

## Automatisierungspotenziale in der qualitativen Diskursanalyse. Das Prinzip des „Filterns“

**Koch, Gertraud**

gertraud.koch@uni-hamburg.de  
Universität Hamburg, Deutschland

**Franken, Lina**

lina.franken@uni-hamburg.de  
Universität Hamburg, Deutschland

Diskursanalytische Verfahren werden in vielen Disziplinen verwendet, so dass die Frage nach Automatisierungspotentialen in diesem Bereich für ganz unterschiedliche Geisteswissenschaften und auch qualitativ arbeitende Sozialwissenschaften relevant ist (vgl. grundlegend Foucault 1971 und 1973; Keller 2011). Die Frage wird aktuell in einem Teilprojekt des Verbundforschungsprojektes hermA (Gaidys et al. 2017) erforscht, aus dem dieser Beitrag hervorgeht.<sup>1</sup> Angesichts steigender digital vorliegender Textmengen, die nicht zuletzt über das Internet verfügbar sind, stellt sich die Frage nach Automatisierungspotentialen immer drängender. Die Perspektive der wissenssoziologischen bzw. -anthropologischen Diskursanalyse, die hier eingenommen wird, kann insofern als exemplarisch angesehen werden, wird aber je nach Schwerpunktsetzung auch Variationen aufweisen.

In der Soziologie und Wissensanthropologie wird die Diskursanalyse als Methode eingesetzt, um zu analysieren, wie sich (neue) gesellschaftliche Phänomene herausbilden und etablieren können, beispielsweise für ein vergleichsweise junges Phänomen wie die Telemedizin. Diese wird zunehmend als Lösung für den Ärztemangel in ländlichen Räumen aber auch zur Betreuung von chronisch kranken Patienten diskutiert und zwar in ganz unterschiedlichen Diskursarenen und durch verschiedene Akteure, vom Bundestag über Ärzte, Krankenkassen bis hin zu Patientenverbänden und den Patienten selbst. Mittels der Diskursanalyse kann herausgearbeitet werden, wie angesichts heterogener Interessenslagen der beteiligten Gruppen die verschiedenen Auffassungen von Telemedizin sowie ihrer Notwendigkeit verhandelt werden, ob sich ggf. ein allgemeines gesellschaftlich weitgehend akzeptiertes Verständnis von Telemedizin herausbildet, sich schließlich konkrete Arbeitsweisen (Praktiken), institutionelle Zusammenhänge oder Organisationsformen sowie gesetzliche Regelungen verfestigen (materialisieren).

**Spezifik der diskursanalytischen Datengrundlage**

Für Diskursanalysen wird heterogenes Quellenmaterial verwendet, es können ganz unterschiedliche Textsorten eine Rolle spielen. In der dem Beitrag zugrunde liegenden Forschung zu Akzeptanzproblematiken der Telemedizin werden Webseiten (Homepages, Blogs, Foren, etc.), wissenschaftliche Beiträge und Bundestagsprotokolle analysiert. Potentiell können auch multimodale Daten wie Bilder oder audio-visuelles Material hinzugezogen werden. Zunächst erfolgt jedoch eine Beschränkung auf Textquellen, um die Datenmodellierung mit Methoden der Digital Humanities gezielt verfolgen zu können.

Die Korpuserstellung erfolgt für Diskursanalysen iterativ nach dem in der Grounded Theory (GT) formulierten Prinzip des Theoretischen Sampling, mit dem eine tendenziell unüberschaubare Datenmenge epistemologisch geleitet reduziert wird. Dieses Verfahren zeichnet sich dadurch aus, dass in mehreren zyklischen Prozessen Daten erhoben, annotiert und interpretiert werden. Dabei wird nach jedem Zyklus anhand der vorliegenden Ergebnisse über die weitere Datenerhebung entschieden (vgl. Glaser/Strauss 1967). „Manuelle“ Filterprozesse sind somit eine iterative Abfolge verschiedener interpretativer, methodisch geleiteter Arbeitsschritte: theoretisches Sampling, offenes Kodieren, selektives Kodieren, axiales Kodieren, theoriegeleitete Interpretation (Bryant/Charmaz 2007; Götzö 2014). Die Informationsfülle wird so sukzessive in eine für qualitative Forschungen handhabbare Größenordnung gebracht. Im Laufe des Forschungsprozesses entsteht so in interaktiven Prozessen der Datenerhebung und -interpretation ein relativ kleiner Datenkorpus, meist aus unterschiedlichen Quellen, die aussagekräftig für die Fragestellung sind (vgl. Strauss/Corbin 1996). Es stellt sich die Frage, inwieweit diese „manuellen“ Filterprozesse von heterogenen Textsorten durch automatisierte Verfahren im Sinne einer höheren Effizienz oder verbesserten analytischen Qualität ergänzt werden können. Die GT zielt nicht auf Repräsentativität der Daten, sondern auf Viabilität also eine hohe Passung in der Erklärungskraft (Glaserfeld 1997) unabhängig davon ob Daten digital oder analog vorliegen.

### **Filtern als Prinzip methodisch geleiteter Analyse**

Im Sinne der iterativen Korpuserstellung wird untersucht, wie sich manuelles und automatisiertes Filtern miteinander verbinden lassen und was die jeweilige Form des Filterns auszeichnet. „Filtern“ wird dabei als ein Arbeitsschritt in der wissenschaftlichen Analyse von Daten verstanden, welches nach methodischen Prinzipien umgesetzt und zur Reduktion der im Alltag beobachtbaren Komplexität eingesetzt wird. Bei den automatisierten Ansätzen werden neben Information Retrieval (vgl. Klink 2017; Manning et al. 2009) auch strukturelle Ansätze des Data und Text Mining für die Reduktion verfügbarer Informationen eingesetzt. Manuell wird entsprechend der Grounded Theory nach dem Prinzip des Theoretical Sampling gearbeitet, um wesentliche, also soziale Wirklichkeit setzende Dokumente bzw. Textabschnitte für die Analyse zu identifizieren. Während es bei automatisierten Ansätzen in der Tendenz darum geht, einen möglichst vollständigen

Korpus aller relevanten Dokumente zu generieren, der wiederum eine Basis für weitere automatische Filteransätze bietet, erfolgt die manuelle Korpuserstellung hochgradig selektiv nach Relevanz bzw. Viabilität.

#### Filtern manuell

Beim manuellen Filtern in der qualitativen Diskursanalyse ist ein hohes Maß an Vorwissen notwendig, welches sich auf mögliche Akteure, Diskursarenen und Kontexte des Themas, hier der Telemedizin bzw. der damit verbundenen Akzeptanzproblematiken, bezieht. Auf der Basis dieses Vorwissens wird der Einstieg in die Frage möglich, wo überhaupt Quellen für die Analyse des Phänomens zu finden sind. Dabei wird heute nicht nur „manuell“ gefiltert, sondern die Hilfe von Suchmaschinen im Internet oder Suchabfragen von Stichwörtern in Archivkatalogen in Anspruch genommen. Allerdings sind damit die Informationen weiterhin tendenziell unüberschaubar und auch hinsichtlich ihrer Relevanz höchst heterogen. Ebenso ist offen, inwieweit tatsächlich alle relevanten Akteure und Diskursarenen erfasst worden sind, so dass eine Vielzahl an weiteren manuellen Filterprozessen vorgenommen werden müssen. Dabei wird das Wissen des Forschenden zum Thema stets erweitert, so dass die (stetig wachsende) Expertise der Forschenden in dem Themenfeld eine wesentliche Voraussetzung für eine hohe analytische Qualität der Diskursanalyse darstellt. Die in der Grounded Theory angelegten methodischen Arbeitsschritte profitieren wesentlich von dieser stetig wachsenden Expertise, sind dabei jedoch auch zur Objektivierung der vom Forschenden formulierten Hypothesen, im Sinne von Falsifizierungen oder Bestätigungen, unerlässlich (vgl. Glaser 1978).

#### Filtern maschinell

Maschinelles Filtern beruht auf strukturellen Analysen von Sprache und bedarf vielfältiger Ressourcen (Ontologien, Wörterbücher, Tools, Korpora). Ein rein automatisiertes Filtern zum Thema Telemedizin ist aufgrund fehlender Ressourcen auf dem aktuellen Wissensstand nicht möglich. Insgesamt darf man davon ausgehen, dass dies meist der Fall ist, wenn neue gesellschaftliche Phänomene auftreten und sich die soziale Wirklichkeit, die Themen und die Sprache wandeln. Es geht insofern darum auszuloten, unter welchen Umständen und wo im Forschungsprozess automatisierte Filterprozesse, zielführend aufgegriffen werden können.

Entsprechend dieser Hypothese wurden bisher unterschiedliche halbautomatisierte Verfahren für die Unterstützung der qualitativen Diskursanalyse mit ihren verschiedenen Arbeitsschritten erprobt: Suchmaschinen, Suchfunktionen, verschiedene Webcrawler, sowie manuelle und automatisierte Annotationen mit proprietären und open-source Tools. Auch die Vorbereitung automatisierter Filterprozesse spielt für die Verfahren eine wichtige Rolle, insbesondere a) die Erstellung von Wortfeldern zur Spezifizierung der Filter-Anwendung, b) die Aufbereitung von Dokumenten für die automatisierte Analyse, c) die Klärung von Arbeitsweisen verschiedener Crawler – etwa iCrawl<sup>2</sup>,

Apify<sup>3</sup> und IssueCrawler<sup>4</sup> oder das auf Crawlern basierende Webarchiv der Deutschen Nationalbibliothek – und wie diese für eine zielführende Diskursanalyse aufgesetzt werden können, d) die Klärung der Datenlage in bestehenden Korpora wie dem Dokumentations- und Informationssystem des Deutschen Bundestages<sup>5</sup>, den alternativ angebotenen Open Data-Beständen<sup>6</sup> sowie dem GermaParlTEI-Korpus<sup>7</sup> und der Aufarbeitung von Crawling-Ergebnissen. Für eine Vorbereitung der Auswertung erfolgt zudem e) die Erprobung und der Vergleich verschiedener Annotationstools von proprietären Programmen der qualitativen Datenanalyse, welche eine Computerunterstützung für Geistes- und Sozialwissenschaftler\*innen zugänglich machen und bereits weit verbreitet sind (vgl. Gasteiger/Schneider 2014; Sattler 2014) und open-source Optionen wie etwa CATMA<sup>8</sup> sowie f) Koreferenzannotationen mittels CoRefAnnotator<sup>9</sup>, um Analysen zur Konkordanz und Netzwerkanalysen vorzubereiten. Dabei stehen die spezifischen Potentiale und Probleme für die Korpuserstellung und -auswertung bei diskursanalytischem Datenmaterial im Zentrum.

#### Filtern als Automatisierungspotential in der Diskursanalyse

Für einen lösungsorientierten Ansatz zur Anreicherung von qualitativen hermeneutischen Verfahren der Diskursanalyse mit den strukturell arbeitenden Ansätzen von Methoden der Digital Humanities hat sich bisher gezeigt, dass es vor allem einfache Verfahren des Information Retrieval sind, die unterstützend für die qualitative Forschung nach GT wirken und auch gegenwärtig eingesetzt werden, allerdings vor allem auf der Ebene generischer Tools (Suchfunktionen, automatisiertes Abrufen und Speichern von Dokumenten, Wordfelder). In dem Moment, in dem Anpassungen von Tools notwendig werden, relativieren sich die Vorteile automatisierter Verfahren, insbesondere weil sich die scheinbare Fülle der Informationen im Sinne sogenannter „big data“ bei näherem Ansehen der Daten bisher nicht erfüllt hat, diese aufgrund weniger valider bzw. viabler Textstellen rasch zu „small data“ werden, die wiederum leichter qualitativen Analysen zugänglich sind. Dies hängt wohl zentral mit der Neuheit des Phänomens Telemedizin zusammen, was für Untersuchungsgegenstände der Kulturanthropologie als typisch angesehen werden kann. Gleiches gilt für die Heterogenität unterschiedlicher Textsorten und entsprechend die Herausforderungen automatischer Aufbereitungsschritte. Der gegenwärtige Lerneffekt im Projekt bezieht sich so insbesondere auf die Spezifizierung, wo und wie automatisierte Verfahren in der Diskursanalyse überhaupt sinnvoll eingebettet werden können, sowie auf die Erfahrungen hinsichtlich der Erprobung verschiedener Ansätze des Filterns inklusive der dafür notwendigen Aufbereitungen des generierten Materials und der iterativ ineinander greifenden Schritte der automatischen und manuellen Filterung.

## Fußnoten

1. Der Forschungsverbund „Automatisierte Modellierung hermeneutischer Prozesse – Der Einsatz von Annotationen für sozial- und geisteswissenschaftliche Analysen im Gesundheitsbereich (hermA)“ ist ein interdisziplinäres Projekt an der Universität Hamburg, der Technischen Universität Hamburg und der Hochschule für Angewandte Wissenschaften Hamburg, das durch die Landesforschungsförderung Hamburg finanziert wird. Das Teilprojekt „Automatisierungspotenziale hermeneutischer Prozesse in der Diskursethnographie zu Akzeptanzproblematiken der Telemedizin“ ist ein Beitrag aus der Kulturanthropologie. Vgl. <https://www.herma.uni-hamburg.de/>.
2. <http://icrawl.l3s.uni-hannover.de/>
3. <https://www.apify.com/page-analyzer>
4. <https://www.issuecrawler.net/>
5. <http://dipbt.bundestag.de/dip21.web/bt>
6. <https://www.bundestag.de/service/opendata>
7. <https://github.com/PolMine/GermaParlTEI>
8. <http://catma.de/>
9. <https://github.com/nilsreiter/CorefAnnotator/releases>

## Bibliographie

- Bryant, Antony; Charmaz, Kathy (Hg.):** *The SAGE handbook of Grounded Theory*, Los Angeles 2007.
- Foucault, Michel:** *Archäologie des Wissens*, Frankfurt a.M. 1973.
- Foucault, Michel (1971):** *Die Ordnung des Diskurses*, Frankfurt a.M. 1991.
- Keller, Reiner (2005):** *Wissenssoziologische Diskursanalyse. Grundlegung eines Forschungsprogramms*, 3. Auflage Wiesbaden 2011.
- Gaidys, Uta / Gius, Evelyn / Jarchow, Margarete / Koch, Gertraud / Menzel, Wolfgang / Orth, Dominik / Zinsmeister, Heike:** *Project Description. HerMA: Automated Modelling of Hermeneutic Processes*, in: *Hamburger Journal für Kulturanthropologie* 7 (2017), S. 119–123.
- Glaser, Barney G.:** *Theoretical Sensitivity. Advances in the Methodology of Grounded Theory*, Mill Valley, Calif. 1978.
- Gasteiger, Ludwig / Schneider, Werner:** *Diskursanalyse und die Verwendung von CAQDA-Software*, in: **Angermüller, Johannes / Nonhoff, Martin / Herschinger, Eva / Macgilchrist, Felicitas / Reisigl, Martin / Wedl, Juliette / Wrana, Daniel / Ziem, Alexander (Hg.):** *Diskursforschung. Ein interdisziplinäres Handbuch. Band 2: Methoden und Analysepraxis, Perspektiven auf Hochschulreformdiskurse*, Bielefeld 2014, S. 852–872.
- Glaser, Barney G. / Strauss, Anselm L. (1967):** *Grounded Theory. Strategien qualitativer Forschung*, Bern 2010.

**Glaserfeld, Ernst von:** *Radikaler Konstruktivismus* Frankfurt 1997.

**Götzö, Monika:** *Theoriebildung nach Grounded Theory*, in: **Bischoff, Christine / Oehme-Jüngling, Karoline / Leimgruber, Walter (Hg.):** *Methoden der Kulturanthropologie*, Bern 2014, S. 444–458.

**Keller, Reiner (2005):** *Wissenssoziologische Diskursanalyse. Grundlegung eines Forschungsprogramms*, 3. Auflage Wiesbaden 2011.

**Klinke, Harald:** *Information Retrieval*, in: **Jannidis, Fotis / Kohle, Hubertus / Rehbein, Malte (Hg.):** *Digital Humanities. Eine Einführung*, Stuttgart 2017, S. 268–278.

**Manning, Christopher D. / Raghavan, Prabhakar / Schütze, Hinrich:** *Introduction to Information Retrieval* Cambridge 2009.

**Sattler, Simone:** *Computergestützte qualitative Datenbearbeitung*, in: **Bischoff, Christine / Oehme-Jüngling, Karoline / Leimgruber, Walter (Hg.):** *Methoden der Kulturanthropologie*, Bern 2014, S. 476–487.