

Oral History auf dem Weg zu Big Data: menschliche und maschinelle Annotation lebensgeschichtlicher Interviews im Vergleich

Egger, Nils

nils.egger@soziologie.uni-muenchen.de
LMU München, Deutschland

Franken, Lina

lina.franken@soziologie.uni-muenchen.de
LMU München, Deutschland

Möbus, Dennis

dennis.moebus@fernuni-hagen.de
FernUniversität in Hagen, Deutschland

Schmid, Florian

florian.schmid@gsi.uni-muenchen.de
LMU München, Deutschland

Oral History sieht sich durch die Digitalisierung großen Veränderungen ausgesetzt. Angesichts der digitalen Aufnahmetechnologien stellte sich schon vor Jahren die Frage nach der Archivierung der „born digital“ Quellen. Diese Herausforderung bietet aber auch neue Chancen und Perspektiven: etwa die automatische Spracherkennung digitaler Audiosignale, computergestützte Transkription und die komfortable Suche in Online-Repositories (Leh 2015, Leh 2018, Gref/Köhler/Leh 2017). In den letzten Jahren konnten große Fortschritte im Bereich der automatischen Spracherkennung gemacht werden, weshalb auch die Zahl digitaler Transkripte deutlich angestiegen ist (Köhler et al. 2019). Mit Oral-History.Digital entsteht zurzeit das größte Portal zur Archivierung und Präsentation lebensgeschichtlicher Interviews in digitaler Form im deutschsprachigen Raum - damit sind auch die historische und qualitative Forschung auf dem Weg in die „Big Data“ (Graham et al. 2015).¹ Um die wachsende Quellenbasis erschließen zu können, bedarf es perspektivisch automatisierter Verfahren, die das klassisch hermeneutische Arbeiten ergänzen. Mit den Digital Humanities haben Verfahren maschinellen Lernens Einzug in die Oral History gehalten, um Texte systematisch inhaltlich zu analysieren.

Als Heuristik zur Erschließung großer Textkorpora hat sich etwa das Topic Modeling etabliert (Graham et al. 2015, Lemke/Wiedemann 2016, Adelman et al. 2019).

Mit diesem Verfahren werden über Wahrscheinlichkeitsrechnung Gruppen von Wörtern extrahiert, die miteinander in Zusammenhang stehen. Gut trainierte Topic Models ermöglichen zum Beispiel die Extraktion thematischer Zusammenhänge aus kompletten Sammlungen lebensgeschichtlicher Interviews und vermitteln einen ersten inhaltlichen Überblick (Hodel et al. 2022). Darüber hinaus können bei explorativer Durchsicht der Ergebnisse unerwartete Phänomene an die Oberfläche gespült werden, die in den vielschichtigen Interviews allzu leicht verschütt gehen (Möbus 2022). Gerade für Sekundäranalysen ist das Verfahren deshalb vielversprechend (Franken 2022).

Allerdings wird im Machine Learning noch zu selten qualitativ evaluiert, um die automatisch generierten Ergebnisse zu validieren (Dobson 2021). Des Weiteren mangelt es in den DH an systematischen Studien, die computationale und menschliche Inhaltserschließung vergleichen, auch wenn erste Ansätze bestehen (Andorfer 2017, Baumer 2017, Fechner/Weiß 2017, Andresen et al. 2020). Schließlich ist die Hemmschwelle zum Einstieg in die Verwendung digitaler Methoden bei qualitativ Forschenden besonders hoch (Franken 2020). Um diesen Desiderata und Vorbehalten zu begegnen, wurde das Potential des Topic Modeling für die Aufbereitung größerer Datenmengen nach Grounded Theory durch die Autor:innen systematisch getestet. In einem am Turing-Test orientierten Versuchsaufbau haben wir im Rahmen eines zweitägigen Workshops lebensgeschichtliche Interviews verschlagwortet - einmal klassisch qualitativ (also manuell), einmal auf Grundlage von Topic Modeling (also maschinell). Leitendes Erkenntnisinteresse des Workshops war, wie das maschinelle Verfahren qualitative Analysen bereichern kann, welche Unterschiede entstehen, wenn Interviewmaterialien durch ein Topic Modeling strukturiert gesichtet werden und wie sich Perspektiven auf (unbekannten) Text unterscheiden. Topic Modeling wurde als Verfahren gewählt, weil es einen intuitiven Zugang zu großen Textmengen ermöglicht und zudem mit Interviewtranskripten gut zurechtkommt (was viele computerlinguistische Verfahren aufgrund der abweichenden Satzstruktur bisher leider nur begrenzt tun). Es ging uns dabei nicht darum, Grounded Theory und Topic Modeling als vergleichbar zu setzen, sondern zu prüfen, wie sich die beiden Methoden ergänzen und wie sich Sinnstiftungsprozesse je nach Zugang ausgestalten. Die schlussendlichen Annotationen an den Transkripten wurden mit im Vorfeld vorbereiteten Projekten in Catma (Gius et al. 2022) durch die Gruppen umgesetzt.

Im Mittelpunkt des Workshops stand das konkrete Arbeiten am Textkorpora aus unterschiedlichen Perspektiven sowie die gemeinsame Reflexion dieser Perspektiven. Der Beitrag stellt die Ergebnisse dieses Experiments vor und diskutiert die Mehrwerte sowie weitere Anschlussmöglichkeiten. Dafür wurden insgesamt sechzehn Teilnehmende in vier interdisziplinäre Gruppen mit jeweils gemischten Kompetenzen und Vorwissen unterteilt, sodass beispielsweise Teilnehmende mit Erfahrung in qualitativer Forschung und Teilnehmende, die bereits mit Topic Modeling vertraut waren, zusammenarbeiteten. Insgesamt setzten sich die Gruppen sehr heterogen aus Studierenden, Promovierenden und Promovierten zusammen.

Gearbeitet wurde parallel in den Gruppen mit einem umfangreichen, digital erschlossenen Korpus von Transkripten lebensgeschichtlicher Interviews zur Sozialgeschichte des 20. Jahrhunderts, dem Bestand Lebensgeschichte und Sozialkultur im Ruhrgebiet (LUSIR). Dieser geht zurück auf das erste großangelegte Oral-History-Projekt in Deutschland, das von Lutz Niethammer und Alexander von Plato 1981 bis 1988 durchgeführt wurde (Niethammer 1983). Die Interviews sind transkribiert und mit Timecodes versehen und können über das Archiv "Deutsches Gedächtnis" online eingesehen und angehört werden.² Für das dem Experiment zugrunde liegende Sample wurden 166 Volltexte herangezogen. Durch die Laufzeit der Interviews von bis zu acht Stunden hat das Korpus einen Umfang von 3,7 Millionen Wörtern, wovon nach Stoppwordbereinigung gut 700.000 übrigbleiben. Das Topic Modeling wurde mit Mallet, allerdings mit Hilfe des Gensim-Wrappers in Python, umgesetzt, nutzt somit LDA mit Gibbs-Sampling als Inferenzalgorithmus. Nach der Evaluation verschiedener Modelle und umfangreichem Parametertuning konnte festgestellt werden, dass eine Aufteilung der Interviews in kürzere Einheiten (Chunks) zu 25 Sätzen inhaltlich konsistente und aussagekräftige Topics hervorbringt, als optimale Topic-Anzahl hat sich 50 herausgestellt (vgl. ausführlich: Hodel et al. 2022, 188f., 194f.).

Als Sample hatten die Organisator:innen im Vorfeld in einer Pilotstudie zahlreiche Interviewpassagen gesichtet, die mit arbeitsspezifischen Topics gelabelt waren, und ein Sample zusammengestellt. In zwei der vier Gruppen wurden drei ausgewählte Interviewpassagen zunächst manuell annotiert (Kategorienvergabe oder Codierung), während die beiden anderen Gruppen mit einer Erschließung der noch unbekannten Texte durch das vortrainierte Topic Modeling starteten. Danach tauschten die Gruppen die Rollen, um die Unterschiede systematisch vergleichen zu können. Als Fragestellung wurde das Thema „Arbeit“ in all seinen Facetten gesetzt, um vergleichbare Ergebnisse zu erhalten. Der exemplarische Zugang wurde gewählt, weil der arbeitsspezifische Wandel im 20. Jahrhundert gesellschaftlich relevant ist (Stichwort: vom Normalarbeitsverhältnis zu prekärer und/oder entgrenzter Arbeit). In Oral-History-Interviews finden sich fast durchgängig Aussagen zur persönlichen Arbeitsbiografie und Einschätzungen zum Wandel des eigenen Arbeitsumfeldes. Sie sind deshalb als Daten für die computationale Analyse zum Thema Arbeit besonders vielversprechend.

Die qualitative Annotation erfolgte auf Grundlage der Grounded Theory (Glaser/Strauss 2010 [1967]; Charmaz 2014). Diese ist als Analyseverfahren in der qualitativen Sozialforschung und den empirischen Kulturwissenschaften weit verbreitet und auch für die Erweiterung qualitativer Forschungsprozesse gut geeignet (Franken 2022). Sie ermöglicht ein induktives Vorgehen, das aus dem Material heraus Bedeutungen und Kontexte erschließt und besonders in offenen und selektiven Annotationen als Prozess (Franken/Koch/Zinsmeister 2020) Teil des Erkenntnisprozesses ist. Allerdings wurde das Methodensetting im Workshop nicht in seiner vollen Komplexität realisiert, sondern für den experimentellen Aufbau auf den Schritt des offenen Annotierens reduziert. Dabei werden Textabschnitte gelesen und aus dem hermeneutischen Sinnerschließen heraus Katego-

rien gebildet. Diese werden direkt an einzelne Textstellen vergeben, so dass das Kategoriensystem bei der Texterschließung nach und nach wächst und strukturiert wird (Holton 2007). Weitere Schritte, wie das theoretische Sampling, auf dessen Grundlage zentrale Quellen aus Korpora ausgewählt werden (Morse 2007), oder das anschließende selektive Annotieren, mit dem das Kategorienset je nach Erkenntnisinteresse wieder reduziert wird, wurden aufgrund der begrenzten Zeit im Workshop nicht realisiert. Auch eine die Analyse begleitende Verschriftlichung von Gedanken in Memos (Lempert 2007) wurde nicht umgesetzt, da eine umfassende Auswertung der im Vorfeld ausgewählten Textstellen nicht Ziel war, sondern der Vergleich der unterschiedlichen Texterschließungen im Mittelpunkt stand. Konkret war der Arbeitsauftrag an die Gruppenmitglieder, die Textstellen zu lesen und offene Kategorien zu vergeben, um die im Text enthaltenen Inhalte möglichst gut zu beschreiben. Im Anschluss an einzeln umgesetzte Annotationen vergaben die Gruppen dann in einer Diskussion gemeinsame Kategorien an den Textstellen, abstrahierten also von der individuellen Interpretation.

Die Analyse der Topic Models erfolgte zunächst auf globaler Ebene. Dazu standen in vorher vorbereiteten Jupyter-Notebooks verschiedene Funktionen zur Verfügung: Topiclisten mit variabler Keyword-Anzahl, Balkendiagramme, welche die Topic-Verteilung zeigen, und Heatmaps, die einerseits die globale Verteilung der Topics auf die Interviews, andererseits die Verteilung der Topics über die in Textpassagen zerteilten Interviews im zeitlichen Verlauf zeigen. Zunächst wurden die Ergebnisse des Topic Modelings - also je eine Wortliste zu jedem der fünfzig Topics - einzeln gesichtet und dann in der Gruppe diskutiert. Topics mit Bezug zum Thema "Arbeit" wurden anschließend mit einem Schlagwort gelabelt. So wurde dem Topic „'chef', 'büro', 'angestellt', 'abteilung', 'thysen', 'sekretärin', 'abteilungen', 'arbeit', 'herren', 'damen' [...]“ von einer Gruppe etwa die Kategorie "Anstellung/Verwaltung" zugewiesen, dem Topic „'betriebsrat', 'gewerkschaft', 'betrieb', 'kollegen', 'gewerkschaften', 'betriebsräte', 'vorsitzend', 'wählen', 'metall', 'belegschaft' [...]“ die Kategorie "Interessenvertretung/Betriebsperspektive". Anschließend wurden mit Hilfe des Jupyter-Notebooks stichprobenartig Interviewpassagen ausgegeben, die den relevanten Topics zugeordnet waren, um die Qualität der Topics qualitativ zu überprüfen. Im nächsten Schritt gab die Workshopleitung die vorher ausgewählten Interviewpassagen bekannt, um den Vergleich mit den qualitativ arbeitenden Gruppen sicherzustellen.³ Die Teilnehmenden sollten beurteilen, ob sie anhand ihrer in der freien Exploration gesammelten Eindrücke auch zu diesem Sample gelangt wären. Abschließend wurden die Oberbegriffe der stärksten Topics einer Textpassage (Chunk) in Catma an die Interviewtranskripte getagged. Die Vergabe der Topics für die jeweiligen Interviewpassagen konnten die Teilnehmenden ebenfalls dem Jupyter-Notebook entnehmen.

Insgesamt stehen aus den Gruppen acht verschiedene Schlagwort-Sets pro Chunk zum Vergleich zur Verfügung (vier Gruppen, pro Gruppe rein menschliche Schlagwörter sowie Oberbegriffe für die Topics). Im Ergebnis lässt sich feststellen, dass für ein Topic, also eine Wortliste, zwar unterschiedliche, aber dennoch vergleichbare Oberbegriffe gewählt wurden. Beispiels-

weise wurden für das bereits erwähnte Topic, für das eine Gruppe den Begriff „Anstellung/Verwaltung“ vergab, von den anderen Gruppen die Oberbegriffe „Positionen in einer Firma“, „Organisation von Arbeit“ und „Arbeitsorganisation“ festgelegt. In den rein menschlich vergebenen Schlagwörtern für eine Textstelle ergeben sich ebenfalls Unterschiede, die jedoch grundsätzlich auf ein vergleichbares Textverständnis schließen lassen. Die Unterschiede der gewählten Kategorien bestehen hier eher in der thematischen Schwerpunktsetzung. Während eine Gruppe „Technische Entwicklung“ benannte, wählte eine andere Gruppe „Wandel Arbeitstechnik“ für den gleichen Textabschnitt. Bei den durch eine Gruppe benannten „Arbeitskonflikte[n]“ annotierte eine andere Gruppe „Arbeitsbedingungen“, nahm also ebenfalls ähnliche Interpretationen vor, wenn auch in geringerer Deutlichkeit. In der Annotation wurden für die maschinengenerierten Topics wesentlich allgemeinere Kategorien vergeben, die menschliche Annotation erfolgte zielgenauer und hat in verschiedenen Dimensionen die den Forschenden bekannten Kontexte (etwa zu historischen Ereignissen) einbezogen.

Im Anschluss an die Arbeit in den Gruppen wurde in einer Abschlussdiskussion das unterschiedliche Vorgehen verglichen und diskutiert. Besonders sticht hervor, dass die Teilnehmenden - unabhängig von Qualifikationsstufe und Vorwissen - übereinstimmend der Meinung waren, dass Topic Modeling die qualitative, textnahe Arbeit vorbereiten und anleiten kann. Es ermöglicht, so eine Teilnehmerin, das „Springen“ zwischen Nähe und Distanz und damit einen anders informierten Umgang mit dem Quellenmaterial. Die Teilnehmenden waren sich einig, dass die aus dem Topic Modeling erzeugten Ergebnisse ihren Interpretationsvorgang angeregt haben. Das Verfahren wurde auch als „kreative Methode“ bezeichnet. Mit ihrer Einschätzung stimmen die Teilnehmenden damit bisherigen theoretisch-konzeptionellen Überlegungen (Jacobs/Tschötschel 2019; Nelson et al. 2018) zu.

Die Gruppen, die zuerst die Topic Models betrachtet hatten, gingen gezielter an den qualitativen Arbeitsschritt. Allerdings wurde mehrfach darauf hingewiesen, dass parallel oder im Anschluss „Rücksprache mit dem Text gehalten“ werden müsse. Denn allein aus den Mustern des Topic Modelings könne man sich nicht erschließen, was die Interviewpartner:innen gemeint haben. Zudem würden sich hierdurch nur beschreibende Kategorien entwickeln, die um analytische Kategorien ergänzt werden müssten, wie sie für den qualitativen Interpretationsprozess typisch sind. Die Notwendigkeit der Verbindung von qualitativen und maschinellen Schritten wurde also mehrfach betont. Wie es ein Teilnehmender zusammenfasste: „Topic Modeling ist eine Suchmaschine, bei der ich Parameter gut beeinflussen kann. Es eignet sich, um Themenkomplexe zu finden, die ich mir danach anschauen kann.“ Dennoch kann Topic Modeling als der Grounded Theory in vielen Punkten entsprechend verstanden werden, da es ohne vorher gebildete Kategorien an Text herangeht und Cluster von Bedeutungen aus dem Text selbst herausstellt. Es eignet sich also für induktives Vorgehen, wie Kitchin (2014, 5) es in seiner *data driven science* fordert und auch Salganik als *empirically driven theorizing* (2018, 61) vorschlägt.

Zur weiteren Bewertung der im Workshop generierten Schlagwörter wurde im Nachgang eine Befragung unter 13 Bachelor-Studierenden der Soziologie sowie mit einigen Monaten Abstand auch unter 10 der Teilnehmenden des Workshops selbst durchgeführt. Über die Schlagwortmengen und Chunks hinweg entstanden so 1.148 Bewertungen. Es sollte die Passung der acht verschiedenen Mengen an Schlagwörtern zu jedem Textabschnitt bewertet werden. Dazu lasen die Teilnehmenden sieben Textabschnitte und bewerteten die jeweiligen Schlagwörter auf einer Skala von eins bis fünf, wobei bei den Teilnehmenden des Workshops die Schlagwörter der eigenen Gruppe jeweils ausgelassen wurden. Mit Kontrolle der kategorialen Variablen (1) Gruppe im Workshop, (2) Befragungsgruppe (Bachelor-Studierende als Referenzkategorie) und (3) Chunk hatten die mithilfe des Topic Modelings erstellten Schlagwörter im multiplen linearen Regressionsmodell eine um 0,94 ($p < 0,001$) schlechtere Bewertung auf der Skala von 1-5 als die rein menschlichen Schlagwörter. Das heißt, die Schlagwörter des Topic Modeling wurden statistisch höchst signifikant als schlechter bewertet. Allerdings variiert der Performance-Unterschied je nach Chunk: Für den Textabschnitt mit der geringsten Differenz zwischen maschinengestützter und menschlicher Annotation beträgt der Unterschied lediglich 0,18 und ist statistisch nicht signifikant. Daran wird deutlich, dass je nach Kontext des Textes Topic Modeling sehr gute Ergebnisse der inhaltlichen Vorstrukturierung liefern kann.

Die nachgängige Befragung der Workshop-Teilnehmenden umfasste zudem eine Art Turing-Test, bei dem die Befragten bestimmen sollten, ob die ihnen präsentierten Schlagwörter jeweils durch manuelle Annotation oder über das Topic Modeling entstanden sind. Cramérs V zwischen tatsächlicher und zugeschriebener Annotationsform beträgt 0,28 und beschreibt damit einen moderaten Zusammenhang.

Selbst Personen, welche die unterschiedlichen Generierungsprozesse der Schlagwörter kennen, können bei separater Betrachtung also nicht mehr eindeutig auf den Generierungsprozess schließen. Die über das Topic Modeling generierten Schlagwörter besitzen für menschliche Betrachter folglich durchaus eine sinnhafte Qualität.

Als Fazit kann festgehalten werden, dass, wenig überraschend, epistemologische Unterschiede zwischen der computationalen und der manuellen Texterschließung bestehen. Gerade am Vergleich wurde jedoch sichtbar, und von den Teilnehmenden so diskutiert, dass in beiden Zugängen wir als Menschen es sind, die subjektiv Sinn zuschreiben und neu ordnen. Wie kontrovers das Thema „Sinnzuschreibung“ ist, zeigte sich in einer Diskussion während des Abschlussplenums: Während einige die Art der statistischen Kondensierung, wie sie im Topic Modeling durchgeführt wird, bereits als einen Akt der Interpretation auffassten, widersprachen andere vehement, dass eine Interpretation Verstehen - konkreter: Sinnverstehen - voraussetze, was bei computationalen Auswertungen nicht der Fall sei. Ein Indiz, das letzteres Argument stützt, ist die Tatsache, dass letztlich den rein deskriptiven Topics analytische Kategorien zugeordnet werden, um die Texte zu annotieren. Immerhin konnte die Auswertung des nachgelagerten Surveys zeigen, dass sich die direkt aus dem Text und die aus den Wortlisten generierten Kategorien nicht grundlegend unterschieden.

den, wenngleich die analytische Tiefe teils signifikant abwich. Zielführend im Sinne einer Erschließung digitaler Großbestände qualitativer Daten erscheint daher ein Mixed-Method-Approach, der die komplementären analytischen Zugänge kombiniert und etablierte Forschungsprozesse ergänzt.

Gleichzeitig hat sich die hohe Relevanz systematischer Evaluation digitaler Methoden gezeigt. Im Setup des Workshops haben sich die verschiedenen Ansätze gut ergänzt und zu einer vertieften Reflexion der Vor- und Nachteile der Zugänge geführt: So können mit Hilfe der Grounded Theory - insbesondere im diskursiven Austausch innerhalb einer Gruppe - Sinnstiftungsprozesse minutiös und akkurat herausgearbeitet werden. Das Topic Modeling hingegen spielte seine Stärken in der rasanten Verschlagwortung ganzer Interviewkorpora aus. Auf der anderen Seite ist das Arbeiten nach den Regeln der Grounded Theory äußerst zeitintensiv und konsistente Topics können Sinnstiftung suggerieren, wo letztlich reine Statistik am Werk ist.

Die Schwächen des Experiments liegen am Ende vor allem in der notwendigerweise gesetzten Ausschnitthaftigkeit des Materials. Um in der begrenzten Zeit vergleichbare Ergebnisse zu erzeugen, wurden nur Interviewausschnitte durch die Gruppen gelesen und annotiert. Das ist in der Grounded Theory unüblich, da das Ausschnitthafte den Blick auf das Material verzerrt. Auch für Ansätze in den Digital Humanities wird üblicherweise das gesamte Korpus analysiert. Für Folgeprojekte sollte ein größeres Zeitbudget eingeplant werden, um komplette Interviewtranskripte analysieren zu können.

Fußnoten

1. <https://www.oral-history.digital/> [03.08.2022]
2. <http://www.deutsches-gedaechtnis.fernuni-hagen.de> [03.08.2022]
3. https://github.com/moebusd/mensch_und_maschine.

Bibliographie

- Adelmann, Benedikt, Franken, Lina, Gius, Evelyn, Krüger, Katharina, Vauth, Michael.** 2019. "Die Generierung von Wortfeldern und ihre Nutzung als Findeheuristik. Ein Erfahrungsbericht zum Wortfeld 'medizinisches Personal'". 6. *Tagung des Verbands Digital Humanities im deutschsprachigen Raum e.V. (DHd 2019)*, hg. von Patrick Sahle. 114–116. DOI: 10.5281/zenodo.4622122.
- Andorfer, Peter.** 2017. "Turing Test für das Topic Modeling. Von Menschen und Maschinen erstellte inhaltliche Analysen der Korrespondenz von Leo von Thun-Hohenstein im Vergleich". *Zeitschrift für digitale Geisteswissenschaften* 2. DOI: 10.17175/2017_002.
- Andresen, Melanie, Vauth, Michael, Zinsmeister, Heike.** 2020. "Modeling Ambiguity with Many Annotators and Self-Assessments". *Proceedings of the 14th Linguistic Annotation Workshop*, 48–59. Barcelona. <https://aclanthology.org/2020.law-1.5/>.
- Baumer, Eric P. S., Mimno, David, Guha, Shion, Quan, Emily, Gay, Geri K.** 2017. "Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence?". *Journal of the Association for Information Science and Technology* 68 (6), 1397–1410. DOI: 10.1002/asi.23786.
- Charmaz, Kathy.** 2014. *Constructing Grounded Theory. Introducing Qualitative Methods*. Los Angeles u.a.: SAGE.
- Dobson, James.** 2021. "Interpretable Outputs. Criteria for Machine Learning in the Humanities." *Digital Humanities Quarterly* 15 (2). <http://www.digitalhumanities.org/dhq/vol/15/2/000555/000555.html>.
- Fechner, Martin, Weiß, Andres.** 2017. "Einsatz von Topic Modeling in den Geschichtswissenschaften: Wissensbestände des 19. Jahrhunderts". *Zeitschrift für digitale Geisteswissenschaften* 2. DOI: 10.17175/2017_005.
- Franken, Lina.** 2020. "Kulturwissenschaftliches digitales Arbeiten. Qualitative Forschung als 'digitale Handarbeit'?" *Berliner Blätter – Ethnographische und ethnologische Beiträge* 82: 107–118. DOI: 10.18452/22125.
- Franken, Lina.** 2022. "Digitale Daten und Methoden als Erweiterung qualitativer Forschungsprozesse. Herausforderungen und Potenziale aus den Digital Humanities und Computational Social Sciences". *Forum Qualitative Sozialforschung* 23 (1). DOI: 10.17169/fqs-22.2.3818.
- Franken, Lina, Koch, Gertraud, Zinsmeister, Heike.** 2020. "Annotationen als Instrument der Strukturierung". In *Annotations in Scholarly Editions and Research. Function, Differentiation, Systematization*, hg. von Julia Nantke und Frederik Schlupkoth, 89–108. Berlin/München/Boston. DOI: 10.1515/9783110689112-005.
- Gius, Evelyn, Meister, Jan Christoph, Meister, Malte, Petris, Marco, Bruck, Christian, Jacke, Janina, Schumacher, Mareike, Gerstorfer, Dominik, Flüh, Marie, Horstmann, Jan.** 2022. *CATMA 6 (Version 6.5)*. Zenodo. DOI: 10.5281/zenodo.1470118.
- Glaser, Barney G., Strauss, Anselm L.** 2010 [1967]. *Grounded Theory. Strategien qualitativer Forschung*. Bern: Huber.
- Graham, Shawn, Milligan, Ian, Weingart, Scott.** 2015. *Exploring Big Historical Data: The Historian's Macroscopic*. London: Imperial College Press.
- Gref, Michael, Köhler, Joachim, Leh, Almut.** 2017. "KA³. Weiterentwicklung von Sprachtechnologien im Kontext der Oral History." *BIOS* 30 (1+2). DOI: 10.3224/bios.v30i1-2.05.
- Hodel, Tobias, Möbus, Dennis, Serif, Ina.** 2022. "Topic Modeling im Vergleich: Aufbereitung, Umsetzung und Interpretation unterschiedlicher historischer Textkorpora". In *Von Menschen und Maschinen - Mensch-Maschine-Interaktion in digitalen Kulturen*, hg. von Selin Gerlek, Sarah Kissler, Thorben Mämecke und Dennis Möbus, 181–205. Hagen. DOI: 10.57813/20220623-153139-0.
- Holton, Judith A.** 2007. "The Coding Process and Its Challenges". In *The SAGE Handbook of Grounded Theory*, hg. von Antony Bryant und Kathy Charmaz, 265–289. Los Angeles. DOI: 10.4135/9781848607941.
- Jacobs, Thomas, Tschötschel, Robin.** 2019. "Topic Models Meet Discourse Analysis. A Quantitative Tool for a Qualitative Approach". *International Journal of Social Research Methodology* 22 (5), 469–485. DOI: 10.1080/13645579.2019.1576317.
- Kitchin, Rob.** 2014. "Big Data, New Epistemologies and Paradigm Shifts". *Big Data & Society* 1 (1), 1–12. DOI: 10.1177/2053951714528481.

Leh, Almut. 2015. "Vierzig Jahre Oral History in Deutschland. Beitrag zu einer Gegenwartsdiagnose von Zeitzeugenarchiven am Beispiel des Archivs 'Deutsches Gedächtnis'". *Westfälische Forschungen. Zeitschrift des LWL-Instituts für westfälische Regionalgeschichte* 65: 255-268.

Leh, Almut. 2018. "Zeitzeugenkonserven. Interviews für nachfolgende Forschergenerationen im Archiv 'Deutsches Gedächtnis'". *Archivar* 71 (2): 155-157.

Lemke, Matthias, Wiedemann, Gregor. 2016. *Text Mining in den Sozialwissenschaften: Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*. Wiesbaden: Springer VS.

Lempert, Lora Bex. 2007. "Asking Questions of the Data. Memo Writing in the Grounded Theory Tradition". In *The SAGE Handbook of Grounded Theory*, hg. von Antony Bryant und Katy Charmaz, 245-264. Los Angeles: SAGE.

Möbus, Dennis (2022): "Holleriths Vermächtnis – ein Beitrag zur Geschichte von Frauen in der EDV. Topic Modeling als Methode digitaler Sekundäranalyse lebensgeschichtlicher Interviews". *BIOS* 33 (1). DOI: 10.3224/bios.v33i2.01#.

Morse, Janice M. 2007. "Sampling in Grounded Theory". In *The SAGE Handbook of Grounded Theory*, hg. von Antony Bryant und Katy Charmaz, 229-244. Los Angeles: SAGE.

Nelson, Laura K., Burk, Derek, Knudsen, Marcel, McCall, Leslie. 2018. "The Future of Coding. A Comparison of Hand-Coding and Three Types of Computer-Assisted Text Analysis Methods". *Sociological Methods & Research* 18 (4). DOI: 10.1177/0049124118769114.

Niethammer, Lutz (Hg.). 1983. *Die Jahre weiß man nicht, wo man die heute hinsetzen soll'. Faschismuserfahrungen im Ruhrgebiet. Lebensgeschichte und Sozialkultur im Ruhrgebiet 1930-1960, Bd. 1*. Berlin, Bonn: Dietz.