

Zeitschrift für digitale Geisteswissenschaften

Beitrag aus:

Zeitschrift für digitale Geisteswissenschaften

Titel:

Mensch und Maschine als Team. Exploratives Topic Modeling und manuelle Annotation in der qualitativen Sozialforschung

Autor*in:

Lina Franken

Kontakt: lina.franken@uni-vechta.de

Institution: Universität Vechta

GND: 136526705 ORCID: 0000-0002-2587-4068

Contribution (CRediT): [Conceptualization](#) | [Formal analysis](#) | [Funding acquisition](#) | [Investigation](#) | [Methodology](#) | [Project administration](#) | [Supervision](#) | [Validation](#) | [Visualization](#) | [Writing – original draft](#) | [Writing – review & editing](#)

Autor*in:

Dennis Möbus

Kontakt: dennis.moebus@fernuni-hagen.de

Institution: FernUniversität in Hagen

GND: 1238567746 ORCID: 0009-0008-9064-7460

Contribution (CRediT): [Conceptualization](#) | [Data curation](#) | [Formal analysis](#) | [Funding acquisition](#) | [Investigation](#) | [Methodology](#) | [Project administration](#) | [Resources](#) | [Software](#) | [Supervision](#) | [Validation](#) | [Visualization](#) | [Writing – original draft](#) | [Writing – review & editing](#)

Weitere Beteiligte:

Florian Schmid ([Conceptualization](#) | [Investigation](#) | [Methodology](#) | [Project administration](#) | [Visualization](#)), Nils Egger ([Conceptualization](#) | [Investigation](#) | [Methodology](#) | [Project administration](#) | [Visualization](#))

DOI des Artikels:

[10.17175/2024_003](https://doi.org/10.17175/2024_003)


Nachweis im OPAC der Herzog August Bibliothek:

[1889562653](#)

Erstveröffentlichung:

23.05.2024

Lizenz:

Sofern nicht anders angegeben 

Medienlizenzen:

Medienrechte liegen bei den Autor*innen

Letzte Überprüfung aller Verweise:

08.05.2024

Format:

PDF ohne Paginierung, Lesefassung

GND-Verschlagwortung:

[Oral history](#) | [Qualitative Methode](#) | [Digitale Hermeneutik](#) | [Annotation](#) | [Empirische Kulturwissenschaft](#) | [Topic Modeling](#) |

Empfohlene Zitierweise:

Lina Franken / Dennis Möbus: Mensch und Maschine als Team. Exploratives Topic Modeling und manuelle Annotation in der qualitativen Sozialforschung. In: Zeitschrift für digitale Geisteswissenschaften 9 (2024). 23.05.2024. HTML / XML / PDF. DOI: [10.17175/2024_003](https://doi.org/10.17175/2024_003)

Lina Franken, Dennis Möbus

Mensch und Maschine als Team. Exploratives Topic Modeling und manuelle Annotation in der qualitativen Sozialforschung

Abstracts

In einem experimentellen Setting wird das Zusammenspiel von *Topic Modeling* und Annotation nach *Grounded Theory* bei der Erschließung umfangreicher qualitativer Forschungsdaten mit einer Kombination aus quantitativer Vorstrukturierung und qualitativer Feinanalyse untersucht. So werden Fragen der Epistemologie, Sinnstiftung, Eigenheiten, Kompatibilität und unterschiedlicher Dynamiken beider Methoden adressiert. *Mixed-Methods*-Ansätze bieten sich an, wobei Topic-Modeling-Angebote unterbreitet und anschließend von Forschenden interpretiert und zur Strukturierung qualitativer Analysen genutzt werden. Zu reflektieren ist, dass Forschende hierbei Handlungsmacht abgeben. Die Verbindung der Perspektiven und ihre ethnografische Begleitung setzt auch Forderungen nach Evaluation im *Machine Learning* um.

In an experimental setting, the interplay of *topic modeling* and *grounded theory annotation* in the exploitation of large-scale collections of qualitative research data is approached via a combination of quantitative pre-structuring and qualitative in-depth analysis. This raises issues of epistemology, sense-making, characteristics, compatibility, and different dynamics of both methods. A *mixed-methods* approach suggests that topic modeling provides a proposition that can be interpreted by researchers for a first structuring of a qualitative analysis. Nevertheless, researchers give away agency through the use of topic modeling. The combination of perspectives and their ethnographic reflection can also meet the demand for evaluation in *machine learning*.

1. Einleitung: manuelle und maschinelle Annotation anhand von Oral-History-Interviews vergleichend untersuchen

Mit der digitalen Verfügbarkeit von *Oral-History*-Interviews wird es möglich, diese mit maschinellen Annotationen zu bearbeiten, also computationale Methoden anzuwenden, um große Mengen der vorliegenden Daten zu analysieren.¹ Bisher wurden vorrangig menschliche Annotationen zur qualitativen Interpretation einzelner Fragestellungen für diese Quellen genutzt. Durch den sich wandelnden Zugang ergeben sich neue Herausforderungen, und es stellt sich die Frage, wie beide Verfahren miteinander verbunden werden können. Außerdem fehlen systematische Überprüfungen der Veränderungen im Interpretationsprozess. Deshalb bietet es sich an, einen experimentellen Ansatz zu verfolgen, der die konkreten Annotations- und Interpretationsschritte ethnografisch dokumentiert.

Der vorliegende Beitrag bearbeitet die leitende Fragestellung, ob maschinelle, computationale Verfahren (und hier insbesondere das *Topic Modeling*) qualitative, manuelle Analysen (hier insbesondere aus Perspektive der *Grounded Theory*²) bereichern können. Dazu untersuchen wir die Unterschiede, die entstehen, wenn Interviewtranskripte durch Topic Modeling strukturiert gesichtet oder durch händische, offene Annotation bearbeitet werden. Daraus lässt sich ableiten, wie sich Perspektiven auf (unbekannte) Texte mit manueller und maschineller Bearbeitung unterscheiden. Mit unserem Zugang möchten wir dazu beitragen, die Mehrwerte und Anschlussmöglichkeiten von computationalen Methoden im Allgemeinen und Topic Modeling im Speziellen für die Analyse von Oral-History-Daten und anderen Interviewtranskripten herauszuarbeiten.

Um das zu realisieren, haben die Autor*innen in einem Workshop mit insgesamt fünfzehn Teilnehmenden und zwei weiteren Leitenden die oben genannten Methoden systematisch verglichen. Auf Grundlage der in diesem Setting produzierten Daten wird im Folgenden exemplarisch aufgezeigt, wie menschliche und maschinelle Annotation sich unterscheiden, wo sie Interpretationsvorgänge offenlegen, wo Gemeinsamkeiten bestehen und wie existierende Ansätze durch Synergieeffekte erweitert werden können. Das führt auch zu Fragen nach Formen von Sinnstiftung, die mit computationalen Verfahren erreicht werden können.

¹ Vgl. Hodel et al. 2022. Das DFG-Projekt *Oral-History.Digital (OHD)* ist die derzeit größte Initiative zur Digitalisierung und Verfügbarmachung von Oral-History-Quellen. Das *Archiv »Deutsches Gedächtnis« des Instituts für Geschichte und Biographie* baut im Rahmen von OHD ein Pilotarchiv auf.

² Vgl. Strübing 2014; Charmaz 2014.

Oral History ist seit jeher technologieaffin und schon lange mit den Folgen der Digitalisierung konfrontiert, etwa durch Textverarbeitungssoftware oder Neuerungen hin zu digitalen Aufnahmetechniken.³ Mit den großen und vielschichtigen Beständen, die digital vorliegen, ist die Oral History ebenso wie die qualitative Sozialforschung in das Zeitalter von *Big Data* eingetreten. Durch neue Zugänge eröffnen sich eventuell völlig neue Möglichkeiten, ganze Interview-Korpora in der Tiefe zu durchdringen, insbesondere für Sekundäranalysen.⁴ Computationale Verfahren können in diesem Bereich einige Unterstützung bieten,⁵ denn in der Regel liegen die langen und komplexen Transkripte unstrukturiert vor. Aus der qualitativen Sekundäranalyse bestehen bisher nur vereinzelt computationale Annäherungen mit in der Regel einfachen Auszählungsverfahren.⁶ Auch wenn Louise Corti bereits 2007 Verfahren des *Natural Language Processing (NLP)* getestet hat, um aus den Transkripten Metadaten zu generieren⁷, ist seither wenig Forschung in diese Richtung realisiert worden. Eine Ausnahme bilden beispielsweise die Studien von Laura Nelson, die verschiedene computationale Verfahren, unter anderem Topic Modeling, nutzt, um soziologische Fragen zu beantworten – allerdings ohne manuelle Annotation.⁸ Jüngst hat ein Team um Nelson ebenfalls einen Vergleich von qualitativen Annotationen und solche durch Topic Modeling vorgelegt, wobei hier schriftliche Antworten auf eine offene Frage ausgewertet wurden. Die Texte wurden zunächst in Anlehnung an inhaltsanalytische Verfahren durch mehrere Personen annotiert, um anschließend ein Modell aus dem Material zu erstellen. Beide Annotationen wurden dann durch quantitative Metriken miteinander verglichen.⁹ Die Ergebnisse sind zwar relevant, das Vorgehen weicht von unserem allerdings deutlich ab und lässt qualitative Aspekte der Evaluation außen vor. Aus der qualitativen Inhaltsanalyse liegt ein Einführungswerk vor, das wichtige Impulse für die qualitative Forschung setzt, sich jedoch mit dem Verhältnis von manuellen zu computationalen Annotationen an keiner Stelle befasst.¹⁰

Aus dem Bereich des NLP hat sich dabei das Topic Modeling als eine besonders zielführende Heuristik zur Erschließung großer, unbekannter Textkorpora etabliert.¹¹ Topic Modeling ist ein Verfahren des *Machine Learnings*, das durch die Berechnung von Wahrscheinlichkeiten die in Texten vorkommenden Wörter (*Token*) nach häufigem gemeinsamen Vorkommen sortiert. Auf dieser Grundlage werden die Topics als Wortlisten ausgegeben, die, so die Grundannahme, wegen ihrer Nähe zueinander einen thematischen Zusammenhang bilden.¹² Gut trainierte Modelle ermöglichen die Extraktion thematischer Zusammenhänge auch aus lebensgeschichtlichen Interviews, wodurch erste inhaltliche Eindrücke generiert werden können.¹³ Bei der freien Exploration der Ergebnisse vermag das Topic Modeling Phänomene sichtbar zu machen, die nach der initialen qualitativen Auswertung im Rahmen eines thematisch fokussierten Forschungsprojekts oft unter der Oberfläche bleiben.¹⁴

Die hier vorgestellte Studie hat das Ziel, manuelle und computationally generierte Annotationen miteinander zu vergleichen. Unter Annotationen verstehen wir dabei die Auszeichnung von Text zur Organisation von Information und zur Dokumentation von Interpretationsprozessen, sowohl während als auch im Nachgang der Interpretation.¹⁵ Dies kann sowohl durch Menschen als auch durch Maschinen geschehen. Mit unserer Studie adressieren wir damit zwei Desiderata: Einerseits fehlen in den Digital Humanities weitestgehend systematische Studien, die computationale und menschliche Inhaltserschließung vergleichen.¹⁶ Andererseits mangelt es häufig an qualitativen Evaluationsroutinen im Machine Learning, um die automatisch generierten Ergebnisse in ihrer Qualität zu validieren.¹⁷ Das ist insbesondere wichtig, weil die Hemmschwelle zur Anwendung digitaler Methoden bei qualitativ Forschenden weiterhin hoch ist und mit solchen Qualitätssicherungen gesenkt werden kann.¹⁸

Im Folgenden stellen wir zunächst das Vorgehen unseres Workshops vor, bevor wir die Ergebnisse der manuellen und der maschinellen Annotation nacheinander darstellen. Im Anschluss werden diese systematisierend verglichen. Der Text endet mit Überlegungen zu Limitationen des experimentellen Ansatzes und einem Fazit zum Vergleich maschineller und manueller Annotation.

³ Vgl. Leh 2020, S. 4, 16–17.

⁴ Vgl. Apel 2015; Hollstein / Strübing 2018; Dunkel et al. (Hg.) 2019; Franken et al. 2023.

⁵ Vgl. Sloan 2014; Cohen 2013.

⁶ Vgl. Dargentas / Le Roux 2005.

⁷ Vgl. Corti 2007.

⁸ Nelson 2020, vgl. für den Workflow S. 14.

⁹ Miner et al. 2023.

¹⁰ Papilloud / Hinneburg 2018.

¹¹ Vgl. Graham et al. 2016; Lemke / Wiedemann (Hg.) 2016.

¹² Vgl. Franken 2023, S. 166–172.

¹³ Vgl. Hodel et al. 2022.

¹⁴ Vgl. Möbus 2020.

¹⁵ Vgl. Rapp 2017, S. 254; Franken et al. 2020.

¹⁶ Vgl. erste Ansätze bei Andorfer 2017; Fechner / Weiß 2017; Andresen et al. 2020.

¹⁷ Vgl. Dobson 2021; DiMaggio 2015.

¹⁸ Vgl. Franken 2020.

2. Methodisches Vorgehen: vergleichender Workshop mit ethnografischer Dokumentation

Die aufgeworfenen Fragestellungen haben wir in einem experimentellen Workshop untersucht, dessen Geschehen wir mit insgesamt vier Workshopleitenden ethnografisch dokumentierten. Ziel des Workshops war es, konkret am Textkorpus herauszuarbeiten, wie und wo sich die Perspektive des Zugangs unterscheidet, je nachdem ob sich den Quellen zunächst durch Topic Modeling oder durch manuelle Annotation genähert wird. Zudem sollten diese Perspektiven in der Gruppe reflektiert werden. Dies diente der explorativen, aber methodisch kontrollierten Evaluation der beiden Zugänge, die Grundlage für weitere Systematisierungen und Methodenkombinationen ist. Dementsprechend war die Fragestellung des Workshops eng verwandt, aber nicht identisch mit der hier verfolgten Fragestellung. Der Workshop wurde als Kooperation der Computational Social Sciences der LMU München (Lina Franken) und der Forschungsgruppe »digital humanities« der FernUniversität in Hagen (Dennis Möbus) durchgeführt. Um eine heterogene Zusammensetzung von Forschenden mit Hintergrund in der qualitativen Forschung / Oral History und Digital Humanities zu erreichen sowie einen Transfer zwischen Forschung und Lehre herzustellen, wurden gezielt Studierende der LMU und der FernUni aus diesen Bereichen, Mitarbeitende der ausrichtenden Institutionen und Partner*innen aus den eigenen Forschungsnetzwerken eingeladen, sich mit einem kurzen Motivationsschreiben zu bewerben.

Die insgesamt fünfzehn Teilnehmenden wurden in zwei Gruppen aufgeteilt, von denen eine zunächst mit auf Topic Modeling basierenden Annotationen (Gruppe A), die andere zunächst mit manueller Annotation (Gruppe B) das Korpus erkundete. Die beiden Gruppen wurden im Sinne einer guten Betreuung und Dokumentation noch einmal in zwei Teilgruppen geteilt. Die vier Kleingruppen waren dabei bewusst heterogen zusammengesetzt. Auf Grundlage eines mit der Anmeldung eingereichten Motivationsschreibens konnten Kompetenzen und Vorwissen eingeschätzt werden, sodass beispielsweise Teilnehmende mit Erfahrung in qualitativer Forschung und Teilnehmende, die bereits mit Topic Modeling vertraut waren, zusammenarbeiteten. Zudem wurde auf ein ausgewogenes Geschlechterverhältnis so gut wie möglich geachtet, wobei sich mehr Männer als Frauen für den Workshop angemeldet hatten. In jeder Gruppe waren Personen unterschiedlicher akademischer Statusgruppen vertreten, insbesondere Studierende, Promovierende und PostDocs.

In Gruppe A1 trafen eine promovierte Europäische Ethnologin mit Forschungsschwerpunkt Arbeitskulturen, ein frischer PostDoc aus dem Bereich Digital Humanities, ein nicht promovierter Historiker und Informatiker mit langjähriger Berufserfahrung im Mittelbau zeitgeschichtlicher Forschung und ein gerade startender Doktorand der Soziologie aufeinander. Gruppe A2 setzte sich zusammen aus einem Studenten der Kulturwissenschaften im Bachelor, einer Studentin der Soziologie im Bachelor kurz vor Einreichung der BA-Arbeit, einem Doktoranden der Europäischen Ethnologie mit Forschungsschwerpunkt in digitalen Arbeitskulturen und einer Master-Studentin der Geschichte mit Berufserfahrung in der Informatik. In Gruppe B1 arbeiteten ein Student der Kulturwissenschaften im Bachelor mit langjähriger Berufserfahrung als Programmierer, ein Doktorand der Europäischen Ethnologie mit Forschungsschwerpunkt in der Montanindustrie, ein Bachelor-Student der Soziologie und eine Doktorandin der Europäischen Ethnologie mit Forschungsschwerpunkt in digitalen Alltagskulturen zusammen. In Gruppe B2 schließlich trafen eine Doktorandin der Europäischen Ethnologie und Studentin der Informatik, sowie eine Studentin und ein Student der Kulturwissenschaften im Bachelor aufeinander, die vierte Person war krankheitsbedingt kurzfristig ausgefallen.

Die Gruppen arbeiteten mit dem digital bereits erschlossenen Korpus lebensgeschichtlicher Interviews aus dem Bestand [Lebensgeschichte und Sozialkultur im Ruhrgebiet \(LUSIR\)](#), einer wichtigen Quellensammlung zur Sozialgeschichte des 20. Jahrhunderts. Von den insgesamt mehr als 300 geführten Interviews dieses Oral-History-Projektes liegen über 150 bereits akkurat transkribiert und mit Timecodes versehen digital vor. Das Material entstand zwischen 1981 und 1988 in einem Projekt von Lutz Niethammer und Alexander von Plato, in dem Faschismus- und Nachkriegserfahrungen im Ruhrgebiet erforscht wurden.¹⁹

Für das hier beschriebene Experiment wurden die schriftlichen Transkripte verwendet, um mit Methoden des NLP vorgehen zu können, insgesamt 166 Volltexte mit einem Umfang von rund 3,7 Millionen Token (Wörtern). Nach einer Stoppwortbereinigung mit einer kuratierten Stoppwortliste blieben gut 700.000 Token. Das Topic Modeling wurde mit dem Java-Paket **MALLET** durchgeführt und mit Hilfe des Wrappers der Python-Bibliothek **Gensim** umgesetzt. Die MALLET-Implementierung des *Latent-Dirichlet-Allocation*-Verfahrens (*LDA*), die auf das rechenintensive *Gibbs-Sampling* als Inferenzalgorithmus setzt, lieferte in einer vorausgehenden Studie wesentlich konsistentere Topics als die Gensim-Implementierung. Die Studie legte ebenfalls die Aufteilung der Interviews in Chunks zu 25 Sätzen und eine Anzahl von 50 Topics als Grundlage für ein optimales Modell nahe.²⁰

¹⁹ Vgl. Niethammer 1983.

²⁰ Vgl. Hodel et al. 2022, S. 198–199.

Die Realisierung des Topic Modelings im Workshop erfolgte in einem vorab erstellen, auf [Google Colab](#) gehosteten Jupyter Notebook mit anonymisierten Daten.²¹ Im Notebook war das in vorhergehenden Studien als aussagekräftig identifizierte Modell bereits geladen. Die qualitativen Annotationen der Interviewtranskripte erfolgten mit im Vorfeld vorbereiteten Projekten in [CATMA](#).²² Für jede Gruppe waren Textstellen sowie Tagsets bereits im Tool angelegt. Ebenfalls wurde für alle Gruppen ein Padlet zur schriftlichen Dokumentation und Diskussion eingerichtet.

Der Workshop folgte einer festgelegten Strukturierung mit Aufgaben, die den Teilnehmenden gestellt und alleine bzw. in der Kleingruppe bearbeitet wurden. Alle im Folgenden verwendeten direkten Zitate der Teilnehmenden stammen aus den Padlets oder den Mitschriften. Die Gruppen A1 und A2 starteten mit der Erkundung der noch unbekanntem Texte im Jupyter Notebook, B1 und B2 begannen mit der manuellen Annotation in CATMA. Am zweiten Workshopstag wurden die Aufgaben getauscht mit dem Ziel, die Unterschiede im Anschluss systematisch vergleichen zu können. Die Aufgabenfolge war dabei für beide Teilgruppen identisch, egal ob sie die händische Annotation als erste Aufgabe oder erst nach der Beschäftigung mit dem Korpus anhand von Topic Modeling umgesetzt haben. Um vergleichbare Ergebnisse zu erhalten, wurde als inhaltliches Erkenntnisinteresse das Thema »Arbeit« in all seinen Facetten vorgegeben. Das lag auf der Hand, weil die Organisation und Kultur von Arbeit im Ruhrgebiet zentraler Bestandteil der Interviews des LUSIR-Projektes war. Auch wenn dies nicht explizit Thema und Fragestellung der Primärstudie war, sind doch umfangreiche Darstellungen von Arbeitsalltagen in den Materialien enthalten. Dieser Testfall eignet sich somit ideal für die Verbindung von qualitativen und computationellen Methoden und der Erprobung unterschiedlicher *Mixed-Methods*-Ansätze, die wiederum mit ethnografischen Methoden dokumentiert werden, um sie bewerten zu können.

3. Ergebnisse der manuellen Annotation

3.1 Auswahl und Vorbereitung der Quellengrundlage

Für die qualitative Analyse des umfangreichen LUSIR-Quellenmaterials waren im Vorfeld des Workshops drei Interviewpassagen ausgewählt worden, die sich mit unterschiedlichen Schwerpunkten des Themas »Arbeit« beschäftigen. Dies geschah in einem mehrstufigen Auswahlprozess durch die Autor*innen, in dem anhand des bestehenden Modells nach Topics mit Bezug zum Thema »Arbeit« recherchiert und für zwölf interessante Topics einzelne Chunks herausgesucht wurden, die eins der betreffenden Topics in hohem Anteil enthalten. Aus diesen wurden zwei Interviewausschnitte ausgewählt, in denen die unterschiedlichen Dimensionen von Industriearbeit anschaulich und in einer kohärenten Erzählung geschildert werden.²³ Mit dem Vorwissen der Autor*innen zum Korpus wurde dann kontrastiv eine weitere Textstelle herangezogen, in welcher weibliche Büroarbeit thematisiert wird.²⁴ Für das finale Sample wurden bewusst komplette Chunks verwendet, um auch thematisch eher entfernte Narrationen einzubeziehen und die Vielfalt der manuell zu annotierenden Texte zu repräsentieren. Im Ergebnis wurden drei Textstellen von 1.611 Token (aus Interview ADG0054), 3.405 Token (aus Interview ADG0089) und 2.266 Token (aus Interview ADG0042) für die manuelle Analyse ausgewählt, wobei Timecodes und Sprecher*innenkürzel nicht mitgezählt sind.

3.2 Grounded Theory als methodologischer Rahmen der Workshop-Annotationen

Die ausgewählten Textpassagen wurden den Workshop-Gruppen ohne weitere Erläuterung vorgelegt, lediglich Rahmeninformationen zum Korpus insgesamt wurden zu Beginn vorgestellt. Vor dem Start in die eigene manuelle Durchsicht und Annotation erfolgte eine kurze Einführung in die Prinzipien der Grounded Theory²⁵ als Grundlage der weiteren Beschäftigung mit den Textpassagen. In der qualitativen Sozialforschung ist diese Analyseverfahren weit verbreitet: Das induktive Verfahren ermöglicht, aus dem Material heraus Bedeutungen und Kontexte zu erschließen, und ist, besonders in offenen und selektiven Annotationen, Teil des Erkenntnisprozesses.²⁶ Es eignet sich auch für die Erweiterung entsprechender Forschungsprozesse mit computationellen Verfahren.²⁷ Diese Form von Annotationen ist prozessorientiert, sie dient der Unterstützung des Forschungsprozesses und weniger der Dokumentation von diesem.

²¹ Zum Nachvollzug kann ein [Github-Repository mit einem mittels Binder bereitgestelltem Jupyter Notebook](#) genutzt werden. Aus Datenschutzgründen muss der Zugang zu den Daten individuell über das [Archiv »Deutsches Gedächtnis«](#) angefragt werden.

²² Gius et al. 2022.

²³ Interview [ADG0054](#), Chunks 83–90 und Interview [ADG0089](#), Chunks 8–14.

²⁴ Interview [ADG0042](#), Chunks 1–7.

²⁵ Vgl. Glaser / Strauss 2010 [1967]; Charmaz 2014.

²⁶ Vgl. Franken et al. 2020.

²⁷ Vgl. Franken 2022.

Das Vorgehen musste für den Workshop in seiner Komplexität deutlich reduziert werden. Gerade das iterative Vorgehen der fortlaufenden Erweiterung des eigenen Textkorpus durch theoretisches Sampling, auf dessen Grundlage zentrale Quellen aus Korpora ausgewählt werden²⁸, sowie die Verschriftlichung von Gedanken in Memos²⁹ hätten den zeitlichen Rahmen gesprengt.

Zentral für den Workshop war der Schritt des offenen Annotierens, der im Rahmen der Grounded Theory einer ersten Annäherung an das Material dient. Ziel des explorativen Vorgehens im Workshop war nicht eine tiefgehende thematische Analyse des Materials, sondern eine exemplarische Sichtung und Kategorisierung der den Teilnehmenden unbekanntem Texte. Im offenen Annotieren werden Textabschnitte gelesen und aus dem hermeneutischen Sinnerschließen heraus erste thematische Kategorien gebildet. Diese Kategorien werden (oft mittels eines digitalen Annotationstools) direkt an einzelne Textstellen vergeben, sodass das Kategoriensystem bei der Texterschließung prozessual wächst.³⁰ Der Text wird hierdurch in seine verschiedenen Bedeutungen aufgebrochen und diese durch die Kategorienbildung expliziert. Im Falle des Workshops wurde so auch die Grundlage für die später anschließenden Vergleiche der unterschiedlichen methodologischen Zugänge gebildet. In der Grounded Theory würde nach dem offenen Annotieren das axiale Annotieren folgen, in welchem die erarbeiteten Kategorien in Zusammenhänge gebracht werden, um schließlich im selektiven Annotieren die entwickelten unterschiedlichen Perspektiven auf eine zentrale Analyseperspektive zu reduzieren.³¹ Der Schwerpunkt lag zwar auf einer offenen Exploration, wie sie typisch für den Beginn von Forschungsprozessen ist, doch in den Gruppendiskussionen wurden durchaus Kategorien gruppiert, wie es beim axialen Annotieren der Fall ist.

3.3 Ergebnisse der Arbeit in den Gruppen

Im ersten Arbeitsauftrag wurden die Gruppenmitglieder gebeten, die Textstellen einzeln zu lesen und in der Folge offen zu annotieren und dabei im Annotationstool CATMA Kategorien zu erstellen.

Die Gruppen stiegen alle mehr oder weniger unverzüglich in die Erstellung von Kategorien und die konkrete Annotation ein. Diskussionen zur Kategorienvergabe erfolgten dabei meistens anhand der konkreten Textstellen. Alle vier Gruppen diskutierten inhaltlich und versuchten, eine Einigung in Bezug auf die Interpretation von einzelnen Textstellen herzustellen. Dabei gingen sie chronologisch durch den Text, sprangen jedoch bei vergleichbaren Passagen zurück und veränderten Kategorien fortlaufend.

	B1 Manuell (ante TM)	B2 Manuell (ante TM)	A1 Manuell (post TM)	A2 Manuell (post TM)
DISKRIMINIERUNGEN / UNGLEICHBEHANDLUNGEN	Gender	Gender	Gender	Geschlechterrollen
	Intersektionalität	-	-	-
ARBEIT	Karriere	Karriere	Karriere	Werdegang
	Ausbildung / Bildung	Ausbildung	Ausbildung	-
	-	-	-	Zugänge zum Beruf / Betrieb
	-	Arbeit	-	-
	-	-	Arbeitsalltag	Arbeitswelt
	-	Arbeitsprozesse	-	-
	-	Arbeitsethos	-	-
	-	-	-	Werte
	Leistung / Entlohnung	-	-	Geld / Lohn / Finanzielles
	-	Arbeitsunfähigkeit	-	-

Tab. 1: Vergleich der in den Gruppen manuell erstellten Kategorien. Clusterung und Sortierung durch die Autor*innen. Größere thematische Gruppen innerhalb eines Clusters sind grün bzw. orange hinterlegt. Ob diese als synonym zueinander gefasst werden, ist Interpretationssache bzw. nicht eindeutig zu entscheiden. Gleichzeitig ist es zweifelsfrei subjektiv, wie hier Kategorien geclustert werden, was auch aufgrund der Erfahrungen und Kenntnisse der Autor*innen erfolgt ist. Dies war zudem durch das übergreifende thematische Erkenntnisinteresse an Arbeit bedingt, so dass andere Kategorien weniger präzise gefasst und auch in der Folge untersucht wurden.

²⁸ Vgl. Morse 2007.

²⁹ Vgl. Lempert 2007.

³⁰ Vgl. Holton 2007.

³¹ Vgl. Strübing 2014, S. 17-19.

	-	-	-	Ausscheiden aus der Erwerbsarbeit
	-	-	arbeitslos / prekär	-
	-	Arbeitsplatzwechsel	-	-
	-	-	Motivation / Identifikation	-
	-	Sicherheit / Stabilität durch Arbeit	-	-
	-	Macht	-	-
	Hierarchien	-	Hierarchie	-
	Arbeitsbeziehungen	-	-	Arbeitsverhältnisse im Betrieb
	-	-	Arbeitskonflikte	-
	Formen der Kollektivierung	-	-	-
	Professionalisierung	-	-	-
	Verkörperlichung / Körperlichkeit von Arbeit	-	-	-
	-	-	Gesundheit / Körper	-
	Wandel der Arbeitsbedingungen	Arbeitsbedingungen	-	-
	-	-	Kollegen	-
	-	Solidarität	Kollegialität / Solidarität	-
	-	-	Gewerkschaft / Interessensvertretung	-
	-	-	Einstellung	-
	technische Entwicklung	Technologie / technologische Entwicklung	Wandel Arbeitstechnik	Einfluss technischer Fortschritt
	konkrete Unternehmen	-	-	-
WOHNEN	-	Wohnsituation	-	Wohnen / Wohnsituation
	-	-	-	Privatraum / Nicht-Arbeit
SOZIALE HERKUNFT / SOZIALES UMFELD	soziales Feld	-	-	-
	-	-	Soziale Herkunft	-
	-	Soziale Situation	-	-
	Kindheit in Arbeiterfamilie	-	-	-
	Klasse	-	-	-
	Familie / Generation	-	-	-
	-	Familie / Familienleben	-	-
	-	-	Generationalität	-
	-	-	-	Carearbeit
	-	Gemeinschaft	-	-
	-	Spaß / Freude	-	-

Tab. 1: Vergleich der in den Gruppen manuell erstellten Kategorien. Clusterung und Sortierung durch die Autor*innen. Größere thematische Gruppen innerhalb eines Clusters sind grün bzw. orange hinterlegt. Ob diese als synonym zueinander gefasst werden, ist Interpretationssache bzw. nicht eindeutig zu entscheiden. Gleichzeitig ist es zweifelsfrei subjektiv, wie hier Kategorien geclustert werden, was auch aufgrund der Erfahrungen und Kenntnisse der Autor*innen erfolgt ist. Dies war zudem durch das übergreifende thematische Erkenntnisinteresse an Arbeit bedingt, so dass andere Kategorien weniger präzise gefasst und auch in der Folge untersucht wurden.

MIGRATION	Herkunft	-	-	-
	-	-	Migration	-
	-	-	-	Ortswechsel
HISTORISCHE UND GESELLSCHAFTLICHE RAHMENBEDINGUNGEN	Zäsuren	-	-	-
	historischer Hintergrund	Krieg	-	Auswirkungen Krieg
	-	Nachkriegszeit	-	-
	-	Vorkriegszeit	-	-
	Widerstand	-	-	-

Tab. 1: Vergleich der in den Gruppen manuell erstellten Kategorien. Clusterung und Sortierung durch die Autor*innen. Größere thematische Gruppen innerhalb eines Clusters sind grün bzw. orange hinterlegt. Ob diese als synonym zueinander gefasst werden, ist Interpretationssache bzw. nicht eindeutig zu entscheiden. Gleichzeitig ist es zweifelsfrei subjektiv, wie hier Kategorien geclustert werden, was auch aufgrund der Erfahrungen und Kenntnisse der Autor*innen erfolgt ist. Dies war zudem durch das übergreifende thematische Erkenntnisinteresse an Arbeit bedingt, so dass andere Kategorien weniger präzise gefasst und auch in der Folge untersucht wurden.

Insgesamt können wir feststellen, dass die Gruppen – unabhängig davon, ob sie zuerst einen Textzugang via Topic Modeling hatten oder nicht – unterschiedlich annotierten. Die Kategoriensets mit ihren Unter- und Oberkategorien entstanden flexibel und veränderten sich im Zuge der gemeinsamen Arbeit. Damit zeigt sich auch, dass die Bedeutungsdimensionen einzelner Kategorien in manueller Annotation stetig ausgehandelt werden. Die Gruppe A2 annotierte alle Textpassagen vollständig, vergab also an allen Textstellen eine Kategorie. Die anderen Gruppen einigten sich nicht auf einen Anspruch der Vollständigkeit, dementsprechend wurden nur jene Textstellen einer Kategorie zugeordnet, die den Gruppen inhaltlich relevant erschienen.

Tabelle 1 zeigt die in den vier Gruppen entwickelten Kategorien in einer durch die Autor*innen im Nachhinein vorgenommenen thematischen Sortierung, die den thematischen Schwerpunkt auf Arbeit wiederum beibehält und andere Themen eher kursorisch bündelt. Die verschiedenen Bezeichnungen der Kategorien verweisen auf die sehr unterschiedlichen Vorkenntnisse und Schwerpunktsetzungen in der Interpretation durch die einzelnen beteiligten Personen. Die jeweilige Entscheidung für eine Kategorie durch die in sich ebenfalls heterogenen Gruppen ist in Teilen auch auf die Diskussionsfreudigkeit bzw. Zurückhaltung von Einzelnen zurückzuführen: Da Entscheidungen ad hoc getroffen werden mussten, setzten sich Vorschläge von Gruppenmitgliedern mit höherem Redeanteil eher durch und sind teilweise auch Ergebnis der Aushandlung wissenschaftlicher Expertise. Dadurch kam es zu Kategoriensets, die dem Vorwissen der Gesamtgruppe oder Teilen von ihr entsprachen.

Zur Interpretation gewisser Stellen der Interviews wünschten sich die Teilnehmenden mehr Kontext über die zur Verfügung gestellten Ausschnitte hinaus. Dies deutet darauf hin, dass manuelles Annotieren die forschende Neugierde anregt. Die Diskussion in der Gruppe B2 zeigt die verschiedenen Einschätzungen zu Textstellen sehr deutlich, die Kontextwissen erforderlich machen, denn die folgende Textstelle wurde einmal als ›unbezahlte Arbeit / Gender‹ und einmal als ›Familie‹ annotiert:

»INT_MS Was haben Sie denn da so für Arbeiten gemacht.«

»IP_LS Ja, ein bißchen Spülen, Trocken machen. Aber wir haben so zu Hause nicht viel arbeiten müssen, die Mutter war ja den ganzen Tag zu Hause und hat so alles selbst gemacht. Aber, was wir mussten, wir mussten unsere Sachen in Ordnung halten, selbst stopfen und bügeln, ja. Also wir sind nicht verwöhnt worden, wir sind sogar sehr streng erzogen, ja, erzogen worden. Meine Eltern waren sehr streng, aber sehr gerecht, ne.«³²

Dies zeigt wiederum, was für unterschiedliche Zusammenhänge ein Interviewausschnitt evozieren kann und wie abhängig dies vom eigenen Vorwissen ist.

Die Gruppe A1 hingegen war sich in ihren Annotationen häufig einig: So wurden Textteile³³ von allen Einzelpersonen mit Kategorien wie ›Gender‹, ›Ungleichheit Gender‹, ›Frauen in der Arbeitswelt‹, ›Frauen im Job‹, ›Karriere‹, ›Ausbildung‹, ›Ausbildung und Profession‹ verschlagwortet. Als übergeordnete Kategorien einigte sich die Gruppe dann auf ›Ausbildung‹, ›Anstellung Verwaltung‹ und ›Gender‹. Dass diese Konsensbildung bei der entsprechenden Textstelle in den anderen Gruppen nicht so deutlich der Fall war, verweist erneut auf vorhandene oder eben nicht vorhandene geteilte Wissensbestände.

³² ADG0042, Chunk 7.

³³ ADG0042, Chunk 1.

Auffällig ist in der Summe, dass die beiden Gruppen, die sich zunächst über Topic Modeling dem Korpus genähert haben, weniger differenziert annotiert haben und auch weniger Kategorien gebildet haben (17 und 14 in den A-Gruppen zu je 21 in den B-Gruppen). Zudem wurden hier weniger analytische Kategorien gebildet: Die Kategorien der A-Gruppen bleiben eher auf einem beschreibenden Level, auch wenn die Grenzen dabei natürlich wiederum interpretativ gesetzt werden und deshalb hier nicht in absoluten Zahlen ausgedrückt werden.

In der Gruppe B1 wurden hingegen direkt abstrahierende Kategorien diskutiert, nachdem die Gruppenmitglieder die Textpassagen gelesen hatten. Themen wie ›Gender‹ und ›soziale Herkunft‹ wurden explizit genannt und in der Folge auch für die gemeinsamen Kategorien verwendet, wobei hier weitere analytische Kategorien wie ›Intersektionalität‹, ›Professionalisierung‹ oder ›Hierarchien‹ verwendet wurden. Nicht alle Beteiligten der Teilgruppe diskutierten diese Punkte intensiv mit, was wiederum auf die Relevanz der individuellen Vorkenntnisse verweist. Interessanterweise kam die eher ›analytische‹ Kategorie ›Professionalisierung‹ von einer Person, die sonst kaum an der Gruppendiskussion beteiligt war.

Die Textstelle, die mit ›Professionalisierung‹ annotiert wurde, lautet:

»IP_LS (...) im Ausbildungswesen haben wir eine Schule errichtet, ja, und dann haben die Damen und da sind Damen angenommen, die auch vorher einen Test machen mussten und die haben dann, äh, Lochen gelernt und Maschinenschreiben, ein bisschen Buchführung und Rechnen hatten die noch, aber der das Gebrauch wurde hier auf für uns gelegt. Die haben dann eine einjährige Ausbildung gemacht und sind dann zu mir hingekommen. Aber das ist inzwischen vorbei.«

»INT_MS Ist abgeschafft.«

»IP_LS Das ist abgeschafft worden. Äh, ich hab jetzt noch, eins, zwei, eins zwei, drei vier, fünf Damen, das sind dann aber schon ältere, aber was heißt älter, nicht über vierzig, ne, die keine Schul, unsere Schulausbildung haben, ja.«³⁴

Schaut man deskriptiv auf diese Textstelle, so steht hier die Buchhaltung mit Lochkarten im Mittelpunkt. Die erst im Laufe der Zeit erforderliche Schulbildung für die Ausbildung wurde von einem Studierenden im Workshop als ›Professionalisierung‹ in Bezug auf die Einstellungsvoraussetzungen gewertet. Da er sich, wie den Autor*innen bekannt ist, in Seminaren und einer Nebentätigkeit bereits mit Arbeitssoziologie und deren Fachterminologie auseinandergesetzt hatte, fand er im Material offensichtlich ein Beispiel für das ihm bekannte übergreifende Konzept.

Auch in der Gruppe A2 wurden in der offenen Annotation analytische Kategorien gebildet, insbesondere ›Carearbeit‹ und ›Hierarchien‹. Die Gruppenteilnehmenden waren sich einig, dass sie auf Grundlage der Lektüre genauer annotieren würden, als dies auf Grundlage des Modells zuvor geschehen sei. Vor der gemeinsamen Annotation tauschten sie sich darüber aus, dass sie allesamt kleinteilig annotiert und teils ausführliche Kategorien-Hierarchien entwickelt hätten. Der grundsätzliche Ansatz der differenzierten Annotation hatte sich also durch den vorherigen computationell bedingten Textzugang nicht verändert.

In B1 wurde stark diskutiert, wann Kategorien zusammengehören und wie dies im Kategoriensystem durch Unter- und Oberkategorien abgebildet werden könnte. Die Teilnehmenden stellten dabei auch fest, dass sie unterschiedlich feinteilig annotierten, sowohl was die Länge der Textstellen als auch die Genauigkeit und Anzahl der Kategorien angeht. Das gemeinsam gebildete Kategorienset bestand hier maßgeblich aus den im Vorfeld erstellten Kategorien von zwei Gruppenmitgliedern, die bereits über umfangreiche Erfahrungen in der qualitativen Annotation verfügten. Im Laufe der Annotation orientierte sich die Gruppe immer stärker an deren Kategorien.

In A1 hingegen wurde die Kategorisierung der einzelnen Chunks so vorgenommen, dass aus den individuellen Kategoriensystemen zunächst ein gemeinsames entwickelt und in der Folge genutzt wurde, um in aufgeteilter Arbeit die Chunks zu benennen. Zuvor wurde diskutiert, ob gewisse Kategorien vielleicht zusammenzufassen seien und wie dabei Unterkategorien – wie etwa die Übergänge in der Kollegialität zu Selbsthilfegruppen bis hin zu Betriebsräten – sichtbar gemacht werden könnten. Hier wurde also nicht nur die Größe einer Kategorie, sondern auch deren analytischer Gehalt diskutiert.

Wenn es eine Verbindung zwischen zwei Kategorien gab, annotierte die Gruppe B1 den Text immer mit allen in Frage kommenden Kategorien. Die Gruppe war sich in der Regel einig, welche Textstellen als Themenblock zusammengehörig sind. Da es teils auch abweichende Meinungen gab, vergab die Gruppe fast immer mehrere Kategorien, die einzelne Mitglieder mit individuell unterschiedlichen Schwerpunkten setzen wollten. Auch B2 behalf sich bei Mehrdeutigkeiten ebenso wie bei Uneinigkeit zwischen den Gruppenmitgliedern damit, Textstellen mehrfach mit unterschiedlichen Kategorien zu annotieren.

³⁴ ADG0042, Chunk 3.

In diesen Praxen zeigt sich ein Umgang mit Ambiguitäten und individuell verschiedenen Interpretationen³⁵, der in qualitativer Annotation stets neu ausgehandelt wird. Zugleich widersprechen solche Lösungen der computationalen Herangehensweise, die zwar mehrere Bedeutungen durch Gewichtungen von Topics deutlich macht, entsprechende analytische Setzungen jedoch nicht ermöglicht. Die Aushandlungsprozesse sind fragil und schon außerhalb der Gruppe kaum noch nachvollziehbar, sodass diese Form der Kategorisierung auch auf Probleme der Nachnutzung und Automatisierbarkeit von qualitativen Annotationsprozessen verweist, bei denen keine produktorientierten Annotationen³⁶ das Ziel sind. Dies schmälert allerdings nicht den Wert entsprechender Annotationen im eigenen Erkenntnisprozess.

Die diskutierten Bedeutungen verschoben sich beim Annotieren in den Gruppen im Laufe der Zeit, so wurde etwa in B1 festgestellt: »Das ist ja nicht das, was wir ursprünglich mit der Kategorie gemeint haben«. Zudem wurden beim Annotieren weitere Kategorien fortlaufend erstellt. Entsprechend notierte auch die Moderatorin während der Annotation der Gruppe: »Eben haben sie in der Kategorienerstellung gesagt, bei Technik sei Entwicklung wichtig, jetzt vergeben sie die Kategorie nicht an eine Textstelle weil ›ist ja keine Entwicklung.« Es zeigt sich hier besonders die Fluidität qualitativer Annotationen, die im Erkenntnisprozess immer weiter angepasst werden. Auch im experimentellen, sehr kurzen Zeitrahmen des Workshops fanden entsprechende Weiterentwicklungen und der Einbezug neuer Kontexte bereits statt, so dass daraus auch geschlossen werden kann, dass die manuelle Annotation im Versuchsaufbau durchaus vergleichbar mit von diesem Setting unabhängigen qualitativen Annotationsprozessen ist. Zugleich ist sie besonders verschieden von den feststehenden Kategorien maschineller Annotation.

4. Ergebnisse der computationalen Annotation

4.1 Überblick und erste Eindrücke

Die Analyse der Topic Models in den Gruppen erfolgte in einem ersten Schritt übergreifend. Dabei wurden die Wortlisten zu jedem der 50 generierten Topics einzeln gesichtet und in der jeweiligen Gruppe diskutiert. Die Gruppen einigten sich in diesem Schritt gemäß der Aufgabenstellung auf Topics, für die sie einen Bezug zum Thema ›Arbeit‹ sahen, und labelten diese Topics mit einem Schlagwort. Die Anzahl der selektierten Topics war dabei in allen Gruppen vergleichbar: Die beiden Teilgruppen, die zuerst Topic Modeling durchgeführt hatten, entschieden sich für 19 und 13 Topics, die beiden Teilgruppen, die zuerst händisch annotiert hatten, wählten 14 und 19 Topics aus (vgl. Tabelle 2). Die Auswahl von Topics unterscheidet sich also nicht nach der Reihenfolge der maschinellen bzw. manuellen Herangehensweise.

T.-Nr.	Von allen Gruppen gelabelte Topics	A1 TM (ante manuell)	A2 TM (ante manuell)	B1 TM (post manuell)	B2 TM (post manuell)
2	chef, büro, angestellt, abteilung, thyssen, sekretärin, abteilungen, arbeit, herren, damen, einstellen, personalabteilung, hütte, kollegen, beschäftigen	Anstellung / Verwaltung	Positionen in einer Firma	Organisation von Arbeit	Arbeitsorganisation
6	arbeit, menschen, fühlen, verändern, art, beruflich, beruf, entscheiden, wesentlich, schwer, tätigen, hart, praxis, herne, selbstverständlich	<i>Beruf (subjektive Perspektive)</i>	<i>Arbeit und Identität</i>	<i>Einstellung zur Arbeit</i>	<i>persönliche Einstellung zum Beruf</i>
16	betriebsrat, gewerkschaft, betrieb, kollegen, gewerkschaften, betriebsräte, vorsitzend, wählen, metall, belegschaft, gewerk-	Interessenvertretung / Betriebsperspektive	politische Organisation von Arbeit / Betriebsrat	Interessenvertretung Arbeit	Gewerkschaft

Tab. 2: Von allen Gruppen gelabelte Topics. Grün hervorgehobene Label haben auf der Horizontalen eine hohe Übereinstimmung zwischen den Gruppen, blau hervorgehoben sind starke Unterschiede auf der Horizontalen zwischen den Gruppen. Kursiv gekennzeichnete Labels sind besonders differenziert. Gruppierung durch die Autor*innen, wobei die konkrete Gruppierung wiederum auch durch die Kenntnisse und Perspektiven der Autor*innen geprägt ist.

³⁵ Vgl. Andresen et al. 2020.

³⁶ Vgl. Franken et al. 2020, S. 104.

	schaftlich, betrieben, organisieren, ange stellt, vertreten				
22	meter, kohlen, kohle, holz, wasser, wagen, steine, stempel, dick, hand, raus, luft, eisen, kaputt, bohren	<i>Arbeitsvorgänge Mon-tanindustrie</i>	<i>Arbeitspraxis im Berg</i>	<i>Arbeitspraxen im Berg-werk</i>	<i>Bergbau / Arbeitsbedin-gungen</i>
27	krieg, einziehen, ar-beitsdienst, soldat, krieges, freiwillig, an-fang, mitmachen, zu-rückkommen, mona-te, verlieren, solda-ten, entlassen, dienen, schluss	Kriegsdienst	Kriegsdienst	<i>Kriegserlebnis</i>	Kriegsdienst
29	russland, lazarett, rus-sen, schießen, frank-reich, soldaten, wo-chen, stellung, ge-le-gen, offizier	Fronteinsatz / Solda-tenerfahrung	Stationierung als Sol-dat	Kriegspraxis	Kriegseinsatz
30	kinder, kind, eltern, fam-ilie, tochter, kindern, heiraten, sohn, verhei-raten, mädchen	<i>Familie / Nahbereich</i>	<i>Reproduktionsarbeit</i>	familiäre Beziehungen	Familie
46	zechen, bergbau, stei-ger, kohle, dintavpe, bergleute, schicht, bergmann, schacht, kumpel	Bergbau	Bergbau	<i>Gegebenheiten im Koh-lebergbau</i>	Bergbau
	Von drei Teilgruppen gelabelt				
20	krankenhaus, arzt, kranken, sterben, wo-chen, schwer, bein, ge-le-gen, krankheit, dok-tor	-	Krankheit	körperlicher Zustand	Lebensstandard
24	bauen, wohnen, hei-materde, wohnungen, häuser, stadt, krupp, siedlung, grundstück, gelände	-	Wohnsituation / Unter-bringung	Wohnen	Wohnbedingungen
26	arbeit, schwer, arbeits-los, öhm, arbeitsamt, wochen, erwerbslos, entlassen, schlecht, ar-beiter	Erwerbslosigkeit	-	<i>Arbeitslosigkeit / Preka-rität</i>	Arbeitslosigkeit
28	lernen, lehre, geschäft, beruf, laden, ausbil-dung, sohn, firma, selbständig, gesellen	Ausbildung / Berufs-stand	<i>Wege in kaufmännische Ausbildung</i>	Ausbildung	-
42	meistern, arbeit, mensch, maschine, ding, schwer, kaputt, betrieb, drehen, ma-schinen	<i>Industriearbeit / Maschi-nen</i>	-	Fabrikarbeit	Technologie

Tab. 2: Von allen Gruppen gelabelte Topics. Grün hervorgehobene Label haben auf der Horizontalen eine hohe Übereinstimmung zwischen den Gruppen, blau hervorgehoben sind starke Unterschiede auf der Horizontalen zwischen den Gruppen. Kursiv gekennzeichnete Labels sind besonders differenziert. Gruppierung durch die Autor*innen, wobei die konkrete Gruppierung wiederum auch durch die Kenntnisse und Perspektiven der Autor*innen geprägt ist.

Auch auf inhaltlicher Ebene ist kein deutliches Muster zu erkennen, das auf einen Unterschied beim Labeln der Topics mit oder ohne Vorkenntnisse der Quellen schließen lässt. In allen Gruppen finden sich unterschiedliche Differenzierungsgrade in der Kategorisierung. Für Topic 6 wurden in allen Teilgruppen ähnlich differenzierte Kategorien gebildet: »Beruf (subjektive

Perspektive), »Arbeit und Identität« (Gruppen A1 und A2), »Einstellung zur Arbeit«, »Persönliche Einstellung zum Beruf« (Gruppen B1 und B2). Auch Topic 22 ist über alle Gruppen hinweg differenziert gelabelt: »Arbeitsvorgänge Montanindustrie«, »Arbeitspraxis im Berg« (Gruppen A1 und A2), »Arbeitspraxen im Bergwerk«, »Bergbau / Arbeitsbedingungen« (Gruppen B1 und B2). Topic 26 wurde von B1 als einer der Teilgruppen, die zuerst händisch annotiert hatten, mit »Arbeitslosigkeit / Prekarität« differenzierter kategorisiert als von den anderen Gruppen. Topic 28 wurde wiederum von A2 als einer der Teilgruppen, die zuerst die Topics gelabelt hatten, mit »Wege in die kaufmännische Ausbildung« detailliert bezeichnet.

Damit zeigt sich, dass eine Exploration mit Topic Modeling auch ohne vorherige Kenntnis der Quellen zuverlässig möglich ist und die *Topic-Term-Matrix* erweist sich als guter Einstieg in Filterprozesse.³⁷ Zu bedenken ist hier allerdings, dass das verwendete Modell im Vorfeld intensiv getestet und angepasst worden war³⁸ und dass die Themensetzung »Arbeit« gut zugänglich und allgemeinverständlich ist. Die hohe Differenzierung der Labels für einzelne Topics verweist wiederum (vergleichbar zu den Schritten in der manuellen Annotation) auf die vorhandenen Wissensbestände der Workshopteilnehmenden und macht deutlich, dass diese auch bei der computationellen Analyse relevant sind.

4.2 Einstieg in die Daten

Nach der übergreifenden Exploration der Topics wurden mit Hilfe des Jupyter Notebooks stichprobenartig Interviewpassagen ausgegeben, die den relevanten Topics zugeordnet waren. In allen Gruppen wiesen die Teilnehmenden auf die voraussetzungsreiche Entstehung der Ergebnisse des Topic Modeling und die Schwierigkeit hin, diese inhaltlich einzuordnen. Jedoch waren die Vorbehalte in den Gruppen größer, die sich den Quellen zunächst über das Topic Modeling näherten (A1 und A2). Bei der ersten Begutachtung der Topics wurde dort angemerkt, dass eine Kontextualisierung des Textes fehle, was auf die Relevanz des Einbezugs von Metadaten verweist.³⁹

In den Gruppen, die zunächst manuell annotierten (B1 und B2), wurden die Chancen maschineller Verfahren bei der Erschließung von Textkorpora deutlicher benannt als in den anderen Gruppen. Gleichzeitig wurde dort besonders über die epistemologischen Herausforderungen einer computationellen Textauswertung diskutiert. Die Vogelperspektive sei verführerisch, erfordere aber das wiederholte Hineinzoomen in den Text, um ein Gespür sowohl für die Qualität und Aussagekraft der Topics als auch das Format der Quelle zu bekommen. Diffizil sei insbesondere, die Lenkung des Interviews durch die Interviewenden aus den Topics herauszulesen. Damit stimmen die Erkenntnisse der Gruppen mit denen aus anderen Forschungsprojekten überein.⁴⁰

4.3 Evaluation der Topic-Listen

Nachdem Topics identifiziert und mit Labels versehen wurden, widmeten sich die Gruppen der Differenzierung ihrer Auswahl. Insgesamt wurde hervorgehoben, dass die Topics zum Thema »Arbeit« einen hohen Grad an Differenzierung aufwiesen und dass »Topics keine geschlossenen Entitäten« seien. Es konnten zum Teil spezifische Perspektiven in zunächst sehr ähnlich anmutenden Topics ausgemacht werden. Das systematische Gegenüberstellen »verwandter« Topics führte zur Wahrnehmung teils subtiler Nuancen, die einerseits als »Draufblick« und andererseits als »Innenansicht« auf die Arbeit im Ruhrgebietsbergbau in den Topics 22 und 46 (vgl. Tabelle 2) bezeichnet wurden. Entlang der Zuordnung entfaltete sich eine Reflexion über die semantischen Zusammenhänge in den statistisch generierten Listen und ausgehend davon eine epistemologische Diskussion, wie man das richtige Label für eine Liste von zehn, 15 oder 30 Wörtern finden könne. Dabei spiele nicht nur Hintergrundwissen, sondern auch Fantasie eine entscheidende Rolle, so die übergreifende Einschätzung. Das Einigen auf ein Label sei ein stetiger Verhandlungsprozess, in dem sich die individuellen Blicke auf die Topic-Liste verändern könnten.

Zur Differenzierung wurde auch Kontextwissen herangezogen, wie etwa zur Einordnung und Benennung von Topic 16, das in der Gruppe, die zunächst Topic Modeling angewandt hat, etwas differenzierter ausgefallen ist (vgl. Tabelle 2). Dabei mag von Bedeutung gewesen sein, dass der Anteil von Historiker*innen und Forschenden mit entsprechendem Schwerpunkt in dieser Gruppe höher war als in der anderen Gruppe – was zugleich kritisch reflektiert wurde. Abseits von Fachwissen half auch die Topic-inhärente Analyse der Zusammensetzung auf Grundlage von Wortlisten mit 30 oder mehr Top-Keywords. Gegenstand tiefgehender Auseinandersetzung war hier z. B. Topic 6 (vgl. Tabelle 2), wo in Gruppe A1 festgestellt wurde, dass »sehr unterschiedliche Begriffe zu finden sind und gerade bei Fällen, wo der wichtigste Begriff (Arbeit) nicht dominant war, es mir schwer fiel nachzuvollziehen, inwiefern dieses Topic zu Arbeit passt. Es müssen also unterschiedliche kleinere Begriffe dazu geführt haben, dass [...] dennoch Topic 6 gewählt wurde.« Hier zeigt sich, dass Topic Modeling auch das implizite

³⁷ Vergleichbare Ergebnisse bei Koch / Franken 2020, S. 133.

³⁸ Vgl. Hodel et al. 2022.

³⁹ Vgl. Apel 2015, S. 249.

⁴⁰ Vgl. Welzer 2000, S. 52–53; Hornung 2010, S. 32–33.

Sprechen einfangen kann und aus der Wahrnehmung, das Topic sei ›schwammig‹, bei näherer Betrachtung die Erkenntnis erwuchs, es würde einen bestimmten »Blick auf [ein] Phänomen beschreib[e]n«. Es wird deutlich, wie wertvoll eine intensive Auseinandersetzung mit der Topic-Term-Matrix für den Erkenntnisgewinn sein kann – was mit der Gruppendiskussion von Kategoriensets in der manuellen Annotation korrespondiert. Bemerkenswert ist, dass das Topic 6, das von Gruppe A1 zunächst schwammig und dann als Erweiterung des Blicks auf Arbeit eingeordnet wurde, von B1, die die Quellen bereits kannte, als ›Einstellung zur Arbeit‹ gelabelt wurde, nachdem ›Soziales Leben‹, ›Lebenslauf‹ und ›Berufswahl‹ diskutiert und verworfen worden waren. Die Gruppe hatte scheinbar keine Schwierigkeiten, das Topic in Zusammenhang mit dem Erkenntnisinteresse des Workshops zu bringen und stieg direkt in die Kategorisierung ein.

4.4 Exploration und kritische Reflexion

In der Folge wurden die für die manuelle Annotation im Vorfeld ausgewählten Interviewpassagen betrachtet. Die Gruppen sollten diese beurteilen und überprüfen, ob die Textstellen auf Grundlage der freien Exploration der Topics und der dabei gesammelten Eindrücke von ihnen ebenfalls als relevant für ein Sample eingeschätzt werden. An diese exemplarischen Chunks wurden schließlich in CATMA die Oberbegriffe der drei im Model am stärksten gewichteten Topics annotiert. Die Vergabe der Labels für diese Topics an die jeweiligen Interviewpassagen konnten die Teilnehmenden dem Jupyter Notebook entnehmen, das die gewichteten Topics für jeden Textabschnitt ausgeben kann. Heatmaps der Topic-Document-Listen, die sowohl die Verteilung von Topics über Interviews (vgl. Abbildung 1) als auch die Chronologie und Schichtung der Topics innerhalb einzelner Interviews (vgl. Abbildung 2) darstellen, ermöglichten den kleinteiligen Rückbezug auf die sehr umfangreichen einzelnen Interviewtranskripte und somit ein textnäheres Arbeiten.

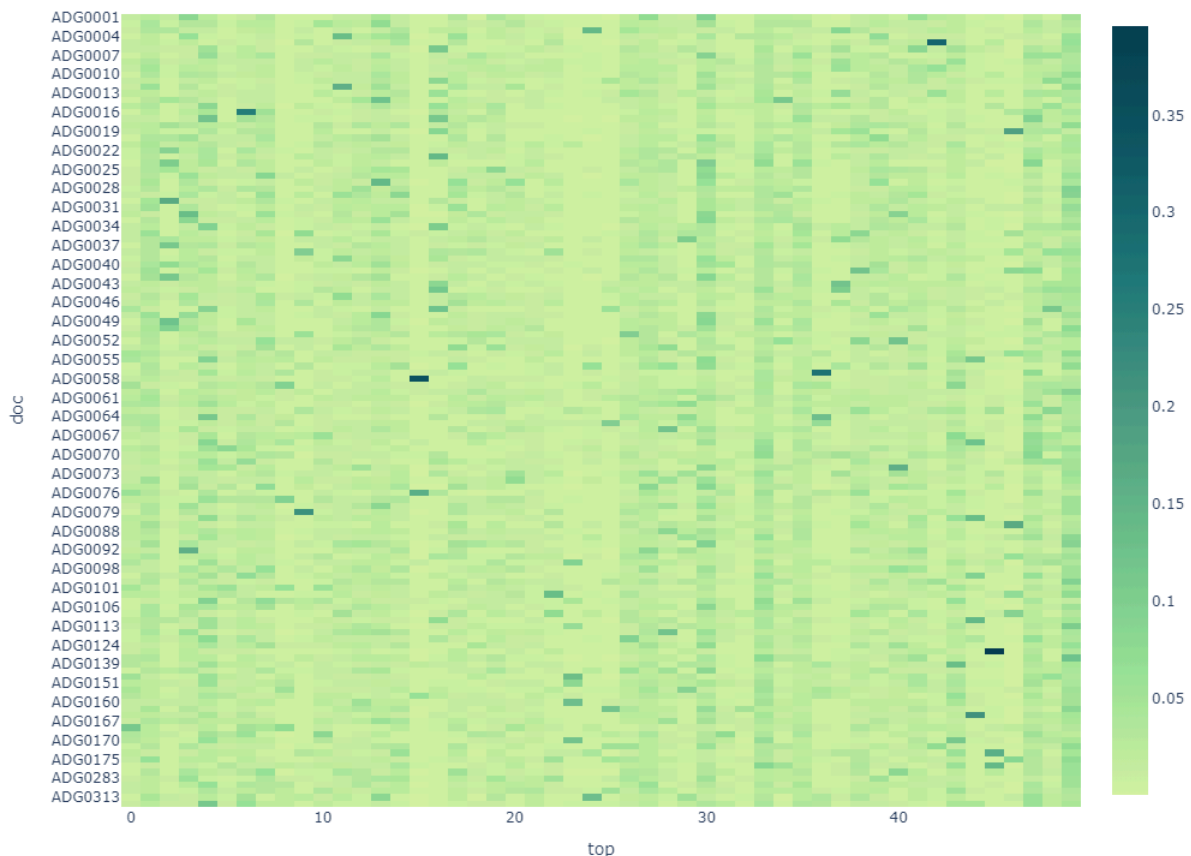


Abb. 1: Globale Heatmap mit der Verteilung aller Topics (X-Achse) über alle Interviews (Y-Achse). Zugrunde liegen jeweils die summierten Topic-Weights aller Chunks eines Interviews. [Grafik: Dennis Möbus 2024]



Abb. 2: Die Verlaufsmatrix zeigt die Topics anhand der ersten zehn Wörter (Y-Achse) im chronologischen Verlauf des Interviews, gegliedert nach Chunks (X-Achse). [Grafik: Dennis Möbus 2024]

Der Wunsch, direkt über die visuelle Oberfläche Muster auszuwählen und weiter zu filtern, wurde in den beiden Gruppen geäußert, die erst an Topic Models arbeiteten (A1 und A2). Das zog jedoch die Kritik eines Historikers nach sich, der anmerkte, die »Heatmap lenk[e] den Blick zwangsläufig auf Stellen, die gesuchte Inhalte thematisieren. Dabei geht verloren, was nicht angesprochen wird.« Der Verweis auf Leerstellen im Material ist jedoch kein Alleinstellungsmerkmal dieses Verfahrens und für manuelle ebenso wie für computationale Ansätze relevant. Interessant ist jedoch seine Erkenntnis, »über die Matrix erschließen sich eher die Randthemen des gesamten Korpus, die in einzelnen Interviews hervorstechen.« Somit kann Topic Modeling zwar den Blick auf die Gewichtung von Themen verzerren, hat aber Potenzial, Ausreißer zu lokalisieren, die unter dem Radar des vordergründigen Forschungsthemas einer Interviewsammlung fliegen.⁴¹ Hier wird die Nutzung als Findeheuristik für spezifische Fragestellungen⁴² besonders deutlich – was für Sekundäranalysen relevant ist und von den Forschungsinteressen der Primärforschung deutlich abweichen kann, weshalb Ausreißer interessieren.

Während der Auseinandersetzung mit dem Material in den durch den Workshop vorstrukturierten Schritten wurde eingeräumt, dass sich die Differenzierung der Topics durch wiederholte Lektüre verknüpfter Textpassagen zunehmend erschließe. Man kann entsprechend festhalten, dass ein iteratives Oszillieren zwischen *Distat* und *Close Reading*, auch als *Scaleable Reading*⁴³ gefasst, eine effektive Methode ist, um die Qualität eines Topic Models zu evaluieren und dieses in qualitative Erkenntnisprozesse einzubinden. Der Perspektivwechsel führte in einer der Gruppen, die bereits mit den Quellen vertraut war (B2), zu einer sehr grundlegenden Diskussion: Einerseits wurde die Anwendung quantitativer Analysen im Rahmen qualitativer Forschung in Frage gestellt. Andererseits wurde argumentiert, dass thematische Häufungen als erste Interpretation genutzt werden können und Topic Modeling somit die Forschungsperspektive auf Quellen ändere. Daran zeigt sich, dass bereits nach kurzer Beschäftigung mit den so veränderten Zugängen grundsätzliche epistemologische Fragen aufgeworfen werden und angenommen werden kann, dass dies besonders durch die enorme Bedeutung des Hinterfragens eigener Methoden in qualitativer und kulturwissenschaftlicher Forschung bedingt ist.⁴⁴

Die Prämisse, die Methode des Topic Modeling verstehen zu müssen, um fundierte Forschung damit betreiben zu können, wurde von mehreren Teilnehmenden der Gruppen geäußert, die zunächst manuell annotiert hatten (B1 und B2). Eine brisante Frage, die vor diesem Hintergrund aufkam, war, ob Machine Learning generell Zusammenhänge *erzeuge* oder *abbilde*, was die entscheidende Frage nach sich zog, ob Topic Modeling schon ein Akt der Interpretation sei. Die vorherrschende Meinung der Workshopeteilnehmenden war, dass ein statistisches Computerprogramm nicht *verstehen* könne – was eine Grundbedingung für jede Art von Interpretation sei. Als Gegenargument wurde angeführt, dass eine Selektion von Daten sehr wohl ein Akt der Interpretation sei. Wenngleich nicht sinnstiftend, sondern nach mathematischen Regeln herbeigeführt, handele es sich doch

⁴¹ Vgl. Möbus 2020.
⁴² Vgl. Adelman et al. 2019.
⁴³ Vgl. Müller 2020.
⁴⁴ Vgl. Franken 2022.

um einen Akt der Kondensierung. Auch hier zeigt sich, wie schnell grundsätzliche Diskussionen um die Handlungsmacht von Algorithmen und Machine Learning⁴⁵ aufleuchten und wie relevant diese für die kritische Übernahme der entsprechenden Methoden in neue Forschungskontexte wie die Oral History und qualitative Forschung sind.

5. Vergleich und Diskussion der Ergebnisse

5.1 Unterschiede in den Annotationen

Durch die Arbeit in den Kleingruppen stehen insgesamt acht Kategorien-Sets zur Verfügung, die sich auf jeden Chunk hin abbilden lassen: Jede der vier Gruppen hat sowohl manuelle Kategorien erstellt als auch Oberbegriffe für die Topics vergeben. Im Folgenden sollen diese Annotationen, die auch Interpretationen aufzeigen können, vergleichend einem Close Reading unterzogen werden. Es wird in der Übersicht (vgl. Tabelle 2) schnell deutlich, dass ein Topic, basierend auf einer Liste von Wörtern, im Labeling durch die Gruppen zwar unterschiedliche, aber in der Summe recht ähnliche Oberbegriffe evoziert hat. So wurden etwa für das Topic 2 von Gruppe A1 ›Anstellung / Verwaltung‹ vergeben, von Gruppe A2 die Kategorien ›Positionen in einer Firma‹. Gruppe B1 bezeichnete dieses Topic als ›Organisation von Arbeit‹, Gruppe B2 als ›Arbeitsorganisation‹. Ähnlich verhielt es sich mit den manuell vergebenen Kategorien (vgl. Tabelle 1): Gruppe B1 bezeichnete ihre Kategorie als ›Technische Entwicklung‹, Gruppe A1 hingegen als ›Wandel Arbeitstechnik‹. Die durch Gruppe A1 benannten ›Arbeitskonflikte‹ annotierte Gruppe B1 mit ›Arbeitsbedingungen‹, nahm also ebenfalls ähnliche Interpretationen vor, wenn auch in geringerer Deutlichkeit. Insgesamt wurden die computationellen Topics wesentlich allgemeiner annotiert, während die manuelle Annotation zielgenauer erfolgte und bekannte Kontexte (etwa zu historischen Ereignissen) einbezogen wurden.

Um die Unterschiede und Gemeinsamkeiten in der Annotation am Beispiel nachzuvollziehen, werden im Folgenden drei aufeinander folgende Chunks aus einer im Workshop analysierten Textstelle einem Close Reading unterzogen und die von den Gruppen jeweils vergebenen Kategorien genauer untersucht. Es handelt sich dabei um die ersten drei Chunks der vorgelegten Textstelle.

ADG0054, Chunk 83

00:30:38.17	INT_NM	Ach so, jaja. [lacht]
00:30:39.22	INT_NM	Ähm, was wollte ich noch fragen.
00:30:44.16	INT_NM	Was waren denn damals die Konfliktpunkte überhaupt. Also ...
00:30:48.15	IP_HB	Ja Gott, Arbeitszeiten, natürlich.
00:30:51.17	IP_HB	Und dann die Frage ...
00:30:53.18	IP_HB	Die Frage der, der, immer hier, der, der Piepen, nich, das war immer ein großer Streitfall.
00:30:59.17	IP_HB	(...) zu meiner Zeit auch schon.
00:31:02.18	IP_HB	Oh ja, hab ich auch immer den Kopf drüber geschüttelt.
00:31:06.00	IP_HB	Da hab ich weiß Gott noch– Als ich ein junger Mann war und, äh, als Student noch im Bergbau tätig war, gabs relativ kleine Förderwagen, kleine Wagen.
00:31:13.26	IP_HB	Ging ne Tonne rein, meistens nicht mehr.
00:31:16.05	IP_HB	Waren so hoch.
00:31:17.28	IP_HB	Die Leute wurden nach Wagen bezahlt.
00:31:20.18	IP_HB	Kleine Wagen, nich.
00:31:22.13	IP_HB	Ich weiß nicht, da kamen ... (...) Schicht kam ... nach Wagen bezahlt.
00:31:25.24	IP_HB	Gut, und jeder Wagen kriegte eine Nummer.
00:31:28.24	IP_HB	Die Leute, die da unten ... die Füller, die da am Wagen, die hatten so Drähte, die wurden da vorne reingesteckt, da war dann die Nummer drauf.
00:31:37.00	INT_NM	Ja.
00:31:38.05	IP_HB	Und der Mann, der oben über Tage die Wagen kippte, zog die Nummer raus und sammelte die.

⁴⁵ Vgl. die theoretischen Diskussionen dazu etwa bei Carlson et al. 2021; Introna 2016; Kitchin 2014; Seaver 2017; Jatón 2021.

ADG0054, Chunk 83

00:31:43.00	IP_HB	Jedes Revier hatte seine eigene Nummer.
00:31:45.08	IP_HB	Und nach den geförderten Wagen wurde dann auch bezahlt.
00:31:49.02	IP_HB	Ja, und eines Tages kamen kluge Betriebsführer dahinter, oder Inspektoren dahinter, dass es doch ganz sinnvoll wäre, wenn man so n Kragen auf den Wagen setzte.
00:31:59.20	IP_HB	Dann wurde der höher.

	Topic	A1	A2	B1	B2
Topic Modeling	22	Arbeitsvorgänge Montanindustrie	Arbeitspraxis im Bergbau	Arbeitspraxen im Bergwerk	Bergbau / Arbeits- bedingungen
	46	Bergbau	Bergbau / Arbeitsall- tag	Gegebenheiten im Kohlebergbau	Bergbau
	33	-	-	-	-
manuelle Annota- tion		Arbeitskonflikte	Arbeitswelt	Leistung / Entloh- nung	Arbeitsbedingungen
		-	Geld / Lohn / Finan- zielles	technische Entwick- lung	Technologie / tech- nologische Entwick- lung
		-	-	-	Macht

Tab. 3: Topics und manuelle Kategorien, die in den Gruppen der Textstelle ADG0054, Chunk 83 zugeordnet wurden. Farblich identisch markierte Kategorien zeigen inhaltliche Übereinstimmung an. Eine Liste aller Topics, abgebildet anhand der Top-30-Keyworts, ist auf [GitHub](#) einsehbar.

Die Gewichtung der Topics zu dieser Interviewpassage – 22 und 46 dominieren in diesem Abschnitt – bildet den inhaltlichen Schwerpunkt der manuellen Annotation gut ab. Die Labels deuten zwar die Innenansicht der Arbeitenden im Kohlebergbau an, der in diesem Chunk angedeutete Konflikt wird durch die reine Ansicht der Topics allerdings nicht sichtbar. Diese eher statische Perspektive spiegelt sich in den manuellen Annotationen der Gruppen wider, die zuerst Topic Modeling durchgeführt haben (A1 und A2). Ob nach einem Topic Modeling insgesamt auch in der händischen Annotation reduzierter annotiert wird, sollte in weiteren Studien aufmerksam verfolgt werden, um den hier zu erkennenden Trend zu überprüfen.

Auffällig ist, dass der Vorgabe, die jeweils drei am stärksten gewichteten Topics einer Passage zu annotieren, von keiner der Gruppen entsprochen wurde. Das liegt daran, dass im Ranking auf Platz drei eines der verrauschten, opaken Topics (33) lag, das weder im initialen Labeling der Topic-Listen mit dem Phänomen ›Arbeit‹ in Verbindung gebracht werden konnte, noch diesem Chunk eine adäquate inhaltliche Nuance hinzugefügt hätte.⁴⁶ Dementsprechend wird deutlich, dass menschliches Verstehen auch durch entsprechend statistische Bewertungen nicht unterdrückt wird, denn die Workshopteilnehmenden interpretierten lediglich jene Topics, denen sie Sinn entnehmen bzw. zuschreiben konnten.

Was durch das Topic Modeling nicht abgedeckt wird und auf die Schwächen der Methode verweist, ist das Nicht-Einfangen dynamischer Prozesse in den genannten Beispielen.⁴⁷ Das liegt daran, dass aus lebensgeschichtlichen Erzählungen oft erst im interpretierenden Nachvollzug Dynamiken und Entwicklungen herausgelesen werden. Die Erzählenden verwenden nicht immer analytische Begriffe wie »Prozess«, »Entwicklung« oder »Dynamik«, um Episoden zu kennzeichnen, die von Wandel geprägt sind. Gleichwohl lassen sich Dynamiken einfangen, wenn beispielsweise unterschiedliche Topics zu einem Thema entstehen, die einen Status quo ante und einen Status quo post einer Entwicklung darstellen. In bestimmten Fällen – wenn beispielsweise Interviews zur Untersuchung gesellschaftlicher Transformationen durchgeführt werden – kann die erzählende Darstellung analytischer ausfallen. Gleichzeitig lassen sich immer wieder einzelne Topics finden, in denen Zeigewörter für Dynamiken gruppiert sind, auch wo diese nicht in entsprechend analytischen Begriffen explizit auftauchen. In den wesentlich differenzierteren manuellen Annotationen der Gruppen, die vor dem Topic Modeling händisch ausgewertet hatten, finden sich Bewegungsbegriffe wie ›technische‹ und ›technologische Entwicklung‹, das dynamische Verhältnis ›Leistung / Entlohnung‹ und das analytische Label ›Macht‹. Die manuelle Annotation deutet zudem über die Gruppen hinweg auf Konflikte in den Arbeitsbedingungen hin.

Dies wird auch im weiteren Textverlauf und dessen Annotation deutlich:

⁴⁶ Die ersten zehn Wörter von Topic 33 sind: moment, schluss, schlecht, sachen, stimmen, lieb, überlegen, versuchen, würd, gedenken.

⁴⁷ Vgl. Althage 2022, S. 272.

ADG0054, Chunk 84

00:32:01.01	IP_HB	Dann ging da mehr rein.
00:32:02.10	INT_NM	Mhm.
00:32:02.28	IP_HB	Nich.
00:32:03.20	IP_HB	Da müsste man meines Erachtens auch mehr dafür bezahlen.
00:32:06.14	IP_HB	Aber damit hatten die ... das wollten die gar nicht.
00:32:08.15	IP_HB	Die haben nur gesagt:
00:32:09.11	IP_HB	Machen wir weniger, haben wir mehr ... weniger Wagen, aber mehr Inhalt, nich.
00:32:12.26	IP_HB	Und schlechteren Lohn für die Leute.
00:32:14.16	IP_HB	Ist doch ganz eindeutig, nich.
00:32:15.20	IP_HB	Und die wurden dann immer wieder, immer wieder höher gemacht, immer wieder höher gemacht.
00:32:19.19	IP_HB	Und das gab Ärger, mit Recht.
00:32:21.20	INT_NM	Mhm, ja sicher.
00:32:22.00	IP_HB	Unkorrekt meines Erachtens, völlig unkorrekt, völlig unkorrekt.
00:32:25.06	IP_HB	Wir haben das nachher geändert, wir haben die Leute nach Quadratmetern bezahlt.
00:32:29.15	IP_HB	Quadratmetern, die sich an Fläche im Streb, im Kohlegewinnungsfeld, nich.
00:32:35.12	INT_NM	Ja.
00:32:36.00	IP_HB	Äh, von Kohle frei machten, über die Grundfläche, nich.
00:32:39.12	IP_HB	Die wurden vermessen, die Länge war bekannt und der Steiger ...
00:32:43.00	IP_HB	Und der Steiger konnte dann zu Ende der Schicht immer aufschreiben:
00:32:47.09	IP_HB	der hat soviel Meter, der hat soviel Meter, der hat soviel Meter.
00:32:49.22	INT_NM	Ja.
00:32:50.00	IP_HB	Da gabs Leute dabei, die waren stämmig und die verstanden ihr Geschäft.
00:32:53.15	IP_HB	Die machten deswegen 10 Quadratmeter und es gab welche, die machten nur 6.
00:32:57.28	IP_HB	Ah, es gab auch welche, die machten vielleicht 15.

	Topic	A1	A2	B1	B2
Topic Modeling	46	Bergbau	Bergbau / Arbeitsalltag	Gegebenheiten im Kohlebergbau	Bergbau
	22	Arbeitsvorgänge Montanindustrie	Arbeitspraxis im Bergbau	Arbeitspraxen im Bergwerk	Bergbau / Arbeitsbedingungen
	43	-	Erzähleinleitung / Aufrechterhaltungsfragen	-	-
manuelle Annotation		Arbeitskonflikte	Geld / Lohn / Finanzielles	Verkörperlichung / Körperlichkeit von Arbeit	Arbeitsbedingungen
		Motivation / Identifikation	Arbeitsverhältnisse im Betrieb	Leistung / Entlohnung	Technologie / technologische Entwicklung
		-	-	Widerstand	Macht

Tab. 4: Topics und manuelle Kategorien, die in den Gruppen der Textstelle ADG0054, Chunk 84 zugeordnet wurden. Farblich identisch markierte Kategorien zeigen inhaltliche Übereinstimmung an.

Die Vergabe der Topics verändert sich hier im Vergleich zum vorherigen Chunk nicht, noch immer geht es um die Innenansichten von Arbeitenden im Bergbau. Interessant ist, dass die Gruppe A2 als einzige Gruppe das für diese Textstelle computationell ermittelte und nicht arbeitsbezogene Topic 43 gelabelt hat, sogar recht analytisch als ›Erzähleinleitung/ Aufrechterhaltungsfragen‹. Hier können wir selbstkritisch mitnehmen, dass in Zukunft – unabhängig von inhaltlicher Fokussierung solcher Workshops – konsequent alle Topics gelabelt werden sollten. Denn anders als im vorherigen Chunk, kann Topic 43 hier Nuancen hinzufügen. Die ersten Keywords ›geschichte‹, ›sachen‹, ›interessant‹, ›zeiten‹, ›sozusagen‹ muten zwar nicht sonderlich konsistent an, verweisen im Kontext der Passage aber auf die historische Bedeutung der Erzählung: Vor dem Krieg wurde das Gedinge, die Bezahlung Untertage, solidarisch für die Gruppe ausgegeben und geteilt, nach dem Krieg wurden die individuellen Meter Kohle, die man Untertage freilegte, für die Abrechnung herangezogen.⁴⁸ Diese Erzählung wird durch das Meta-Topic 43 zumindest annäherungsweise eingefangen.

Die manuellen Kategorien differenzieren sich hier zwischen den Gruppen weiter aus. A1, B1 und B2 versuchen, die Konflikte, den Widerstand bzw. die Machtverhältnisse im Betrieb, die in der Erzählung deutlich werden, mit eigenen Kategorien zu greifen. Mit ›Motivation / Identifikation‹ hat Gruppe A1, die zunächst Topic Modeling durchgeführt hat, nun auch eine analytische Kategorie angelegt. Bemerkenswert ist hingegen, dass bei der expliziten Darstellung von unterschiedlichen körperlichen Fähigkeiten nur eine Gruppe (B1) die Körperlichkeit auch als Kategorie vergeben hat, dies also trotz der Thematisierung im Text und unabhängig vom methodischen Zugang nicht als bedeutsam erachtet wurde. Es kann nur vermutet werden, dass dies ggf. am Vorwissen der Workshopteilnehmenden oder an der Fokussierung auf ›Arbeit‹ lag.

Hier ergibt sich außerdem eine weitere Limitierung durch das Topic Modeling: Werden Labels einmal pro Topic vergeben, geht das oft zu Lasten der teils in sich recht differenzierten Topics. Denn während Topic 46 eindeutig auf Strukturen und Organisation im Bergbau fokussiert, treten bei Topic 22 auch persönliche Aspekte, wie eben körperliche Belastung, hervor: ›dick‹, ›hand‹, ›luft‹, ›kaputt‹, ›bohren‹, ›dicke‹, ›hacken‹, ›winter‹, ›rutschen‹, ›kalt‹, ›gas‹, ›schlagen‹ sind Begriffe, die sich unter den 50 häufigsten Keywords befinden und durch die Labels ›Arbeitsvorgänge‹, ›Arbeitspraxis/en‹ und ›Arbeitsbedingungen‹ nur grob eingefangen werden.

Der weitere Textverlauf zeigt eine noch differenziertere Annotation:

ADG0054, Chunk 85

00:33:01.08	IP_HB	Also ganz völlig unterschiedlich.
00:33:02.10	IP_HB	Die wurden natürlich unterschiedlich auch bezahlt.

⁴⁸ Vgl. Tenfelde 1981, S. 297–298; Brüggemeier 1983, S. 99.

ADG0054, Chunk 85

00:33:04.00	IP_HB	Jeder nach seiner Leistung, nich.
00:33:05.19	IP_HB	Das war an sich n korrektes Gedinge, nur hat das natürlich auch einen Fehler, insofern ...
00:33:10.28	IP_HB	Ich hab immer auch gepredigt, ich sag:
00:33:12.20	IP_HB	Leute, ihr redet immer hier von Kameradschaft bei euch im Pütt, ich bin da ganz anderer Auffassung.
00:33:17.14	IP_HB	Von Kameraden ...
00:33:18.10	IP_HB	Wenn einer unter ...
00:33:19.11	IP_HB	Wenn einer hier untern Bruch kommt, zu Deutsch unter Gestein kommt, dann ist die Kameradschaft bei euch groß.
00:33:25.00	IP_HB	Da riskiert ihr euer eigenes Leben um den Mann zu retten.
00:33:28.08	IP_HB	Da seid ihr Kameraden.
00:33:29.08	IP_HB	Aber ihr seid ja auch nicht alle Tage gleich stark und gleich wieder, gleich kräftig und gleich gesund.
00:33:36.00	IP_HB	Ich kann mir gut vorstellen, dass ein schwächlicher Mann mit 5 Kindern zu Hause, ich übertreib mal immer, damit ihr es auch noch merkt, nun, äh, nicht 10 Meter machen kann.
00:33:45.20	INT_NM	Mhm.
00:33:46.20	IP_HB	Und da sagt dann der Steiger:
00:33:49.00	IP_HB	Heute nur 7.
00:33:50.10	IP_HB	Oder der sich 10 Meter vornimmt, aber bis Ende der Schicht damit nicht fertig wird.
00:33:56.04	IP_HB	Das ist auch lästig, denn man musste früher die, die, die Rutschen umlegen.
00:34:01.03	IP_HB	Das konnte man nur, wenn alle Leute ihre Kohlen raus hatten.
00:34:04.05	INT_NM	Ja.
00:34:04.08	IP_HB	Nich.
00:34:04.11	INT_NM	Jaja.
00:34:05.21	IP_HB	Dann würde ich ja sagen: da müssen dann auch Kerle, die so 15 Meter machen können, sagen:
00:34:12.02	IP_HB	Komm, hör mal, ich ... Komm, lass mal, ich helf dir mal eben. Nich.

	Topic	A1	A2	B1	B2
Topic Modeling	49	-	-	-	-
	22	Arbeitsvorgänge Montanindustrie	Arbeitspraxis im Bergbau	Arbeitspraxen im Bergwerk	Bergbau / Arbeits- bedingungen
	46	Bergbau	Bergbau / Arbeitsall- tag	Gegebenheiten im Kohlebergbau	Bergbau
manuelle Annota- tion		Motivation / Identifi- kation	Arbeitsverhältnisse im Betrieb	Leistung / Entloh- nung	Solidarität
		Gesundheit / Körper	Werte	Verkörperlichung / Körperlichkeit von Arbeit	Macht
		Kollegialität / Solida- rität	-	-	Arbeitsbedingungen
		Gender	-	-	-

Tab. 5: Topics und manuelle Kategorien, die in den Gruppen der Textstelle ADG0054, Chunk 85 zugeordnet wurden. Farblich identisch markierte Kategorien zeigen inhaltliche Übereinstimmung an.

Die Gewichtung der Topics bleibt auch in diesem dritten Chunk kontinuierlich: Noch immer geht es um den Bergbau (46) und die Arbeitsbedingungen (22) vor Ort. Die manuellen Annotationen entwickeln sich jedoch wiederum weiter – auch in den Gruppen, die zuerst Topic Modeling durchgeführt haben: ›Gesundheit / Körper‹ wird nun von Gruppe A1 angelegt, ›Werte‹ von Gruppe A2. Dieser Aspekt ist eng verbunden mit dem Konzept ›Solidarität‹ (und ›Kollegialität‹), annotiert von den Gruppen A1 und B2, was im Topic Modeling nicht abgebildet wurde – allenfalls das abgeschlagene Topic 35 liefert hier einen entsprechenden Hinweis.⁴⁹ Auch das Thema der Körperlichkeit wurde nun zusätzlich herausgestellt und von Gruppe A1 explizit in Zusammenhang mit Gesundheit gebracht (›Gesundheit / Körper‹). Nachdem im vorherigen Chunk das Thema ›Gesundheit‹ nur von Gruppe B2 gesetzt wurde, ergänzt dies nun A1, sodass die Reihenfolge des methodischen Zugangs auch hier offenbar nicht die entscheidende Rolle spielt. Ebenso wurde das Thema der Solidarität sowohl von Gruppe A1, die sich zunächst mit Topic Modeling auseinandersetzte, als auch von B2, die zunächst manuell annotierte, an diesem Chunk vermerkt. Das für dieses Chunk dominante Topic 49 wurde nicht als relevant für das Thema ›Arbeit‹ erkannt und entsprechend nicht gelabelt.⁵⁰

Wenn man sich die Labels für alle drei Chunks gesammelt anschaut, fallen die Differenzierungsgrade der manuellen Annotation zwischen den Gruppen weniger stark ins Gewicht. Es wäre eine weitere Untersuchung wert, inwiefern die sukzessive Differenzierung und Analyse in den Gruppen, die zuerst Topic Modeling durchgeführt haben, durch eben diese Vorgehensweise bedingt ist. Eine Hypothese ist, dass die über Topic-Listen vergebenen Labels zunächst auch im manuellen Annotieren aus der Erinnerung gerufen und angelegt werden und während der Lektüre – also Chunk für Chunk – um weitere Beobachtungen und analytische Feststellungen ergänzt werden.

In der Summe zeigt sich an diesem Close Reading, dass die manuelle Annotation feinteiliger ist als die Topic-Labels und deutlicher sowie schneller auf Veränderungen in der Erzählung reagiert. Obwohl auch dort eine rein technisch vorgegebene Textpassage (die Chunks) annotiert wurde, die in sich keine geschlossene Erzähleinheit bildet, registrierten alle Gruppen die Veränderung im Bericht von einer sachlichen Darstellung hin zu einer eigenen Einschätzung, die deutlich wertend ist. Dementsprechend verschieben sich die Kategorien weg von ›Arbeitsbedingungen‹ hin zu ›Solidarität‹ und ›Körperlichkeit‹. Letztlich zeigt sich, dass die menschlichen Annotationen tiefgehender sind. Nichtsdestotrotz bilden die computationell erzeugten und manuell gelabelten Topics einen sehr allgemeinen, aber überzeugend korrekten thematischen Überblick.

⁴⁹ Die ersten zehn Wörter von Topic 35 sind: kontakt, verhältnis, freunde, kollegen, familie, wohnen, nachbarn, helfen, leuten, nachbarschaft.

⁵⁰ Die ersten zehn Wörter von Topic 49 sind: gedenken, mensch, gott, angst, passieren, raus, wussten, lieb, reden, ruhig.

5.2 Vergleich der Annotationen in den Gruppen: Kategorienbildung in manueller und maschineller Annotation

Es kann festgehalten werden, dass die Gruppen, die erst nach der Beschäftigung mit den Topics an das Material gingen (A1 und A2), differenzierte und analytische Kategorien erst an späteren Chunks angelegt haben. Das korrespondiert mit der Feststellung, dass die Kategoriensets der beiden Gruppen kleiner ausgefallen sind als die Sets der Gruppen, die mit der manuellen Auswertung eingestiegen sind (B1 und B2). Zudem wurden von den Gruppen, die zuerst Topic Modeling angewendet haben, auch insgesamt weniger analytische Kategorien gebildet. Analytische Kategorien sind, bereits im offenen Annotieren, zentral für die weitergehende Interpretation der untersuchten Quellen.⁵¹ In der qualitativen Sozialforschung werden unterschiedliche Annotationspraxen realisiert, die mehr oder weniger analytische Anteile haben, etwa in der Grounded Theory⁵², der darauf aufbauenden wissenssoziologischen Diskursanalyse⁵³ oder der qualitativen Inhaltsanalyse⁵⁴. Gleichzeitig werden auch im qualitativen Annotieren immer beschreibende Kategorien gebildet. Die im Topic Modeling generierten Topics sind rein textimmanent, was zunächst unabhängig von der Art des Zugangs zum Text der Fall ist. Es liegt nahe, dass beschreibende Kategorien durch die Topics beeinflusst wurden: Da im Topic Modeling nur Begriffe aus den Texten selbst verwendet werden, bleiben diese beschreibend. Analytische und damit auch abstrahierende Darstellungen sind hier nicht vorgesehen und müssen von den Forschenden ergänzt werden. Was für die Gruppen ohne vorherige Beschäftigung mit Topic Modeling selbstverständlich in analytische Begriffe gefasst wurde, bleibt bei den Gruppen, die nach dem Topic Modeling annotierten, zunächst eher deskriptiv. Ein Zusammenhang mit dem ersten, beschreibenden Zugang ins Material ist sehr wahrscheinlich, sollte in weiteren Erhebungen allerdings vertieft untersucht werden. Gleichzeitig scheint die Bildung von analytischen Kategorien durch eine vorhergehende Beschäftigung mit Topic Modeling nicht ausgeschlossen zu werden, da die Gruppen im weiteren Verlauf ihrer manuellen Annotation dazu übergingen, ebenfalls analytische Kategorien zu bilden.

Die Anzahl der analytischen Kategorien und deren Abstraktionsgrad hängen zudem stark von der vorhergehenden Intensität der Beschäftigung mit dem Forschungsthema ab und sind auch je nach Forscher*in individuell unterschiedlich: Das Kontextwissen, insgesamt das individuelle Vorwissen der Annotierenden, hat hier einen großen Einfluss. Vergleichende Studien zu solchen qualitativ erstellten Kategoriensets bestehen nicht. Der Unterschied zum Topic Modeling liegt allerdings darin, dass kaum eine qualitative Annotation auf Dauer ohne analytische Kategorien arbeiten wird. Im Laufe der Beschäftigung mit dem Material und der parallelen Rezeption von Sekundärliteratur und theoretischen Konzepten werden die analytischen Perspektiven wachsen und sich auch in der (Re-)Strukturierung von Kategorien widerspiegeln. In der Grounded Theory zeigt sich dies in der Sättigung von Kategorien und Quellenkorpus.⁵⁵ Wenn eine qualitative Studie also gründlich am Material arbeitet und die entsprechenden Kontexte berücksichtigt, werden Unterschiede in der Annotation, die auf dem Vorwissen der forschenden Person beruhen, im Laufe der Forschung ausgeglichen werden, was zum in qualitativer Forschung üblichen Qualitätsstandard der intersubjektiven Überprüfbarkeit der Interpretation⁵⁶ führt.

Dieser Prozess ist auch auf einer weiteren Ebene sichtbar geworden. Die lokalisierte Statik des Topic Modeling, also die mangelhafte Abbildung zeitlicher Zusammenhänge und historischer Prozesse im Material, ist nicht allein dem statistischen Verfahren geschuldet. Denn in der manuellen Annotation gingen alle Gruppen zwar im Grundansatz chronologisch durch die Texte, sprangen jedoch in verschiedenem Ausmaß dann zurück, wenn sie eine neue Kategorie am Material bzw. in der Diskussion entwickelten und diese auch für vorhergehende Textstellen passend fanden. Auch wurden Kategorien im Laufe ihrer Verwendung immer wieder umgenutzt, also mit fortschreitender Annotation anders verwendet als zu Beginn. Dies zeigt die iterative, aber kontextbezogene Beschäftigung mit dem Material. Diese wird im Topic Modeling zwar auch, aber anders umgesetzt: Der nichtlineare Blick auf das Material ist dort durch die Bayes'sche Wahrscheinlichkeitsermittlung zwar bereits algorithmisch angelegt, jedoch wurden die auf Grundlage der Topic-Listen erstellten Schlagworte während der Lektüre und Annotation der Textpassagen nicht weiter iterativ verfeinert.⁵⁷ Und auch bei der manuellen Annotation innerhalb der Gruppen, die zunächst Topic Modeling angewendet hatten, lösten die Annotierenden sich erst sukzessive von den Schlagworten der zuvor gelabelten Topics. In der rein manuellen Annotation entstehen die Kategorien hingegen anlassbezogen und dynamisch direkt am Text und, wie bereits dargestellt, in ihrem Grad der Analytik vom Kontextwissen der Forschenden ebenso abhängig wie vom Stadium der Forschung selbst.

⁵¹ Vgl. Rapp 2017, S. 254.

⁵² Vgl. Charmaz 2014.

⁵³ Vgl. Keller 2011.

⁵⁴ Vgl. Mayring 2015.

⁵⁵ Vgl. Strübing 2014, S. 32; Morse 2007.

⁵⁶ Vgl. Strübing et al. 2018.

⁵⁷ Vgl. Blei 2012, S. 79–80.

Dabei ist die intersubjektive Nachvollziehbarkeit von Kategorien bei qualitativen Annotationen in der Praxis nicht immer gegeben, selbst wenn die Kategorien durch eine feste Gruppe erstellt und in dieser intersubjektiv ausgehandelt wurden. Ein Beispiel hierfür ist die ausführlich dokumentierte Herangehensweise bei der Entstehung der Kategoriensets zweier Gruppen. So hat B2 in der manuellen Annotation lange Kategorien diskutiert, bevor die konkrete Annotation vorgenommen wurde. Das ist eine mögliche Erklärung dafür, warum die Gruppe ein ausgefeiltes und eher abstrahierendes Kategoriensystem hat. B1 ist viel schneller in die Annotation eingestiegen, hat dementsprechend mehr annotiert und das Kategoriensystem »on the fly« erstellt. Da beide Gruppen erst manuell annotiert haben, liegt der Unterschied hier offensichtlich vielmehr in der jeweiligen Zusammenstellung der Forschendenpersönlichkeiten sowie in deren Erfahrungswissen zu Annotationspraxen als in der Reihenfolge der methodischen Annäherung an das Material begründet: Wer bisher stärker mit der Inhaltsanalyse gearbeitet hat, ist ein vorab erstelltes Kategoriensystem gewohnt, während nach Grounded Theory Kategorien erst im offenen Annotieren entstehen. Ob die unterschiedlichen Ergebnisse der Herangehensweisen nur oder primär dem Einfluss von Topic Modeling geschuldet sind, kann somit nicht in allen Fällen eindeutig belegt werden, zeigt aber unabhängig davon, wie sehr die individuelle Forschungspraxis die Annotation bzw. Entwicklung von Kategorien beeinflusst. Das belegt auch eine weitere Diskussion: In der Gruppe B1 ging es um die Frage, ob mit Nomen oder mit Verben kategorisiert werden sollte, da Verben stärker an die Perspektive der Praxistheorie⁵⁸ anknüpfen würden. An diesen und an anderen Diskussionsprozessen der Gruppen wird deutlich, dass individuelle Annotationspraxen weiterer Erforschung bedürfen, insbesondere wenn sie im Kontext von Digital Humanities modelliert werden sollen.

Ausgehend von diesen Beobachtungen scheint eine ergänzende, iterative Verzahnung von manueller und maschineller Annotation auf mehreren Ebenen ratsam. Angelehnt an den hermeneutischen Zirkel⁵⁹, können sich Weitblick und Einblick, Kontext und Inhalt, Allgemeines und Besonderes im Wechselspiel von Topic Modeling und händischer Annotation gegenseitig stützen – oder wie es eine Teilnehmerin formulierte: Es ermögliche das »Springen« zwischen Nähe und Distanz im Sinne eines Scalable Reading⁶⁰.

Insgesamt herrschte die Meinung vor, das Topic Modeling habe das manuelle Annotieren vorbereitet, indem es Anregungen geliefert (B2), erste Kenntnisse vermittelt (B1) und einen Überblick zu unbekanntem Material generiert (A1) habe. Festgehalten werden kann, dass es für die manuelle Annotation durchaus relevant ist, wann sich mit Topic Modeling befasst wird. Ausgehend von unseren Beobachtungen scheint es ratsam, ergänzend und schon zu einem frühen Zeitpunkt der Auseinandersetzung mit unbekanntem Korpora via Topic Modeling auch in erste konkrete Textstellen einzusteigen und diese einer textnahen, manuell-qualitativen Analyse zu unterziehen, also etwa offen zu annotieren. Danach kann dann wieder in iterativen Prozessen in die Auseinandersetzung mit den Topics eingestiegen werden, um etwa die weitere Auswahl von Passagen zu leiten. Die Workshopteilnehmenden bestätigten, dass das Oszillieren zwischen Topics und Textpassagen zur Qualität der Kategorien – und damit auch zur Evaluation der Topics während des Trainings – beiträgt, insbesondere aber die Analysekategorien bereichert. In Bezug auf die Reihenfolge der Beschäftigung wurde als zielführend angesehen, zunächst mit Topic Modeling und dann in der Folge mit manueller Annotation nach Grounded Theory zu arbeiten. Dabei werde aus einem unbekanntem Korpus eine Vorsortierung vorgenommen, die dann im zweiten Schritt Kategorienbildung evoziere. Anders als DiMaggio⁶¹ schlagen wir also vor, die computationalen Verfahren weniger zum Skalieren auf große Textengen, sondern iterativ in den qualitativen Forschungsprozess verschränkt zu nutzen. Manuelle und maschinelle Annotation können, so ein Ergebnis unserer Vergleichsstudie, also in gegenseitiger Ergänzung genutzt werden.

6. Limitationen des eigenen Ansatzes

Obwohl mit dem von uns durchgeführten experimentellen Vorgehen zahlreiche Erkenntnisse möglich wurden, fanden zeitgleich andere Aspekte weniger Beachtung und sollten in weiteren Forschungen weitergehend untersucht werden. Drei Bereiche sind dabei besonders zu berücksichtigen:

Ein zentraler Punkt hierbei ist die künstliche Begrenzung der manuellen Annotation. Diese zieht sich in der Regel über längere zusammenhängende Textpassagen und wurde im experimentellen Setting stark begrenzt. Alle Gruppen haben in der manuellen Annotation fast durchgängig mehrere Kategorien pro Chunk vergeben, was vielen Teilnehmenden dennoch nicht ausreichte, um den Pluralismus der Interviewpassagen abzubilden. Die Annotation der Chunks mit pauschal drei Kategorien bzw. Topics wurde als stark reduzierend bemängelt. Auch das Annotieren abgegrenzter Chunks, die keine hermeneutisch ermittelten Sinnabschnitte darstellen, wurde als »sehr unnatürliches« Setting für den qualitativen Analyseprozess wahrgenommen, wodurch auch quer

⁵⁸ Vgl. Reckwitz 2003; Hui et al. 2017.

⁵⁹ Vgl. Kurt / Herbrink 2019.

⁶⁰ Vgl. Müller 2020.

⁶¹ Vgl. DiMaggio 2015.

über die Chunk-Grenzen hinweg verlaufende Inhalte verloren gehen könnten. Dementsprechend sollte in weiteren Erhebungen die manuelle Annotation freier gestaltet werden, wengleich hierdurch die Vergleichbarkeit zur maschinellen Annotation eingeschränkt wird.

Die Zusammenarbeit in den Gruppen war ebenfalls stark begrenzt und begrenzend. Diese wurden nur für den Workshop zusammengestellt und verfügten so über kaum gemeinsames Vorwissen, sondern mussten dies im Zuge der Annotation im Workshop aushandeln. Teilweise führte dies zu besonders interessanten Ergebnissen gerade aufgrund der so dokumentierten Diskussionen, in anderen Fällen reichte die Zeit nicht für einen abschließenden Austausch. So blieben die manuellen Annotationen noch unvollständiger, als sie es in ihrer Grundanlage der theoretischen Sättigung ohnehin sind und im experimentellen Zeitkorsett bleiben müssen. Die Gruppenzusammensetzung in künftigen Workshops sollte deshalb intern homogener sein, denn die Heterogenität führte in einigen Fällen zu prägnanten Stellungen erfahrener Personen. In Verbindung mit dem gegebenen Zeitdruck konnten deshalb nicht alle Teilnehmenden ihre Einschätzungen vollumfänglich einbringen. Verbleibende Limitierungen unseres Konzepts könnten durch ergänzende Befragungen (im Sinne qualitativer Interviews⁶²) oder Walkthroughs⁶³ bestehender Annotationen ergänzt werden.

Für das Topic Modeling wurde auf ein im Vorfeld trainiertes Modell zurückgegriffen, sodass das Verfahren für die Teilnehmenden intransparenter blieb, als dies in Forschungen der Fall sein sollte. Durch die thematische Fokussierung wurden außerdem nicht alle generierten Topics betrachtet, in weiteren Workshops sollten jedoch konsequent alle Topics gelabelt werden, um potenzielle Querbezüge zum interessierenden Thema nicht aus dem Blick zu verlieren. Es ist außerdem zu prüfen, ob pro Topic mehrere Labels zugelassen werden sollten, um die teils starke Binnendifferenzierung besser abzudecken. Zudem wurde mit einem sehr großen und bereits gut aufbereiteten Korpus gearbeitet, wie es in der qualitativen Forschung nur selten vorhanden ist. Dieses wurde in seiner Komplexität durch die Vorauswahl von Textpassagen in Form von Chunks durch die Workshopleitenden zudem stark reduziert und fragmentiert. Dieses Vorgehen war pragmatisch bedingt, um eine kleine Auswahl für den Workshopaufbau treffen zu können. Da genau diese Auswahlprozesse allerdings zur Genese qualitativer Erkenntnisse gehören, sind hierzu weitere Explorationen auf Grundlage der im Workshop gewonnenen Ergebnisse notwendig. Darüber hinaus könnten auch die Topics selbst von Workshopteilnehmenden stärker auf ihre Konsistenz hin bewertet werden, um so bessere Aussagen darüber zu erhalten, inwiefern diese hilfreich sind. Auch könnte dies mit Metriken oder Maßzahlen, wie *Exclusivity* oder *Coherence*, verbunden werden, um diese qualitativ zu evaluieren und Empfehlungen auszusprechen.⁶⁴

Für künftige Workshops und andere Evaluationsschritte sollten Anpassungen der Zeitplanung sowie der Gruppenzusammensetzung vorgenommen werden. Besonders beschränkend war der notwendige Ausschnitt aus dem Material, der aufgrund der begrenzten Zeit notwendig war, aber weder für Ansätze der Grounded Theory und qualitativen Forschung noch für die Digital Humanities üblich ist. Künftig sollten deshalb größere Ausschnitte in längeren Zeitabschnitten vergleichend untersucht werden, auch eine Ergänzung um stärker befragende als begleitende Erhebungsmethoden ist zu prüfen. Die Heterogenität der Gruppenzusammensetzung muss ambivalent bewertet werden und bietet sich für weitere vergleichende Untersuchungen mit divergierenden Zusammensetzungen an. Schließlich wären in weiteren Workshops auch andere Verfahren für den Vergleich menschlicher und maschineller Annotationen interessant und die Ansätze könnten auf andere Korpora übertragen werden.

7. Fazit

Unser experimentelles Vorgehen zielte darauf ab, die Unterschiede und Gemeinsamkeiten von menschlicher und maschineller Annotation empirisch zu überprüfen und auszudifferenzieren. Dafür hat sich das Workshop-Setting in weiten Teilen als zielführend erwiesen. Die Annotationen der Gruppen und deren genaue Dokumentation mit ethnografischen Methoden haben umfangreiche Einblicke in manuelle und maschinelle Annotationen, deren Unterschiede und Gemeinsamkeiten ermöglicht. Gerade der Vergleich zweier konkreter verschiedener Zugänge war aufschlussreich und hat gezeigt, dass ein iteratives, komplementäres Zusammenspiel aus quantitativer Vorstrukturierung und qualitativer Feinanalyse neue Perspektiven auf große Datenbestände liefert.

Besonders relevant ist die Frage nach der sich verändernden Epistemologie und Sinnstiftung, die im Rahmen des Workshops ebenso wie in der Interpretation des dabei erzeugten Datenmaterials immer wieder deutlich zu Tage trat. Sie zeigt die grundsätzliche Veränderung und auch Verunsicherung im Umgang mit computationellen Zugängen auf. Auch wenn die Ergebnisse des Topic Modeling konsistent und »sinnvoll« erscheinen, ist es letztlich der Mensch, der den Wortlisten durch die

⁶² Vgl. Schmidt-Lauber 2007; Spiritova 2014.

⁶³ Vgl. Light et al. 2013; Amelang 2023.

⁶⁴ Vgl. Weston 2023.

Vergabe von Kategorien Sinn zuschreibt. Zur Erschließung digitaler bzw. digitalisierter Großbestände qualitativer Daten bietet sich entsprechend ein Mixed-Methods-Ansatz an, bei dem das Topic Modeling ein Angebot unterbreitet, das von den Forschenden interpretiert und zur Strukturierung genutzt werden kann.⁶⁵

Wo die Sinnstiftung anfängt, löste bereits im Workshop kontroverse Diskussionen aus, etwa um die Frage, ob bereits die Kondensierung des Korpus hin zu strukturierten Wortlisten ein Akt der Interpretation ist. Zwar sei dies kein Akt der Sinnstiftung, doch die Selektion der Quellen anhand computationeller Vorstrukturierung beeinflusse definitiv die Auswertung der Quellen. Diese und andere maschinelle Verarbeitungsschritte sind eng verbunden mit Fragen von Agency, denn durch die Nutzung von Topic Modeling geben die Forschenden Handlungsmacht ab.⁶⁶

Dennoch – oder gerade deshalb – korrespondiert Topic Modeling mit der Grounded Theory, da es ohne vorher gebildete Kategorien an Text herangeht und Cluster von Bedeutungen aus dem Text selbst herausarbeitet. Gleichzeitig kann die Grounded Theory helfen, die Ergebnisse maschineller Annotation qualitativ am Text zu überprüfen und zu kontrastieren. Daraus ergibt sich im besten Fall ein Wechselspiel gegenseitiger Bereicherung, was auch den Vorschlägen einer *Data Driven Science*⁶⁷ oder des »empirically driven theorizing«⁶⁸ entspricht: Mit Hilfe der Grounded Theory können – insbesondere im diskursiven Austausch innerhalb einer Gruppe – Sinnzusammenhänge im Detail und in der Tiefe herausgearbeitet werden. Die besondere Stärke des Ansatzes liegt in der Generierung analytischer Kategorien, welche im textimmanent bleibenden Topic Modeling kaum erzeugt werden. Erstaunlich ist jedoch, dass Miner et al. in ihrer Evaluation zum gegenteiligen Ergebnis kommen und aus den Topics abstraktere Kategorien bilden, als sie das in der qualitativen Annotation getan hatten.⁶⁹ Ob dies an den Unterschieden im methodischen Vorgehen liegt oder an den annotierenden Personen bzw. dem annotierten Quellenmaterial, wäre weiter zu prüfen. An anderer Stelle merken sie jedoch an, dass die Topics insbesondere mit konkreten Kategorien übereinstimmen, was wiederum unseren Erkenntnissen entspricht.⁷⁰ Zudem wird in der manuellen Annotation gründlicher und differenzierter annotiert, wie auch die Workshopeteilnehmenden selbst feststellten. Dies ist allerdings äußerst zeit- und auch personalaufwendig, was umso mehr gilt, wenn die dabei wirksamen Sinnstiftungsprozesse dokumentiert werden sollen, wie es für eine Operationalisierung in den DH notwendig wäre. Wenngleich die Dokumentation in der Grounded Theory (insbesondere in Form von Memos) umgesetzt wird, findet diese zudem auf einer abstrakten und bereits interpretierenden Ebene statt. Dies ist wiederum nicht übertragbar auf Dokumentationen zur Vorbereitung einer Operationalisierung in computationale Schritte, wie sie auch für Annotationen vorgeschlagen wird.⁷¹

Das Topic Modeling hingegen spielt seine Stärken in der rasanten Verschlagwortung ganzer Interviewkorpora aus – doch suggerieren konsistente Topics zunächst semantische Zusammenhänge, wo letztlich reine Statistik am Werk ist. Allein durch die Auseinandersetzung mit dem Modell können jedoch spezifische Perspektiven des Korpus ebenso herausgearbeitet werden wie systematische Gegenüberstellungen von verschiedenen Themenbereichen. Diese intensive Beschäftigung insbesondere mit der Topic-Term-Matrix kann für den Erkenntnisprozess produktiv genutzt werden und ist strukturell vergleichbar mit der Diskussion von Kategoriensets aus der manuellen Annotation. Die Vogelperspektive verhindert dabei allerdings die kontextualisierende Auseinandersetzung mit spezifischeren Textstellen. Eine eigene, wenn auch nur sehr ausschnittshafte Kenntnis der zugrunde liegenden Quellen durch Close Readings einzelner Textpassagen kann zu einem fokussierten Durchdringen des Modells maßgeblich beitragen. Durch die Verbindung der quantitativ-computationellen und der qualitativ-manuellen Perspektiven und deren expliziter ethnografischer Begleitung können die jeweiligen Vorteile kombiniert und die Nachteile – zumindest in Teilen – kompensiert werden. Als zielführend zeigt sich ein iteratives Oszillieren zwischen den beiden Verfahren im Sinne eines Mixed-Methods-Ansatzes. Dieses kann insbesondere von Scalable Readings getragen werden, indem ausgewählte Topics bis zu konkreten Textstellen zurückverfolgt werden und diese manuell annotiert und anschließend mit den Topics verglichen werden.

Mit diesem Vorgehen könnte auch einem grundlegenden Unterschied zu qualitativer Forschung begegnet werden: der unterschiedlichen Dynamik. Während die auf das Topic Modeling gestützten Kategorien nach der Berechnung statisch bleiben und an die Textstellen angebracht werden, liegt dem qualitativen Annotieren insbesondere auf Grundlage der Grounded Theory ein fortlaufender Prozess der Sinnstiftung in Auseinandersetzung mit dem Material zugrunde. Das bezieht sich auch auf die Methode selbst. Werden beim manuellen Annotieren entlang der Kategorien eigene Regeln sukzessive entwickelt, besteht die Interpretation der Topics auf dem Weg zum passenden Label zum guten Teil in der Suche nach den zugrundeliegenden (algorithmischen) Regeln und deren Nachvollzug. Topics behalten im Textverlauf eher eine statische Perspektive bei und passen sich Veränderungen im Erzählfluss nur träge an. Es ist weiter zu untersuchen, ob eine manuelle Annotation im Anschluss an

⁶⁵ Vergleichbar auch der Vorschlag bei Miner et al. 2023, S. 13.

⁶⁶ Vgl. Bächle 2015.

⁶⁷ Vgl. Kitchin 2014, S. 5.

⁶⁸ Salganik 2018, S. 61.

⁶⁹ Vgl. Miner et al. 2023, S. 7.

⁷⁰ Vgl. Miner et al. 2023, S. 13.

⁷¹ Vgl. etwa Rapp 2017.

die Sichtung und das Labeling entsprechender Topics ebenfalls statischer und damit reduzierter vorgenommen wird, was den Mehrwert einschränken würde. Insgesamt kann jedoch festgehalten werden, dass die manuellen Annotationen tiefergehender und analytischer sind. Die maschinellen Annotationen bilden dennoch einen zutreffenden Einstieg ins Material.

Unser experimentelles Setting und die daraus gewonnenen Erkenntnisse zu manueller und maschineller Annotation legen nahe, dass Mensch und Maschine im Team – im Sinne einer Mensch-Maschine-Relation – Mehrwerte für Forschungsperspektiven generieren und diese in Form von Mixed-Methods-Ansätzen weiter erprobt und umgesetzt werden sollten. Damit kann auch der Forderung nach vermehrter qualitativer Evaluation im Machine Learning entsprochen werden.⁷² Diese Verknüpfungen evozieren zudem eine epistemologische Reflexion der eigenen Erkenntnis, was gerade zur Methodenentwicklung sehr gewinnbringend genutzt werden kann.

⁷² Vgl. Dobson 2021.

Bibliografie

- Benedikt Adelmann / Lina Franken / Evelyn Gius / Katharina Krüger / Michael Vauth: Die Generierung von Wortfeldern und ihre Nutzung als Findeheuristik. Ein Erfahrungsbericht zum Wortfeld »medizinisches Personal«. In: Patrick Sahle (Hg.): 6. Tagung des Verbands Digital Humanities im deutschsprachigen Raum e. V. (DHd 2019, Frankfurt / Main und Mainz, 25.–29.03.2019) 2019, S. 114–116. DOI: [10.5281/zenodo.4622122](https://doi.org/10.5281/zenodo.4622122)
- Melanie Althage: Potenziale und Grenzen der Topic-Modellierung mit Latent Dirichlet Allocation für die Digital History. In: Karoline Dominika Döring / Stefan Haas / Mareike König / Jörg Wettlaufer (Hg.): Digital History. Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft. Berlin/Boston 2022, S. 255–277. DOI: [10.1515/9783110757101-014](https://doi.org/10.1515/9783110757101-014)
- Katrin Amelang: Wie Apps erforschen? Zum Zusammentreffen neuer Forschungsgegenstände und alter Methoden. In: Hamburger Journal für Kulturanthropologie 16 (2023), S. 11–28. [\[online\]](#)
- Peter Andorfer: Turing Test für das Topic Modeling. Von Menschen und Maschinen erstellte inhaltliche Analysen der Korrespondenz von Leo von Thun-Hohenstein im Vergleich. In: Zeitschrift für digitale Geisteswissenschaften 2 (2017). 25.04.2017. PDF. DOI: [10.17175/2017_002](https://doi.org/10.17175/2017_002)
- Melanie Andresen / Michael Vauth / Heike Zinsmeister: Modeling Ambiguity with Many Annotators and Self-Assessments. In: Proceedings of the 14th Linguistic Annotation Workshop. (LAW 14, Barcelona / online, 12.12.2020), S. 48–59. Stroudsburg, US-PA 2020. PDF. [\[online\]](#)
- Linde Apel: Oral History reloaded. Zur Zweitauswertung von mündlichen Quellen. In: Westfälische Forschungen 65 (2015), S. 243–254. PDF. [\[online\]](#)
- Thomas Christian Bächle: Mythos Algorithmus. Die Fabrikation des computerisierbaren Menschen. Wiesbaden 2015. [\[Nachweis im GVK\]](#)
- David Meir Blei: Probabilistic Topic Models. Surveying a Suite of Algorithms That Offer a Solution to Managing Large Document Archives. In: Communications of the ACM 55 (2012), H. 4, S. 77–84. DOI: [10.1145/2133806.2133826](https://doi.org/10.1145/2133806.2133826)
- Franz-Josef Brüggemeier: Leben vor Ort. Ruhrbergleute und Ruhrbergbau 1889–1919. München 1983. [\[Nachweis im GVK\]](#)
- Rebecca Carlson / Ruth Dorothea Eggel / Lina Franken / Sarah Thanner / Libuše Hannah Vepřek: Approaching Code as Process. Prototyping Ethnographic Methodologies. In: Kuckuck. Notizen zu Alltagskultur und Volkskunde (2021), H. 1, S. 13–17. [\[Nachweis im GVK\]](#)
- Kathy Charmaz: Constructing Grounded Theory. 2. Auflage. Los Angeles u. a. 2014. [\[Nachweis im GVK\]](#)
- Steve Cohen: Shifting Questions. New Paradigms for Oral History in a Digital World. In: The Oral History Review 40 (2013), H. 1, S. 154–167. DOI: [10.1093/ohr/ohr036](https://doi.org/10.1093/ohr/ohr036)
- Louise Corti: Smart Qualitative Data: Methods and Community Tools for Data Mark-Up (SQUAD) Full Research Report. ERSC End of Award Report, Reference No. RES-346-25-3019. Swindon 2007. PDF. [\[online\]](#)
- Magdalini Dargentas / Dominique Le Roux: Potentials and Limits of Secondary Analysis in a Specific Applied Context. The Case of EDF—Verbatim. In: Forum Qualitative Social Research 6 (2005), H. 1, S. 1.01.2005. DOI: [10.17169/fqs-6.1.505](https://doi.org/10.17169/fqs-6.1.505)
- Paul DiMaggio: Adapting Computational Text Analysis to Social Science (and Vice Versa). In: Big Data & Society 2 (2015), H. 2. PDF. DOI: [10.1177/2053951715602908](https://doi.org/10.1177/2053951715602908).
- James Dobson: Interpretable Outputs. Criteria for Machine Learning in the Humanities. In: Digital Humanities Quarterly 15 (2021), H. 2. [\[online\]](#)
- Wolfgang Dunkel / Heidemarie Hanekop / Nicole Mayer-Ahuja (Hg.): Blick zurück nach vorn. Sekundäranalysen zum Wandel von Arbeit nach dem Fordismus (= International Labour Studies, 25). Frankfurt / Main u. a. 2019. [\[Nachweis im GVK\]](#)
- Martin Fechner / Andreas Weiß: Einsatz von Topic Modeling in den Geschichtswissenschaften: Wissensbestände des 19. Jahrhunderts. In: Zeitschrift für digitale Geisteswissenschaften 2 (2017). 19.12.2017. PDF. DOI: [10.17175/2017_005](https://doi.org/10.17175/2017_005)
- Lina Franken: Kulturwissenschaftliches digitales Arbeiten. Qualitative Forschung als »digitale Handarbeit«. In: Berliner Blätter. Ethnographische und ethnologische Beiträge 82 (2020), S. 107–118. DOI: [10.18452/22125](https://doi.org/10.18452/22125)
- Lina Franken: Digitale Daten und Methoden als Erweiterung qualitativer Forschungsprozesse. Herausforderungen und Potenziale aus den Digital Humanities und Computational Social Sciences. In: Forum Qualitative Sozialforschung 23 (2022), H. 2. DOI: [10.17169/fqs-22.2.3818](https://doi.org/10.17169/fqs-22.2.3818)
- Lina Franken: Digitale Methoden für qualitative Forschung. Computationale Daten und Verfahren. Münster u. a. 2023. DOI: [10.36198/9783838559476](https://doi.org/10.36198/9783838559476)
- Lina Franken / Nils Egger / Luis Fischer / Katharina Lillich / Florian Schmid: Nachnutzung von Forschungsdaten für qualitative Forschungen. Text Mining als Ansatz zur Exploration transkribierter Interviews. In: Digital[ität] Ethnografieren. Forschungsmethoden für den digitalen Alltag. Kulturanthropologie Notizen 85 (2023), S. 188–222. DOI: [10.21248/ka-notizen.85.16](https://doi.org/10.21248/ka-notizen.85.16)
- Lina Franken / Gertraud Koch / Heike Zinsmeister: Annotationen als Instrument der Strukturierung. In: Julia Nantke / Frederik Schlupkothen (Hg.): Annotation in Scholarly Editions and Research. Berlin u. a. 2020, S. 89–108. DOI: [10.1515/9783110689112-005](https://doi.org/10.1515/9783110689112-005)
- Evelyn Gius / Dominik Gerstorfer / Mari Akazawa / Malte Meister (Hg.): CATMA. Computer Assisted Text Markup and Analysis. 2014. HTML. [\[online\]](#)
- Evelyn Gius / Jan Christoph Meister / Malte Meister / Marco Petris / Christian Bruck / Janina Jacke / Mareike Schumacher / Dominik Gerstorfer / Marie Flüh / Jan Horstmann: CATMA. Version 6.5.0 vom 06.04.2022. DOI: [10.5281/zenodo.6419805](https://doi.org/10.5281/zenodo.6419805)
- Barney Galland Glaser / Anselm Leonard Strauss: Grounded Theory. Strategien qualitativer Forschung. 3. Auflage. Bern 2010 [1967]. [\[Nachweis im GVK\]](#)
- Google (Hg.): Google Colaboratory. Letzter Zugriff: 23.04.2023. HTML. [\[online\]](#)
- Shawn Graham / Ian Milligan / Scott Weingart: Exploring Big Historical Data: The Historian's Macroscope. London 2016. [\[Nachweis im GVK\]](#)
- Tobias Hodel / Dennis Möbus / Ina Serif: Topic Modeling im Vergleich: Aufbereitung, Umsetzung und Interpretation unterschiedlicher historischer Textkorpora. In: Selin Gerlek / Sarah Kissler / Thorben Mämecke / Dennis Möbus (Hg.): Von Menschen und Maschinen – Mensch-Maschine-Interaktion in digitalen Kulturen. Hagen 2022, S. 181–205. DOI: [10.57813/20220623-153139-0](https://doi.org/10.57813/20220623-153139-0)
- Betina Hollstein / Jörg Strübing: Archivierung und Zugang zu Qualitativen Daten. In: Rat für Sozial- und Wirtschaftsdaten (Hg.): Archivierung und Zugang zu Qualitativen Daten (= RatSWD Working Paper, 267). Berlin 2018, S. 1–13. DOI: [10.17620/02671.35](https://doi.org/10.17620/02671.35)
- Judith A. Holton: The Coding Process and Its Challenges. In: Antony Bryant / Kathy Charmaz (Hg.): The SAGE Handbook of Grounded Theory. Los Angeles 2007, S. 265–289. [\[Nachweis im GVK\]](#)
- Ela Hornung: Die Rede des Anderen. Narrative Interviews versus psychoanalytische Interviews. Überlegungen zum Setting. In: BIOS. Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen 23 (2010), H. 1, S. 127–137. PDF. [\[online\]](#)
- Allison Hui / Theodore R. Schatzki / Elizabeth Shove (Hg.): The Nexus of Practices. Connections, Constellations and Practitioners. London u. a. 2017. [\[Nachweis im GVK\]](#)
- Lucas D. Introna: Algorithms, Governance, and Governmentality. On Governing Academic Writing. In: Science, Technology, & Human Values 41 (2016), H. 1, S. 17–49. DOI: [10.1177/0162243915587360](https://doi.org/10.1177/0162243915587360)
- Florian Jatton: The Constitution of Algorithms. Ground-Truthing, Programming, Formulating. Cambridge, US-MA 2021. [\[Nachweis im GVK\]](#)
- Reiner Keller: Wissenssoziologische Diskursanalyse. Grundlegung eines Forschungsprogramms. 3. Auflage. Wiesbaden 2011 [2005]. [\[Nachweis im GVK\]](#)
- Rob Kitchin: Big Data, New Epistemologies and Paradigm Shifts. In: Big Data & Society 1 (2014), H. 1, S. 1–12. 01.04.2014. DOI: [10.1177/2053951714528481](https://doi.org/10.1177/2053951714528481)

- Gertraud Koch / Lina Franken: Filtern als digitales Verfahren in der wissenssoziologischen Diskursanalyse. Potenziale und Herausforderungen der Automatisierung im Kontext der Grounded Theory. In: Samuel Breidenbach / Peter Klimczak / Christer Petersen (Hg.): Soziale Medien. Interdisziplinäre Zugänge zur Onlinekommunikation. Wiesbaden 2020, S. 121–138. [[Nachweis im GVK](#)]
- Ronald Kurt / Regine Herbrich: Sozialwissenschaftliche Hermeneutik und hermeneutische Wissenssoziologie. In: Nina Baur / Jörg Blasius (Hg.): Handbuch Methoden der empirischen Sozialforschung. Wiesbaden 2019, S. 545–564. DOI: 10.1007/978-3-531-18939-0_33
- Almut Leh: Vierzig Jahre Oral History in Deutschland. Beitrag zu einer Gegenwartsdiagnose von Zeitzeugenarchiven am Beispiel des Archivs »Deutsches Gedächtnis«. In: Westfälische Forschungen 65 (2015), S. 255–268. [[Nachweis im GVK](#)]
- Matthias Lemke / Gregor Wiedemann (Hg.): Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse. Wiesbaden 2016. DOI: 10.1007/978-3-658-07224-7
- Lora Bex Lempert: Asking Questions of the Data. Memo Writing in the Grounded Theory Tradition. In: Antony Bryant / Kathy Charmaz (Hg.): The SAGE Handbook of Grounded Theory. Los Angeles 2007, S. 245–264. [[Nachweis im GVK](#)]
- Ben Light / Jean Burgess / Stefanie Duguay: The Walkthrough Method. An Approach to the Study of Apps. In: New Media & Society 20 (2017), H. 3, S. 881–900. PDF. DOI: 10.1177/1461444816675438
- Mallet: MACHINE Learning for Language Toolkit. Letzter Zugriff: 23.04.2024. HTML. [[online](#)]
- Philipp Mayring: Qualitative Inhaltsanalyse. Grundlagen und Techniken. 12. Auflage. Weinheim u. a. 2015. [[Nachweis im GVK](#)]
- Adam S. Miner / Sheridan A. Stewart / Meghan C. Halley / Laura K. Nelson / Eleni Linos: Formally Comparing Topic Models and Human-Generated Qualitative Coding of Physician Mothers' Experiences of Workplace Discrimination. In: Big Data & Society 10 (2023). DOI: 10.1177/20539517221149106
- Dennis Möbus: Holleriths Vermächtnis – ein Beitrag zur Geschichte von Frauen in der EDV. Topic Modeling als Methode digitaler Sekundäranalyse lebensgeschichtlicher Interviews. In: BIOS. Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen 33 (2020), H. 1, S. 162–180. DOI: 10.3224/bios.v33i2.01
- Janice Morse: Sampling in Grounded Theory. In: Antony Bryant / Kathy Charmaz (Hg.): The SAGE Handbook of Grounded Theory. Los Angeles 2007, S. 229–244. [[Nachweis im GVK](#)]
- Martin Müller: Scalable Reading. 26.04.2020. HTML. [[online](#)]
- Laura K. Nelson: Computational Grounded Theory. A Methodological Framework. In: Sociological Methods & Research 49 (2020), S. 3–42. PDF. DOI: 10.1177/0049124117729703.
- Lutz Niethammer (Hg.): Lebensgeschichte und Sozialkultur im Ruhrgebiet 1930–1960. 3. Bde. Berlin u. a. 1983–1885. Bd. 1 (1983): »Die Jahre weiß man nicht, wo man die heute hinsetzen soll«. Faschismuserfahrungen im Ruhrgebiet. [[Nachweis im GVK](#)]
- Christian Papilloud / Alexander Hinneburg: Qualitative Textanalyse mit Topic-Modellen. Eine Einführung für Sozialwissenschaftler. Wiesbaden/Heidelberg 2018. [[Nachweis im GVK](#)]
- Andrea Rapp: Manuelle und automatische Annotation. In: Fotis Jannidis / Hubertus Kohle / Malte Rehbein (Hg.): Digital Humanities. Eine Einführung. Stuttgart 2017, S. 253–267. DOI: 10.1007/978-3-476-05446-3_18
- Andreas Reckwitz: Grundelemente einer Theorie sozialer Praktiken. Eine sozialtheoretische Perspektive. In: Zeitschrift für Soziologie 32 (2003), H. 4, S. 282–302. DOI: 10.1515/zfsoz-2003-0401
- Radim Rehurek (Hg.): GENESIM. Topic modelling for humans. Letzte Aktualisierung: 21.12.2022. HTML. [[online](#)]
- Matthew Jeffrey Salganik: Bit by Bit. Social Research in the Digital Age. Princeton u. a. 2018. [[online](#)]
- Brigitta Schmidt-Lauber: Das qualitative Interview oder: Die Kunst des Reden-Lassens. In: Silke Göttisch / Albrecht Lehmann (Hg.): Methoden der Volkskunde. Positionen, Quellen, Arbeitsweisen der Europäischen Ethnologie. 2. Auflage. Berlin 2007, S. 169–188. [[Nachweis im GVK](#)]
- Nick Seaver: Algorithms as Culture: Some Tactics for the Ethnography of Algorithmic Systems. In: Big Data & Society 4 (2017) H. 2, S. 1–12. DOI: 10.1177/2053951717738104
- Stephen M. Sloan: Swimming in the Exaflood. Oral History as Information in the Digital Age. In: Douglas Boyd / Mary Larson (Hg.): Oral History and Digital Humanities. Voice, Access, and Engagement. New York 2014, S. 175–186. [[Nachweis im GVK](#)]
- Marketa Spiritova: Narrative Interviews. In: Christine Bischoff / Karoline Oehme-Jüngling / Walter Leimgruber (Hg.): Methoden der Kulturanthropologie. Bern 2014, S. 117–130. [[Nachweis im GVK](#)]
- Jörg Strübing: Grounded Theory. Zur sozialtheoretischen und epistemologischen Fundierung eines pragmatistischen Forschungsstils. 3. Auflage. Wiesbaden 2014. DOI: 10.1007/978-3-531-19897-2
- Jörg Strübing / Stefan Hirschauer / Ruth Ayaß / Uwe Krähnke / Thomas Scheffer: Gütekriterien qualitativer Sozialforschung. Ein Diskussionsanstoß. In: Zeitschrift für Soziologie 47 (2018), H. 2, S. 83–100. DOI: 10.1515/zfsoz-2018-1006
- Klaus Tenfelde: Sozialgeschichte der Bergarbeiterschaft an der Ruhr im 19. Jahrhundert. 2. Auflage. Bonn 1981. [[Nachweis im GVK](#)]
- Harald Welzer: Das Interview als Artefakt. Zur Kritik der Zeitzeugenforschung. In: BIOS. Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen 13 (2000) H. 1, S. 51–63. [[Nachweis im GVK](#)]
- Sara J. Weston / Ian Shryock / Ryan Light / Phillip A. Fisher: Selecting the Number and Labels of Topics in Topic Modeling: A Tutorial. In: Advances in Methods and Practices in Psychological Science 6 (2023), H. 2, S. 1–13. DOI: 10.1177/25152459231160105

Abbildungs- und Tabellenverzeichnis

Tab. 1: Vergleich der in den Gruppen manuell erstellten Kategorien. Clusterung und Sortierung durch die Autor*innen. Größere thematische Gruppen innerhalb eines Clusters sind grün bzw. orange hinterlegt. Ob diese als synonym zueinander gefasst werden, ist Interpretationssache bzw. nicht eindeutig zu entscheiden. Gleichzeitig ist es zweifelsfrei subjektiv, wie hier Kategorien geclustert werden, was auch aufgrund der Erfahrungen und Kenntnisse der Autor*innen erfolgt ist. Dies war zudem durch das übergreifende thematische Erkenntnisinteresse an Arbeit bedingt, so dass andere Kategorien weniger präzise gefasst und auch in der Folge untersucht wurden.

Tab. 2: Von allen Gruppen gelabelte Topics. Grün hervorgehobene Label haben auf der Horizontalen eine hohe Übereinstimmung zwischen den Gruppen, blau hervorgehoben sind starke Unterschiede auf der Horizontalen zwischen den Gruppen. Kursiv gekennzeichnete Labels sind besonders differenziert. Gruppierung durch die Autor*innen, wobei die konkrete Gruppierung wiederum auch durch die Kenntnisse und Perspektiven der Autor*innen geprägt ist.

Abb. 1: Globale Heatmap mit der Verteilung aller Topics (X-Achse) über alle Interviews (Y-Achse). Zugrunde liegen jeweils die summierten Topic-Weights aller Chunks eines Interviews. [Grafik: Dennis Möbus 2024]

Abb. 2: Die Verlaufsmatrix zeigt die Topics anhand der ersten zehn Wörter (Y-Achse) im chronologischen Verlauf des Interviews, gegliedert nach Chunks (X-Achse). [Grafik: Dennis Möbus 2024]

Tab. 3: Topics und manuelle Kategorien, die in den Gruppen der Textstelle ADG0054, Chunk 83 zugeordnet wurden. Farblich identisch markierte Kategorien zeigen inhaltliche Übereinstimmung an. Eine Liste aller Topics, abgebildet anhand der Top-30-Keywords, ist auf [GitHub](#) einsehbar.

Tab. 4: Topics und manuelle Kategorien, die in den Gruppen der Textstelle ADG0054, Chunk 84 zugeordnet wurden. Farblich identisch markierte Kategorien zeigen inhaltliche Übereinstimmung an.

Tab. 5: Topics und manuelle Kategorien, die in den Gruppen der Textstelle ADG0054, Chunk 85 zugeordnet wurden. Farblich identisch markierte Kategorien zeigen inhaltliche Übereinstimmung an.