustering and conceptualization. *Proceedings of the National Academy of Science of the United States of America*, 108(7), 2643-2650, <https://doi.org/10.1073/pnas.1018067108> [Zugriff: 15. Januar 2019].

Günther, Elisabeth & Quandt, Thorsten (2016). Word counts and topic models. *Digital Journalism*, 4(1), 75-88.

Guo, Lei; Vargo, Chris J.; Pan, Zixuan; Ding, Weicong & Ishwar, Prakash (2016). Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. *Journalism & Mass Communication Quarterly*, 93(2), 332-359.

Heyer, Gerhard, Quasthoff, Uwe & Wittig, Thomas (2006). *Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse*. Bochum: W3L.

Hotho, Andreas; Nürnberger, Andreas & Paaß, Gerhard (2005). A brief survey of text mining. *LDV Forum – GLDV Journal for Computational Linguistics and Language Technology*, 20(1), 19-62.

Institut für Deutsche Sprache (2016). *Deutsches Referenzkorpus: Archiv der Korpora geschriebener Gegenwartssprache 2016-II (Release vom 30.09.2016)*. Mannheim: Institut für Deutsche Sprache, <http://www1.ids-mannheim.de/kl/projekte/korpora/releases.html> [Zugriff: 9. Januar 2019].

Kepplinger, Hans Mathias (2001). Der Ereignisbegriff in der Publizistikwissenschaft. *Publizistik*, 46(2), 117-139.

Leech, Geoffrey (1993). 100 million words of English: A description of the background, nature and prospects of the British National Corpus project. *English Today*, 9(1), 9-15.

Lejeune, Christophe (2011). From normal business to financial crisis ... and back again. An illustration of the benefits of Cassandre for qualitative analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 12(1), Art. 24, <http://dx.doi.org/10.17169/fqs-12.1.1513> [Zugriff: 15. Januar 2019].

Maier, Daniel; Waldherr, Annie; Miltner, Peter; Wiedemann, Gregor; Niekler, Andreas; Keinert, Alexa; Pfetsch, Barbara; Heyer, Gerhard; Reber, Ueli; Häussler, Thomas; Schmid-Petri, Hannah & Adam, Silke (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2-3), 93-118.

Mayring, Philipp (2001). Kombination und Integration qualitativer und quantitativer Analyse. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 2(1), Art. 6, <http://dx.doi.org/10.17169/fqs-2.1.967> [Zugriff: 16. Januar 2019].

Mayring, Philipp (2010). Qualitative Inhaltsanalyse. In Günther Mey & Katja Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S.601-613). Wiesbaden: VS.

Mayring, Philipp & Fenzl, Thomas (2014). Qualitative Inhaltsanalyse. In Nina Baur & Jörg Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S.543-556). Wiesbaden: Springer VS.

Moretti, Frank (2013). *Distant reading*. London: Verso.

Rayson, Paul & Garside, Roger (2000). Comparing corpora using frequency profiling. In Association for Computer Linguistics (Hrsg.), *WCC '00 Proceedings of the workshop on comparing corpora* (S.1-6). Hong Kong, <https://doi.org/10.3115/1117729.1117730> [Zugriff: 15. Januar 2019].

Rössler, Patrick (2010). *Inhaltsanalyse*. Konstanz: UVK.

Scharkow, Michael (2012). *Automatische Inhaltsanalyse und maschinelles Lernen*. Berlin: epubli, https://opus4.kobv.de/opus4-udk/frontdoor/deliver/index/docId/28/file/dissertation_scharkow_final_udk.pdf [Zugriff: 15. Januar 2019].

Scharkow, Michael (2013a). Automatische Inhaltsanalyse. In Wiebke Möhring & Daniela Schlütz (Hrsg.), *Handbuch standardisierte Erhebungsverfahren der Kommunikationswissenschaft* (S.289-306). Wiesbaden: VS Verlag für Sozialwissenschaften.

Scharkow, Michael (2013b). Thematic content analysis using supervised machine learning: An empirical evaluation using German online news. *Quality & Quantity*, 47(2), 761-773.

Scheu, Andreas M.; Vogelgesang, Jens & Scharkow, Michael (2018). Qualitative Textanalyse. Blaupause und Potentiale (teil)automatisierter Verfahren. In Andreas M. Scheu (Hrsg.), *Auswertung qualitativer Daten: Strategien, Verfahren und Methoden der Interpretation nicht-standardisierter Verfahren in der Kommunikationswissenschaft* (S.309-322). Wiesbaden: Springer VS.

[Schreier, Margrit](#) (2014). Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. *Qualitative Sozialforschung / Forum: Qualitative Social Research*, 15(1), Art. 18, <http://dx.doi.org/10.17169/fqs-15.1.2043> [Zugriff: 16. Januar 2019].

Sebastiani, Fabrizio (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1-47.

Silge, Julia & Robinson, David (2017). *Text mining with R: A tidy approach*. Sebastopol, CA: O'Reilly.

Sinclair, Stéfan & Rockwell, Geoffrey (2016). *Voyant tools* [Software], <https://voyant-tools.org/> [Zugriff: 9. Januar 2019].

Stone, Philip; Dunphy, Dexter; Smith, Marshall & Ogilvie, Daniel (1966). *The general inquirer: A computer approach to content analysis*. Cambridge, MA: MIT Press.

Stuckardt, Roland (2000). *Qualitative Inhaltsanalyse durch Computer: Ein uneinlösbarer Anspruch?* Berlin: Tenea.

Stulpe, Alexander & Lemke, Matthias (2016). Blended Reading. Theoretische und praktische Dimensionen der Analyse von Text und sozialer Wirklichkeit im Zeitalter der Digitalisierung. In Matthias Lemke & Gregor Wiedemann (Hrsg.), *Text Mining in den Sozialwissenschaften: Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse* (S.17-62). Wiesbaden: Springer VS.

Tribble, Christopher & Jones, Glyn (1997). *Concordances in the classroom*. Houston, TX: Athelstan.

Waldherr, Annie; Maier, Daniel; Miltner, Peter & Günther, Enrico (2016). Big data, big noise: The challenge of finding issue networks on the Web. *Social Science Computer Review*, 35(4), 427-443.

Waldherr, Annie; Heyer, Gerhard; Jähnichen, Patrick; Niekler, Andreas & Wiedemann, Gregor (2016). Mining big data with computational methods. In Gerhard Vowe & Philipp Henn (Hrsg.), *Political communication in the online world: Theoretical approaches and research designs* (S.201-217). New York, NY: Routledge.

Waldherr, Annie; Ostner, Sophia; Müller, Lars-Ole; Miltner, Peter; Stoltenberg, Daniela; Sona, Patrick & Pfetsch, Barbara (2015). *Wissenschaftliche Auswertung von Bürger- und Online-Dialogen zum Thema Lebensqualität in Deutschland*, <https://www.gut-leben-in-deutschland.de/SharedDocs/Downloads/DE/LB/Abschlussbericht-wissenschaftliche-Auswertung-Buergerdialog.pdf> [Zugriff: 9. Januar 2019].

Wettstein, Martin (2012). Term-Mapping zur komparativen Analyse öffentlicher Debatten. Eine Anwendung der Smallest Space Analysis für Inhaltsanalysen. In Birgit Stark, Melanie Magin, Olaf Jandura & Marcus Maurer (Hrsg.), *Methodische Herausforderungen komparativer Forschungsansätze* (S.138-158). Köln: von Halem.

Wettstein, Martin (2014). "Best of both worlds": Die halbautomatische Inhaltsanalyse. In Katharina Sommer, Martin Wettstein, Werner Wirth & Jörg Matthes (Hrsg.), *Automatisierung in der Inhaltsanalyse* (S.16-39). Köln: von Halem.

Wettstein, Martin (2015). Verfahren zur computergestützten Inhaltsanalyse in der Kommunikationswissenschaft. *Dissertation*, Philosophische Fakultät der Universität Zürich, <https://www.zora.uzh.ch/id/eprint/127459/1/20162838.pdf> [Zugriff: 15. Januar 2019].

Wiedemann, Gregor (2013). Opening up to big data: Computer-assisted analysis of textual data in social sciences. *Forum Qualitative Sozialforschung / Forum Qualitative Social Research*, 14(2), Art. 23, <http://dx.doi.org/10.17169/fqs-14.2.1949> [Zugriff: 9. Januar 2019].

Wiedemann, Gregor & Lemke, Matthias (2016). Text Mining für die Analyse qualitativer Daten. Auf dem Weg zu einer Best Practice? In Matthias Lemke & Gregor Wiedemann (Hrsg.), *Text Mining in den Sozialwissenschaften: Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse* (S.397-420). Wiesbaden: Springer VS.

Zu den Autorinnen und Autoren

Annie WALDHERR, Juniorprofessorin für digitalisierte Öffentlichkeiten am Institut für Kommunikationswissenschaft, Fachbereich Erziehungswissenschaft und Sozialwissenschaften, Westfälische Wilhelms-Universität Münster. Forschung und Lehre zu den Themen: digitalisierte Öffentlichkeiten, politische Kommunikation, Technologie- und Wissenschaftsdiskurse, *Computational Social Science*.

Kontakt:

Jun.-Prof. Dr. Annie Waldherr
Institut für Kommunikationswissenschaft,
Fachbereich Erziehungswissenschaft und
Sozialwissenschaften, Westfälische Wilhelms-
Universität
Bispinghof 9-14
D-48143 Münster

Tel.: +49 (0)251 83 23000

Fax: +49 (0)251 83 28394

E-Mail: annie.waldherr@uni-muenster.de

URL: <https://www.uni-muenster.de/Kowi/personen/annie-waldherr.html>

Lars-Ole WEHDEN, wissenschaftlicher Mitarbeiter am Institut für Kommunikationswissenschaft, Fachbereich Erziehungswissenschaft und Sozialwissenschaften, Westfälische Wilhelms-Universität Münster. Forschung und Lehre zu den Themen: digitale Öffentlichkeit, Nachrichten in sozialen Medien, *Virtual and Augmented Reality*.

Kontakt:

Lars-Ole Wehden, M.A.

Institut für Kommunikationswissenschaft,
Fachbereich Erziehungswissenschaft und
Sozialwissenschaften, Westfälische Wilhelms-
Universität
Bispinghof 9-14
D-48143 Münster

Tel.: +49 (0)251 83 21273

E-Mail: lars-ole.wehden@uni-muenster.de

URL: <https://www.uni-muenster.de/Kowi/personen/lars-ole-wehden.html>

Daniela STOLTENBERG, wissenschaftliche Mitarbeiterin am Institut für Kommunikationswissenschaft, Fachbereich Erziehungswissenschaft und Sozialwissenschaften, Westfälische Wilhelms-Universität Münster. Forschung zu digitalisierten Öffentlichkeiten, insbesondere mit Blick auf deren räumliche Dimension, im Sonderforschungsbereich 1265 "Re-Figuration von Räumen".

Kontakt:

Daniela Stoltenberg, M.A.

Institut für Kommunikationswissenschaft,
Fachbereich Erziehungswissenschaft und
Sozialwissenschaften, Westfälische Wilhelms-
Universität
Bispinghof 9-14
D-48143 Münster

Tel.: +49 (0)30 838 63424

E-Mail: daniela.stoltenberg@uni-muenster.de

URL: <https://www.uni-muenster.de/Kowi/personen/daniela-stoltenberg.html>

Peter MILTNER, bis 2017 wissenschaftlicher Mitarbeiter am Institut für Publizistik- und Kommunikationswissenschaft, Fachbereich Politik- und Sozialwissenschaften, Freie Universität Berlin. Derzeit wissenschaftlich-technischer Mitarbeiter beim Projektträger Jülich.

Kontakt:

Peter Miltner

Institut für Publizistik- und
Kommunikationswissenschaft, Fachbereich
Politik- und Sozialwissenschaften, Freie
Universität Berlin
Garystraße 55
D-14195 Berlin

E-Mail: peter.miltner@fu-berlin.de

URL: <https://tinyurl.com/y6vmly4c>

Sophia OSTNER, bis 2016 wissenschaftliche Mitarbeiterin am Institut für Publizistik- und Kommunikationswissenschaft, Fachbereich Politik- und Sozialwissenschaften, Freie Universität Berlin. Derzeit wissenschaftliche Mitarbeiterin, Bundestagsbüro Marja-Liisa VÖLLERS MdB.

Kontakt:

Sophia Ostner

Institut für Publizistik- und Kommunikationswissenschaft, Fachbereich Politik- und Sozialwissenschaften, Freie Universität Berlin
Garystraße 55
D-14195 Berlin

E-Mail: sophiaostner@web.de

URL: <http://www.polsoz.fu-berlin.de>

Barbara PFETSCH, Professorin für Kommunikationstheorie und Medienwirkungsforschung am Institut für Publizistik- und Kommunikationswissenschaft, Fachbereich Politik- und Sozialwissenschaften, Freie Universität Berlin. Forschung und Lehre zu den Themen politische Kommunikation im internationalen Vergleich, Inhalte und Wirkungen der Onlinekommunikation, Europäische Öffentlichkeit.

Kontakt:

Univ.-Prof. Dr. Barbara Pfetsch

Institut für Publizistik- und Kommunikationswissenschaft, Fachbereich Politik- und Sozialwissenschaften, Freie Universität Berlin
Garystraße 55
D-14195 Berlin

Tel.: +49 (0)30 838 57530

E-Mail: pfetsch@zedat.fu-berlin.de

URL: <https://www.polsoz.fu-berlin.de/kommwiss/arbeitsstellen/kommunikationstheorie/mitarbeiterinnen/bpfetsch/>

Zitation

Waldherr, Annie; Wehden, Lars-Ole; Stoltenberg, Daniela; Miltner, Peter; Ostner, Sophia & Pfetsch, Barbara (2019). Induktive Kategorienbildung in der Inhaltsanalyse: Kombination automatischer und manueller Verfahren [83 Absätze]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 20(1), Art. 19, <http://dx.doi.org/10.17169/fqs-20.1.3058>.