

Von Heimat zu Beheimatung - Visuelle Exploration und Sekundäranalyse von lebensgeschichtlichen Interviewsammlungen

Bayerschmidt, Philipp

philipp.bayerschmidt@fernuni-hagen.de
FU Hagen, Deutschland

Die Digitalisierung in der Oral History eröffnet zahlreiche neue Möglichkeiten, die vielschichtigen Quellen einer vertieften digitalen Analyse im Sinne von Distant Reading zu unterziehen. Das Zusammenwachsen verschiedener Interviewsammlungen, etwa durch das Online-Portal Oral-History.Digital (OHD) (OHD, 2023), erlaubt das breit angelegte Vergleichen und Erforschen von lebensgeschichtlichen Interviews. In meinem Promotionsprojekt untersuche ich, wie Personen, die nach 1945 nach Deutschland migriert und entschieden haben, zu bleiben, Heimat definieren und inwiefern sich ihr Verständnis im Laufe ihres Lebens eventuell verschoben oder verändert hat (Egger, 2022; Mitzscherlich 2019; Oesterheld, 2022). Um ein so komplexes und oft auch implizit besprochenes Thema zu analysieren, reicht es nicht, die Interviews mit Volltextsuche oder Metadaten zu durchsuchen, besonders da einige Sammlungen noch nicht vollständig inhaltlich erschlossen sind und lediglich das Transkript vorhanden ist.

Zur Beantwortung der Forschungsfrage wird der gesamte zugängliche Interviewbestand der OHD-Plattform von 986¹ Interviews aus sieben verschiedenen Archiven in ein übergreifendes Transkript-Korpus zusammengeführt. Eine solche umfassende Anzahl an Interviews händisch auszuwerten und zum Thema Heimat zu analysieren, lässt sich zeitlich kaum bewerkstelligen. Deshalb möchte ich einen inhaltlichen Zugang zum Korpus über Topic Modeling mit LDA (Blei et al. 2003) ermöglichen. Da für jedes Topic die Gewichtung pro Textstelle berechnet wird, können über *distant reading* Zusammenhänge und Muster entdeckt werden (Hodel et al., 2022; König, 2017; Althage, 2020).

Für die Analyse habe ich eine Topic-Modeling Pipeline entwickelt, die für die OHD-Transkripte optimiert ist und so auch Nutzer*innen ohne Programmierkenntnisse erlauben soll, Topic Modeling für die OHD Interviews zu nutzen (Bayerschmidt, 2024a, im peer-review). Dies ist besonders im Kontext der Oral History-Forschung zentral, um eine breite Zugänglichkeit der Verfahren zu gewährleisten. Über eine intuitiv bedienbare Oberfläche können alle wichtigen Parameter definiert werden. Hierfür wurde eine eigene Datenstruktur entwickelt, die neben dem gesamten Korpus alle

relevanten Metadaten und Topic-Modeling Ergebnisse beinhaltet und als json-Datei abgespeichert werden kann.

Auf Grundlage dieser Dateistruktur wurde eine Evaluations- und Explorationsoberfläche in Form eines *Plotly Dashboards* entwickelt (Bayerschmidt, 2024b, im peer-review). Dieses Analysetool ist die Grundlage der Untersuchung, um die Zusammenhänge und Muster innerhalb des Korpus sowie innerhalb der Interviews über graphische Darstellungen sichtbar zu machen und so die Auffindbarkeit dieser erst zu ermöglichen.

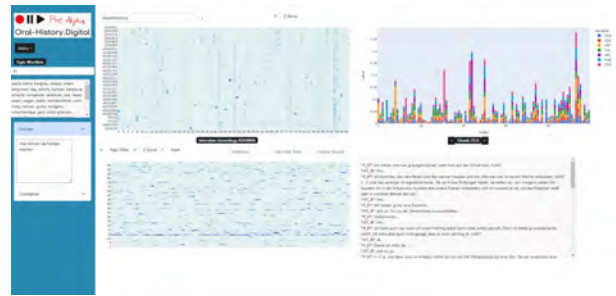


Abbildung 1: Startseite des OHTM-Dash

Um die quantitativen Ergebnisse (*distant reading*) mit den Dokumenten in Verbindungen setzen zu können, ermöglicht das *Plotly Dashboard* die jeweiligen Textstellen (*close reading*) hinter den Ergebnissen aufzurufen und diese so direkt mit in die Analyse einzubeziehen (*scalable reading*) (Müller, 2020). Im Gegensatz zu den gängigen Analysetools wie Voyant, AntConc, Catma oder das in der qualitativen Sozialforschung verbreitete MaxQDA, die umfangreiche Werkzeuge zur manuellen Annotation und Auswertung von Text- und Wortstatistiken bieten, liegt die Stärke des Topic-Modeling darin, automatisch Textstellen mithilfe der berechneten Topics zu annotieren. Dadurch können Inhalte von unbekannten Korpora erschlossen werden. Allerdings gibt es für die Analyse von Topic-Modeling Ergebnissen bisher keine Software-Lösungen, die die Ergebnisse direkt in Verbindungen mit den jeweiligen Textstellen setzt, weshalb eine eigene Lösung entwickelt wurde. Diese wird nachnutzbar zur Verfügung gestellt und hat somit einen eigenen Mehrwert.

Dabei möchte ich zeigen, dass die inhaltliche Arbeit mit Topic-Modeling Ergebnissen, angelehnt an die *Grounded Theory*, explorativ erfolgen kann, indem ein großer, nur zu Teilen bekannter Datensatz, in diesem Fall das OHD Korpus, anhand eines Forschungsinteresses bzw. einer Forschungsthematik untersucht wird, um Informationen zu extrahieren und so Muster und Zusammenhänge zu erkennen, die bei der Erarbeitung und zeitgleichen Beantwortung einer konkreten Fragestellung helfen können. Die *Grounded Theory* bietet ein qualitatives methodisches Vorgehen, um in einem iterativen Prozess unbekannte Daten induktiv zu analysieren und zu kategorisieren und durch weitere Verfeinerung dieser Schritte Zusammenhänge zu erkennen, um eine Fragestellung zu erarbeiten. Die Theorie wird dabei im Material, grounded, entwickelt (Charmaz 2014;

Strübing 2014), erste Annäherungen an computationale Verfahren bestehen (Baumer et al. 2017; Carlsen/Ralund 2022). Da die Berechnung eines Topic Models und die Analyse der Ergebnisse ebenfalls iterative und explorative Prozesse beinhaltet, kann eine Adaption hilfreich sein (Rosenthal, 2015; Przyborski/Wohlrab-Sahr, 2014).

Interessant ist dabei insbesondere auch, wie Interpretationsprozesse durch das Verfahren mitgestaltet werden und welche Verzerrungen in der Perspektive auf das Material sich hierdurch ergeben. Dies wird explorativ an einzelnen Textstellen aufgezeigt. Gleichzeitig ist dies bei der vorhandenen Menge Text nicht umfassend möglich, es werden allerdings andere Erkenntnisse ermöglicht: Entsprechend demonstriert das Promotionsprojekt, wie Topic Modeling nicht nur als Erschließungsmethode von Interviews genutzt werden kann, sondern durch die „distant“ Ergebnisse auch Erkenntnisse ermöglichen, die über eine analoge Auswertung nicht erzielt worden wären. In meinem Vortrag im Doctoral Consortium werde ich neben einem kurzen Überblick über mein Promotionsprojekt genauer auf das von mir entwickelte Explorationstool eingehen. Insbesondere sollen während einer Live-Präsentation die Funktionalitäten anhand des konkreten Forschungsinteresses zum Beheimatungsprozess exemplarisch demonstriert werden. So werden die Toolentwicklung und die inhaltliche Erkenntnisgenese ineinander verschränkt vorgestellt.

Fußnoten

1. Die Zahl bezieht sich nicht auf den tatsächlichen Gesamtbestand von OHD sondern auf die Interviews, deren Transkripte für das Forschungsprojekt bisher freigegeben worden sind.

Bibliographie

- Althage, Melanie.** 2020. Promotionsprojekt HU Berlin. abgerufen unter: <https://www.geschichte.hu-berlin.de/de/bereiche-und-lehrstuehle/digital-history/forschung/mining-the-historians-web>. Stand: 23.07.2024.
- 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? In: *Journal of the Association for Information Science and Technology* 68/6, S. 1397–1410.
- Bayerschmidt, Philipp.** 2024a. OHTM, an Oral History Topic Modeling Pipeline. <https://github.com/bayerschphi/ohtm.git>. Stand: 17.07.2024. Die Pipeline und das Vorgehen werden in kürze in einem Aufsatz veröffentlicht. Im peer-review im *Journal of digital History*.
- Bayerschmidt, Philipp.** 2024b. OHTM-Dash, a Visual Topic Modeling Exploration, [git@github.com:bayerschphi/ohtm_dash.git](https://github.com/bayerschphi/ohtm_dash.git). Stand: 17.07.2024. Die Pipeline und das Vorgehen werden in kürze in einem Aufsatz veröffentlicht. Im peer-review im *Journal of Open Humanities Data*.
- Blei, David M./Andrew Y. Ng/Michael I. Jordan.** 2003. Latent Dirichlet Allocation. In: *Journal of Machine Learning Research* 3. Online: <https://dl.acm.org/doi/10.5555/944919.944937>. Abgerufen am 23.07.2024
- Carlsen, Hjalmar Bang/Ralund, Snorre.** 2022. Computational Grounded Theory Revisited. From computer-led to computer-assisted Text Analysis. In: *Big Data & Society* 9/1
- Charmaz, Kathy.** 2014: *Constructing Grounded Theory. Introducing Qualitative Methods*. Los Angeles u.a.
- Egger, Simone.** 2022. Heimat, in: Schmidt-Lauber, Brigitta/Liebig, Manuel (Hg.): *Begriffe der Gegenwart. Ein kulturwissenschaftliches Glossar*. Wien, Köln.
- König, Mareike.** 2017. Digitale Methoden in der Geschichtswissenschaft. Definitionen, Anwendungen, Herausforderungen, in: *BIOS – Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen*. 1-2. S. 7–21
- Martin Müller.** 2020. Scalable Reading. <https://scalablereading.northwestern.edu>. Abgerufen am 23.07.2024
- Mitzscherlich, Beate.** 2019. Heimat als subjektive Konstruktion. Beheimatung als aktiver Prozess, in: Costadura, Edoardo/ Ries, Klaus/ Wiesenfeldt, Christiane (Hg.): *Heimat global. Modelle, Praxen und Medien der Heimatkonstruktion* (Edition Kulturwissenschaft Band 188). Bielefeld.
- Oesterheldt, Anja.** 2022. *Geschichte der Heimat. Zur Genese Ihrer Semantik in Literatur, Religion, Recht und Wissenschaft* (Studien und Texte Zur Sozialgeschichte der Literatur Ser v.157). Berlin/Boston.
- OHD.** 2023. <https://www.oral-history.digital/>, abgerufen am 16.07.2022.
- Przyborski, Aglaja/ Wohlrab-Sahr, Monika.** 2014. *Qualitative Sozialforschung. Ein Arbeitsbuch* (Lehr- und Handbücher der Soziologie). München.
- Rosenthal, Gabriele.** 2015. *Interpretative Sozialforschung. Eine Einführung* (Grundlagentexte Soziologie). Weinheim.
- Strübing, Jörg.** 2014. *Grounded Theory. Zur sozialtheoretischen und epistemologischen Fundierung eines pragmatistischen Forschungsstils*.
- Tobias Hodel / Dennis Möbus / Ina Serif:** Topic Modeling im Vergleich. 2022. 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. S. 181–205. DOI: 10.57813/20220623-153139-0