

Topic Modeling und interpretative Erzählanalyse

Michael Corsten, Ulrich Heid, Patrick Kahle, Fritz Kliche

Beitrag zur Ad-hoc-Gruppe »Maschinelles Lernen, Natürliche Sprachverarbeitung und Qualitative Methoden«

Vorüberlegungen: Topic Modeling, Qualitative Textanalyse, Narrationen

Erster Gedanke

Topic Modeling als eine Anwendung von Natural Language Processing (NLP) wird bereits erfolgreich zur Ergänzung und Überprüfung von qualitativen Textanalysen eingesetzt. Instrumente der maschinellen Verarbeitung natürlicher Sprache gehören mittlerweile zur standardmäßigen Ausstattung in qualitativer Textanalyse-Software wie MAXQDA oder ATLAS.ti. Beide Programme integrieren darüber hinaus auch KI-Werkzeuge auf der Basis von Large Language Models (LLMs) und bieten automatisierte Lösungen für Textzusammenfassungen, Paraphrasen bis hin zu Kodierungen und Interpretationen an.

Anders gefragt: Warum sollte es einen Grund geben, maschinelles Lernen und computerisierte Verarbeitung von natürlicher Sprache nicht auf qualitativ erhobene Textdaten anzuwenden? Denn erstens sind die NLP-Werkzeuge maschinelle Verfahren zur Analyse natürlicher Sprachdaten, d. h. sie sollen gerade die Zielsetzung erfüllen, Sprache, die in alltäglichen Kontexten erzeugt worden ist, möglichst effizient zu untersuchen, d. h. vor allem große Mengen sprachlicher Daten zu verarbeiten. Zweitens sind auch die Verfahren der Qualitativen Textanalyse zu einem nicht unbedeutenden Teil darauf ausgerichtet, große Textmengen (aus Interviews, Zeitungsartikeln, Gruppendiskussionen, Popsongs) zu kategorisieren, oder wie es in entsprechenden Methodenparadigmen lautet: zu „kodieren“. Und dabei unterscheidet sich drittens das Kodieren (im Sinne der Grounded Theory Methodology nach Anselm Strauss oder der Qualitativen Inhaltsanalyse nach Philipp Mayring) in methodischer Hinsicht nicht grundlegend von dem, was im Rahmen sogenannter „Text-Mining-Verfahren“ (Biemann et al. 2022) als ‚humanes‘ (oder nicht-maschinelles) Annotieren von Texten verstanden wird. Viertens ist auch in den maschinellen Verfahren der Massentextanalyse der Wechsel von (quantitativem) distant reading zu (qualitativem) close reading ein Prüfverfahren für die Güte der Analyse (Wiedemann und Lemke 2016).

Die Kompatibilität von computergestützter Verarbeitung von großen Mengen natürlich produzierter Sprache (Textdaten) mit qualitativer Textanalyse lässt sich auch aus einem anderen Paradigma herleiten. So weisen Verfahren der maschinellen Sprachverarbeitung nicht zufällig Ähnlichkeiten zur strukturalen Sprachanalyse (De Saussure 1995 [1916]) auf, sondern knüpfen in ihren Ursprüngen (Harris 1954) unmittelbar daran an. Und nicht zuletzt entsprechen auch die grundlegenden methodischen Hinweise zur poststrukturalen Diskursanalyse bei Foucault, mit der Anweisung, die „description des énoncés“ an

den Kriterien der „rareté, extériorité, cumul“ (Foucault 1969, S. 155ff.) auszurichten, einer statistischen Idee, die dem maschinellen bzw. statistischen Lernen sehr verwandt erscheint.

Zweiter Gedanke

Wie kommt es denn (trotz der im ersten Gedanken bemerkten Nähe) zu einer Skepsis gegenüber dem Einsatz maschineller Verarbeitung natürlicher Sprache? Der zweite Gedanke bezieht sich auf die Beobachtung, dass sich qualitative Textanalysen einerseits auf „Was-Fragen“ und andererseits auf „Wie-Fragen“ beziehen können. Der Einsatz von maschineller Textverarbeitung und von Topic Modeling im Besonderen ist bei der Beantwortung von „Was-Fragen“ überzeugend.

Wie sieht es mit der Beantwortung von „Wie-Fragen“ aus?

Wie-Fragen beziehen sich *nicht* auf Inhalte, *sondern* auf Formen, Modalitäten, Operative Modi, Gestaltungsweise, Ausdruck oder Artikulationsstil und nicht zuletzt auf die Komposition. Die Unterscheidung von „Was-Fragen“ und „Wie-Fragen“ verweist insofern auf eine paradigmatische Differenz der in den Ausrichtungen der Fragen vorausgesetzten Sprachtheorien, genauer auf den womöglich grundlegenden Unterschied zwischen strukturaler Semantik und realer Pragmatik als Ausgangspunkt eines Verständnisses von Sprache und Sprechen (s. Abbildung 1).

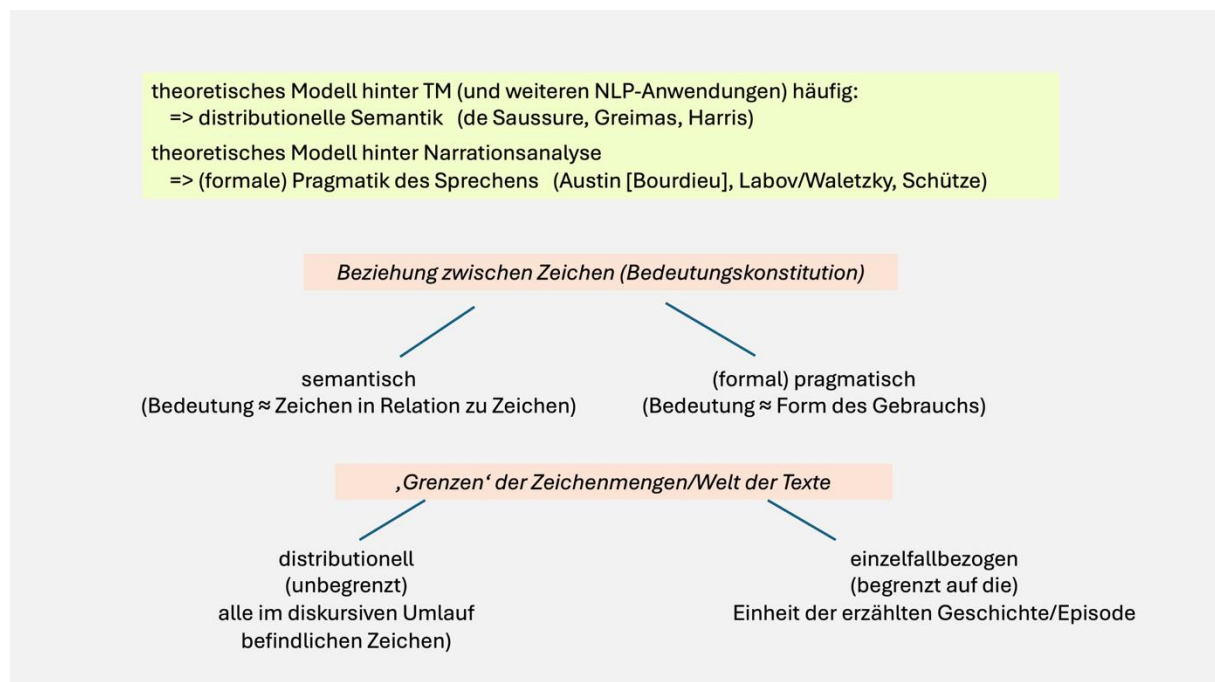


Abbildung 1: „Paradigmatischer Sprung“

Die Strömung der Qualitativen Forschung, die sich vornehmlich auf die Beantwortung von Wie-Fragen konzentriert, wird oft auch als „Interpretative Analyse“ oder „Interpretationsmethode“ bezeichnet. Das Interpretieren richtet sich dabei auf das Sprechen als ein Tun und bezieht sich dabei auf die Muster des Hervorbringens oder Verwendens von Zeichen. Von John L. Austin (1962) stammt die Wendung eines „Tuns mit Worten“ und ihm geht es primär um das „Wie“ dieses Tuns. Das strukturelle Verständnis von Sprache (als Semantik oder Syntax) ist dagegen „distributionell“. Es fragt danach, wie sich einzelne Zeichenelemente in einer größeren Menge von Texten verteilen, oder ob sich solche Verteilungen von Elementen beim Vergleich von unterschiedlichen Texten (als Menge von Texten oder Zeichen) unterscheiden oder ähneln.

Den beiden Ansätzen zum Verständnis der Sprache liegt eine unterschiedliche Bestimmung von Bedeutung zugrunde. In dem einen Fall ergibt sich die Bedeutung aus dem (distributionellen) Verhältnis von Zeichen in Relation zu anderen Zeichen; im anderen Fall ergibt sich die Bedeutung eines Zeichens aus seinem Gebrauch im Sprachkontext, also den Situationen und Konstellationen des Sprechens. Hinzu kommt, dass auch die Mengen der Zeichen, die in einer Analyse zu Rate gezogen werden, unterschiedlich begrenzt sind und jeweils verschiedene ‚Welten‘ von Texten als Kontext voraussetzen. Im Fall der maschinellen Verarbeitung von Sprachdaten ist es die quantitativ unabgeschlossene (im Prinzip unendliche) Menge von Zeichen, die sich im diskursiven Umlauf befinden. Prinzipiell unabgeschlossen bzw. unendlich ist dieser Raum, da immer neue Sprachdaten hinzukommen können, und zwar nicht nur bezogen auf zukünftige Textproduktion, sondern auch aufgrund des Auffindens vergangener Sprachdaten (archäologische Funde) oder bisher unbekannter Sprachkontexte oder Textdatensammlungen. Zwar wird im Rahmen der maschinellen Bearbeitung dann in der Regel ein Textkorpus begrenzt, aber diese Begrenzung ist durch den Rahmen der Untersuchung (oder der Gegenstandsbestimmung) bedingt und nicht prinzipiell durch die Endlichkeit von Texten. Demgegenüber ist im pragmatischen Verständnis die Verwendung von Zeichen stets situationsbezogen. Die Situation als Ausgangspunkt der Textgenerierung wird dabei als maßgeblicher Bezugsraum der Zeichenverwendung und des damit einhergehenden Diskurshorizonts verstanden. Dementsprechend verfährt die interpretative Forschung fallbezogen. Und fallbezogen meint dabei in der Regel die Referenz auf eine spezifische Situation als Fall (und nicht mehr).

Diese Bestimmungen der interpretativen Analyse treffen auch auf den von uns im Besonderen untersuchten Gegenstand der autobiographischen Erzählungen im Interview zu. Das bedeutet: Die soziologische Narrationsanalyse autobiographischer Texte (Fischer-Rosenthal und Rosenthal 1997) bezieht sich zunächst auf das Stegreiferzählen von Personen im je konkreten Fall einer Interviewsituation. Hierbei handelt es sich zudem um eine spezifische kommunikative Gattung, genauer eine bestimmte Gesprächsform in der Konstellation bzw. im Rahmen eines sozialwissenschaftlichen Interviews. Insofern haben wir es mit einem Forschungsproblem zu tun, das in der Kombination von Befunden besteht. Anhand von Interpretationen eines Sprechhandelns werden auf Einzelfallbasis Klassifizierungen unternommen und mit Textanalysen (des gleichen Textmaterials) verbunden, die anhand der maschinellen Verarbeitung natürlicher Sprache generiert wurden. Der Gewinn dieser methodologisch zunächst nicht naheliegenden Kombination könnte im experimentellen Nachweis bestehen, dass trotz der Differenzen in den zugrundeliegenden Sprachanalyseansätzen die Befunde aus der maschinellen Verarbeitung natürlicher Sprache auch für eine Untersuchung pragmatischer Muster des Sprechens, hier: Erzählens, fruchtbar gemacht werden können. Dies erscheint auch deshalb sinnvoll, weil die Erhebungsmethode des biographisch-narrativen Interviews in der Erziehungswissenschaft, der Geschichtswissenschaft und den Sozialwissenschaften relativ beliebt ist.

Frage: Lassen sich auf der Basis des methodologischen Paradigmas der „distributionellen Semantik“, die auf die Detektion von Themensträngen ausgerichtet ist (Biemann et al. 2022, S. 270ff.), Ergebnisse finden, die für eine formal-pragmatische Analyse von Narrationen instruktiv sein könnten?

Antwort: (1) Ja, besonders zur Ermittlung des ‚Was wird erzählt?‘ oder ‚Worüber wird gesprochen?‘ (2) Ja, über Umwege kann distributionelle Semantik formale Strukturelemente detektieren. (3) Ja, auf distributioneller Semantik basierende Techniken lassen sich korrespondierend im Rahmen von Mixed-Methods-Analysen (zur Überprüfung, zur Ergänzung u. ä.) einsetzen.

Interpretative Erzählanalyse – konkret exemplifiziert

Wie sehen nun die Aufgaben einer interpretativen Erzählanalyse konkret aus und wie könnten sie auf hinreichend vergleichbare Weise in Form maschineller Verarbeitung ‚natürlicher‘ Erzähldaten gelöst werden? Wir erläutern dazu erstens wichtige Elemente der Bauweise des Erzählens und illustrieren anhand eines Fallbeispiels die Kennzeichen des Erzählens im sozialwissenschaftlichen Interview.

Bauformen des Erzählens

Die soziologische Erzählanalyse ist in ihrer relativ jungen Geschichte von zwei soziolinguistischen Ansätzen stark beeinflusst: erstens von einem Beitrag von William Labov und Joshua Waletzky (1967), die mündliche Erzählungen von Konfliktgeschichten untersuchen, und zweitens von den Arbeiten von Fritz Schütze, die sich mit den „kognitiven Strukturen des Stegreiferzählens“ systematisch auseinandergesetzt haben. Beide Ansätze machen grundlegende Annahmen zur Erzählorganisation, die den Aufbau von Untereinheiten einer Erzählung (Segmente) und die Form von Erzählsätzen (Propositionen) betreffen. Diese Annahmen dienen der (konstitutionslogischen) Abgrenzung des Erzählens von anderen Formen des Sprechens bzw. Darstellens (Kallmeyer und Schütze 1976).

Demzufolge ist Erzählen in seiner Bauweise kein singulärer Sprechakt, sondern eine darstellende Tätigkeit, die zeitlich aus langen Redebeiträgen besteht und sachlich aneinander anschließende Äußerungsketten enthält, die sozial eher monologisch als dialogisch erzeugt sind und kompositorisch aus Segmenten (Rede- bzw. Erzählteile) und Einzeläußerungen (Aussagen, Sätze, Propositionen) bestehen. In Übereinstimmung mit entsprechenden Ansätzen aus Literatur- und Sprachwissenschaften (Lämmert 1955; Stanzel 1982; Weber 1998; Boyd et al. 2020, S. 3) lässt sich Erzählen als „*serielle Rede von zeitlich bestimmten Sachverhalten*“ (Weber 1998, S. 11) bestimmen, in der (1) erzählend-berichtende Rede konstitutiv ist, während Formen der (2) wörtlichen Redewiedergabe, der (3) Rede zur Sache oder der (4) Rede zur Rede zwar vorkommen können, aber fakultativ sind. In Weiterführung der poetologischen und literaturwissenschaftlichen Tradition wird die Bauweise des Erzählens als eine Abfolge von Teileinheiten begriffen, die der dramaturgischen Gestaltung dienen, und zwar als (1) Intro, als (2) Orientierung, als (3) Komplikation, (4) Auflösung und (5) Koda (s. Abbildung 2).

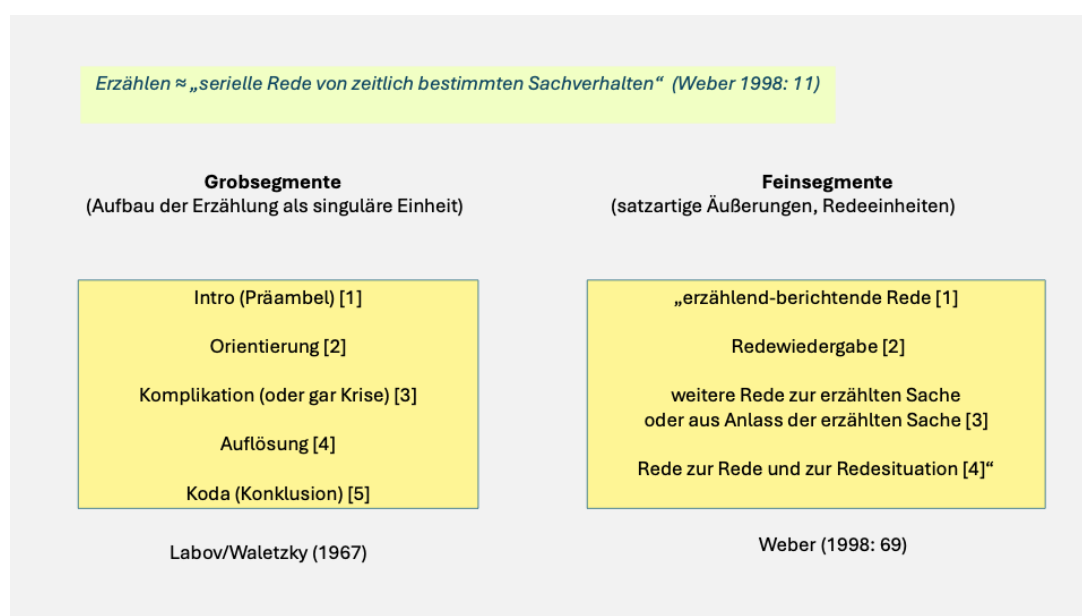


Abbildung 2: Kategorien der Grobsegmentierung und der Feinsegmentierung

Autobiographisches Erzählen im sozialwissenschaftlichen Interview

Wenn wir neueren Arbeiten zur Hervorbringung von Lebensgeschichten (bzw. „life stories“) folgen, dann stellen mündliche Stegreiferzählungen des eigenen Lebens das Phänomen dar, das als Datum empirisch aufgezeichnet wurde, wobei als ein hypothetisches Konstrukt ein „Life-Story-Schema“ die erzählte Lebensgeschichte strukturiert und die Muster dieser Struktur in der Abfolge der „life narrative chapters“ erzeugt, die als Segmente selbst wiederum unterschiedlich gestaltet sein können, z. B. als geraffte chronikförmige Darstellung oder ausgebauter Erzählung einzelner ereignisbezogener Episoden oder auch als Argumentationen (Abbildung 3). Das Schema der Lebenserzählung wird insofern in jedem einzelnen Fall anhand der Bauweise der Segmente rekonstruierbar.

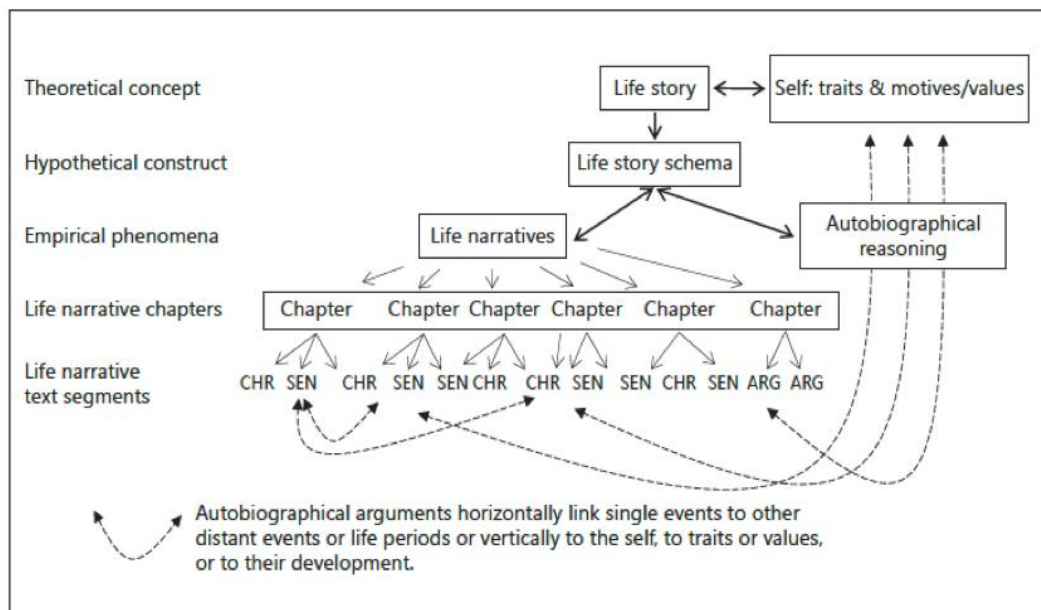
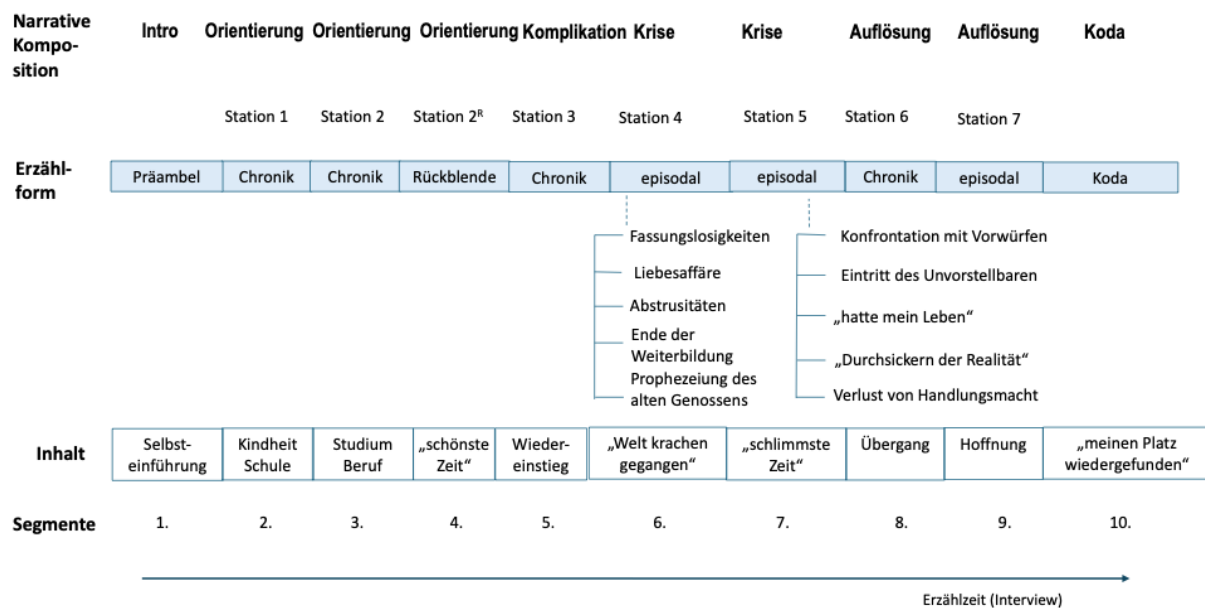


Abbildung 3: Heuristisches Konzept der Lebenserzählung (Habermas und Reese 2015)

Dies sehen wir in Abbildung 4 am Fallbeispiel der konkreten Organisation einer Einstiegserzählung illustriert. Bei dem Fallbeispiel handelt es sich um ein Interview aus der Studie „Politische Kultur und bürgerschaftliches Engagement“ im Rahmen des SFB 580 der Universitäten Halle und Jena.

Insgesamt wurden 114 Interviews in der ersten Welle erhoben. Wir haben diese Interviews im Rahmen einer methodisch orientierten Folgestudie unter Leitung von Audehm, Corsten und Herma (2019–2023, DFG-Projektnr. 405985489) hinsichtlich der Erzählorganisation reanalysiert (Jafke et al. 2022; Corsten und Maleyka 2024). Hierbei haben wir uns im Besonderen mit den Eigenschaften der Einstiegserzählungen befasst, d. h. mit der ununterbrochenen Rede der Interviewpartner:innen im Anschluss an den biographisch-narrativen Eingangsstimulus „Könnten Sie uns bitte Ihr Leben erzählen, möglichst von Anfang an“.

**Abbildung 4: Fallbeispiel**

Die Abbildung 4 zeigt die Form der Grobsegmentierung der Einstiegserzählung, wobei wir lose an die Kategorien anknüpfen, die Tilmann Habermas und Elaine Reese (2015) vorgeschlagen haben. Die Erzählung im Beispielfall umfasste etwa 3.900 Wörter und dauerte etwa 25 Minuten. In dieser Erzählzeit schilderte die interviewte Person ihren gesamten Lebenslauf von der Geburt bis heute. Im Rahmen der Grobsegmentierung konnten wir 10 Abschnitte bzw. Segmente rekonstruieren, also die „chapters“ im Sinne von Habermas und Reese (2015). Diesen 10 Segmenten haben wir jeweils eine dramaturgische Rolle für die narrative Komposition zugeordnet anhand der von Labov und Waletzky postulierten Kategorien. Insbesondere für die Übergänge bzw. Begrenzungen der jeweiligen Segmente kam uns die Methodik von Schütze (1984) zuhilfe. Erstens lässt sich das Intro einer biographischen Erzählung sehr präzise mit Schützes Beschreibung der „biographischen Präambeln“ eingrenzen (dazu ebf. Jafke et al. 2022; Corsten und Maleyka 2024). In der Präambel stellt sich die Person in der Regel kurz anhand zentraler Daten ihrer biographischen Herkunft vor (Geburtsdaten, Kurzcharakterisierung von Familienmitgliedern sowie Hinweise auf das regionale oder soziale Milieu und die historische Zeit des Aufwachsens, Schütze 1984, S. 84). Während die biographische Präambel in Form einer Rede zur Sache vorgestellt wird, ist der Übergang zur eigentlichen Erzählung in der Regel durch eine zeitliche „Rahmenschaltung“ (Schütze 1984) gekennzeichnet, der eine erzählend-berichtende Rede in der Ich-Form folgt¹. Diese temporale Rahmenschaltung (in der Regel durch Konnektoren wie „dann“, „und danach“ usw.) ist neben kodaförmigen Endmarkierungen von Abschnitten die wichtigste Form, anhand derer Grenzen zwischen Segmenten zu erkennen sind. Die grobe Regel dabei lautet, einem Kodaelement eines vorherigen Segments folgt eine temporale Rahmenschaltung, die oft durch redundante Verweise auf einen Zeitwechsel (durch Konnektoren, zeitliche Adverbien) markiert ist.

Die von uns in Abbildung 4 gezeigte Einstiegserzählung entspricht dem konventionellen Schema nach Labov und Waletzky. Nach dem Intro folgen Segmente der Orientierung, bis in Segment 5 eine erste Komplikation (mit nicht ganz erwartungsgemäßen Lebensereignissen) anschließt, der wiederum zwei Segmente folgen, die relativ eindeutig als Krise eingeleitet werden, etwa durch Äußerungen wie

¹ Schütze spricht hier auch vom „Schicksalsfaden der eigenen Lebensgeschichte“ (Schütze 1984, S. 84).

„und dann ist meine Welt es erste Mal krachen gegangen“. Erzähllogisch kohärent ist dann ebenso, dass die Narration mit zwei auflösenden Abschnitten und einem Koda-Abschnitt endet. Die Abschnitte zeigen dabei chronikförmig geraffte und stärker episodisch ausgeschmückte Erzählformen. Die ausgeschmückten Episoden zeichnen sich dadurch aus, dass sie mehrere Untersegmente enthalten. Last but not least lassen sich die Segmente auch grob inhaltlich etikettieren. Interessant ist, dass die gerafften Segmente eher konventionelle Lebensthemen wie Schule, Studium, Berufseinstieg betreffen, während die Krisen und Auflösungen wertende Rahmungen (wie „schlimmste Zeit“, „Hoffnungsschimmer“ usw.) erhalten. Solche Grobsegmentierungen wurden für alle 114 Interviews der ersten Welle manuell annotiert. Für 16 ausgewählte Interviews liegen außerdem Feinsegmentierungen in Propositionen (Einzeläußerungen) vor. 12 von diesen 16 Einstiegserzählungen liegen der nun folgenden Verarbeitung durch maschinelle Textanalyseverfahren zugrunde.

Methodische Vorbereitungen für den Einsatz von NLP-Werkzeugen

Textklassifikationsexperimente mit NLP-Werkzeugen basieren in der Regel auf einem manuell klassifizierten und annotierten (z. B. grobsegmentierten) Teilkorpus, das als ‚Goldstandard‘ verwendet wird (d. h. als ‚ground truth‘ der Klassifikation). In unserer Analyse sind es die genannten 12 Einstiegserzählungen. Sie bestehen aus 22.319 Wörtern, untergliedern sich in 119 Grobsegmente und 2318 Propositionen (Einzeläußerungen) im Sinne der Feinsegmentierung. Wir beschränken uns hier auf die Analyse auf der Basis der Grobsegmentierung.

Die Grobsegmente verteilen sich wie folgt:

- 12 Präambeln
- 42 Orientierungen
- 21 Komplikationen
- 7 Krisen
- 22 Auflösungen
- 13 Koda
- 2 sonstige Segmente („Aushandlungen“)

Unsere Ausgangsüberlegungen konzentrierten sich auf folgende Fragen:

- Können wir Kategorien wie „Krisenepisoden“ über *Topic Modeling* finden?
- Hilft uns *generative KI*? (Problem hier: Sie ist aktuell schlecht in Textklassifikation)
- Handelt es sich um ein Problem, das mit Methoden der Stilometrie gelöst werden kann?

Computerlinguistische Auswertungsstrategien

In diesem Beitrag zeigen wir Befunde in Bezug auf zwei der genannten Fragen, (a) ob sich Kategorien wie „Krisenepisoden“ über *Topic Modeling* (unter Einbezug generativer KI) finden lassen oder (b) ob ein Verständnis der Kategorien als stilometrisch detektierbar angemessener ist. Die Darstellung der Befunde bei Unterstützung durch generative KI hat nur teilweise zu einer Verbesserung geführt (Ziems et al. 2024).

Topic Modeling

Auf Latent Dirichlet Allocation (LDA) beruhendes Topic Modeling (Blei et al. 2003) wird eingesetzt, um Themen und lexikalisches Material in Textsammlungen zu identifizieren und Wahrscheinlichkeitsverteilungen von Wörtern und Themen in einem Korpus zu beschreiben.

Grootendorst (2022) legt mit BERTopic einen Topic-Modeling-Ansatz vor, der auf der Verortung von Wörtern, Sätzen und Abschnitten von Texten im Embedding-Raum mittels einer Transformer-Architektur aufsetzt. Dem Ansatz liegt zugrunde, dass sich Textstücke ähnlich sind, wenn sie im Embedding-Raum nahe beieinanderliegen. In BERTopic werden als ähnlich erkannte Textstücke geclustert. Das statistische Maß TF-IDF (Term-Frequenz im Verhältnis zur inversen Dokument-Frequenz) berechnet die Wahrscheinlichkeit, dass Wörter in den Clustern auftreten.

Anknüpfend an die interpretative bzw. manuelle Grobsegmentierung der Lebenslauf-Erzählungen und die Klassifikation der dramaturgischen Rolle der Segmente für die Erzählorganisation wurde untersucht, welche Topics aus einer Kollektion von Lebenslaufbeschreibungen extrahiert werden, wenn BERTopic in einem Bottom-up-Ansatz verwendet wird, und wie sich die Topics unterscheiden, wenn die Topics durch den Einsatz generativer KI geschärft werden, also bestimmte aus theoretischer Sicht oder aus der vorgängigen Topic-Analyse erwartete Wörter vorgegeben werden, um die Generierung der Topics zu steuern (Kahle und Kliche 2022).

Dies eröffnete die methodische Möglichkeit eines dreifachen Vergleichs der Güte der Resultate einer humanen interpretativen Analyse, eines Bottom-up Topic Modelings und eines KI-gesteuerten Topic Modelings. Im Projekt, das Textkorpora aus der soziologischen Biographieforschung, der Oral History und soziologischen Expert:innen-Interviews exploriert hat, konnten so auf qualitative (regelbasierte) und quantitative (stochastische) Weise Beurteilungen der Güte der human und maschinell vorgenommenen Klassifikationen/Interpretationen vorgenommen werden. Dabei lassen sich Fälle zeigen, in denen Ergebnisse der Topic-Modeling-Verfahren auch in regelbasierter Beurteilung interpretativ angemessene Zuordnungen lieferten bzw. auch statistisch zuverlässige Zuordnungen (z. B. gemäß Krippendorffs Alpha oder Cohens Kappa) durch humane Klassifikation aufwiesen.

Zur Vorbereitung des Topic Modelings zu den Grobsegmenten wurden die Interviewtexte, deren Transkription an der Informalität der gesprochenen Sprache orientiert war (Dialektfärbung, grammatische Fehlkonstruktionen, Satzabbrüche, Fehlaussprache z. B. „g“ wie „j“), in Standardsprache überführt. Hinzu trat die Lemmatisierung der Texte sowie die Entfernung von Stoppwörtern. Die folgende Abbildung 5 zeigt als Ergebnis 13 Topics.

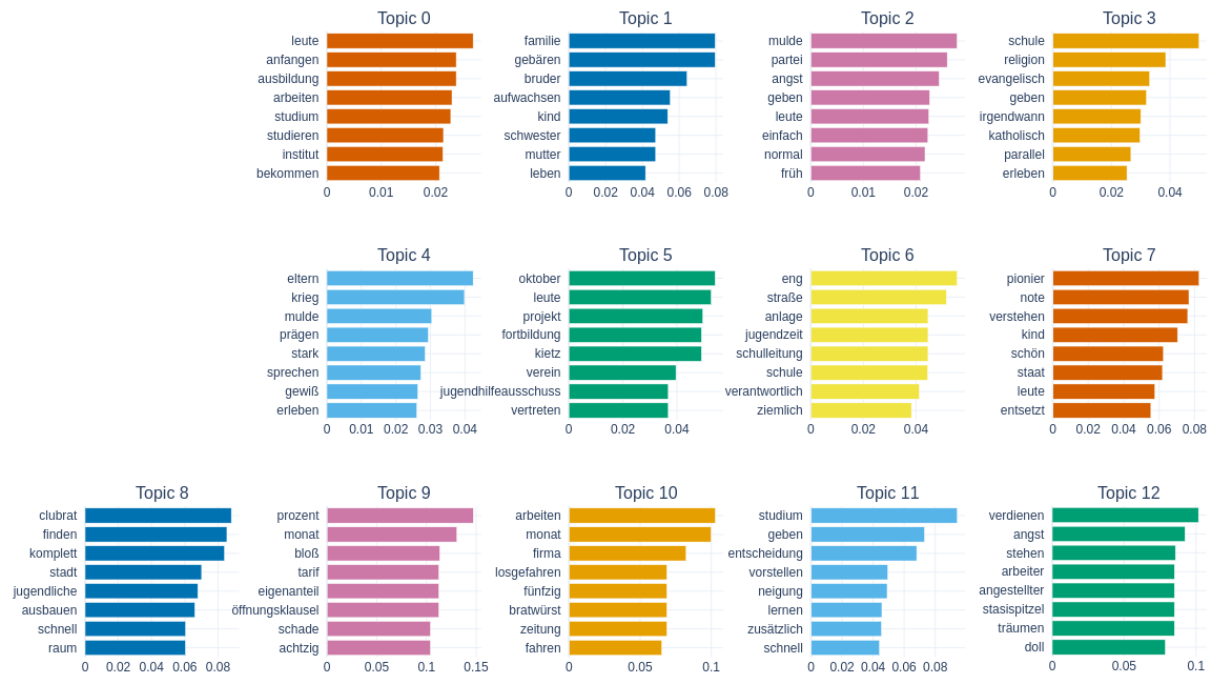


Abbildung 5: Topic Modeling der manuell annotierten Einstiegserzählungen

Die maschinell herauspräparierten Topics weisen starke inhaltliche Bezüge auf. Topic 0 etwa kreist um Begriffe wie „Ausbildung“, „Arbeiten“, „Studium“, Topic 1 um Begriffe des familiären Kontexts („Bruder“, „gebären“, „Mutter“). Aus interpretativer Perspektive weist das Ergebnis die typischen Limitationen von maschineller bzw. statistischer Verarbeitung von Sprachdaten auf. Sie sind einerseits trivial, andererseits bieten sie auch Themenfelder als verschieden an, die sich semantisch als redundant erweisen. Worin unterscheidet sich z. B. der Bedeutungshof der Topics 0, 6, und 11? Alle enthalten mehr oder weniger Bezüge zum Thema Bildung. Aber lässt sich anhand der statistisch signifikanten Wortgruppen deren Differenzierung trennscharf und semantisch plausibel nachvollziehen? Hier scheint nachträglicher interpretativer Spekulation Tür und Tor geöffnet. Gleichwohl dokumentieren die Topics 2 und 12 Wörter (wie „Angst“), die Prägnanz für Krisenschilderungen aufweisen.

Daher erproben wir im nächsten Schritt einen Ansatz, um Textpassagen zu detektieren, die Krisenbeschreibungen enthalten. Wir sehen diese Kategorie als Beispiel für das „Wie“, da mit der Darstellung als Krise ausgedrückt wird, wie die Personen bestimmte erzählte Episoden empfanden. Die Erkennung der Krisendarstellungen erwies sich als schwierig und ein spezifisches Krisenvokabular zeigte sich in den Ergebnissen des Topic Modelings kaum. Hier erproben wir eine Erweiterung unseres Ansatzes unter Verwendung generativer KI. Über die API von OpenAI griffen wir auf GPT-4 Turbo zu, um die 119 untersuchten Segmente zu verarbeiten. So entstanden 119 sprachlich vereinheitlichte Textstücke mit „geschärften“ Krisen-Beschreibungen, auf die wir erneut BERTopic anwendeten. Wir verbinden dieses Vorgehen mit dem strategischen Ziel, daraus ein ‚Krisen-Vokabular‘ im Sinne eines semantischen Repertoires abzuleiten (Kahle und Kliche 2022). Die Ergebnisse sind in Abbildung 6 dargestellt.

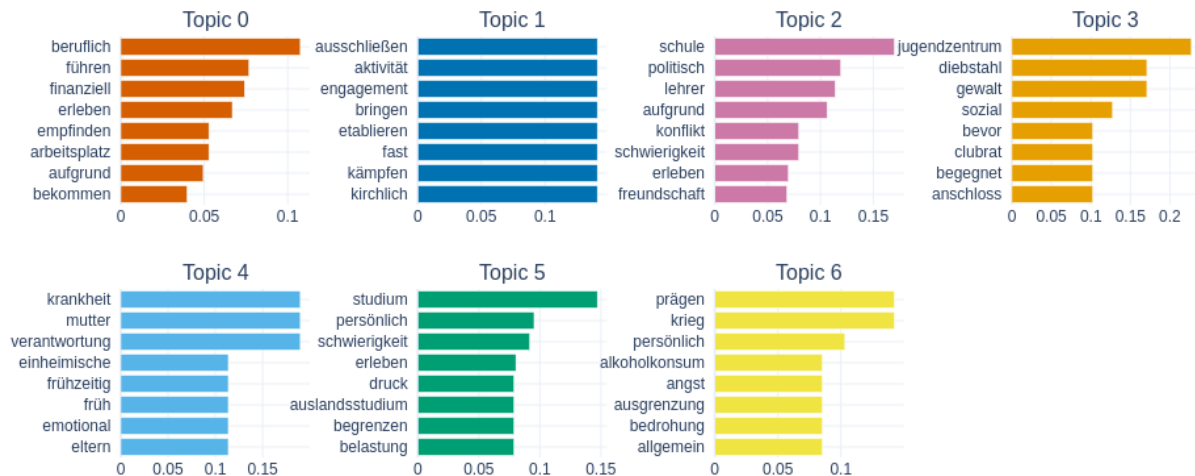


Abbildung 6: Topics der als „Krise“ annotierten Passagen

Die ermittelten Topics weisen nun wieder Gruppen von Wörtern auf, die in statistischer Hinsicht eine höhere Wahrscheinlichkeit besitzen, im Rahmen des latenten Themas aufzutreten, wobei das latente Thema selbst wieder aus einer statistisch erzeugten Gruppe von Wörtern („bag of words“) besteht. Anders ausgedrückt: das Topic Modeling erzeugt in Bezug auf das Korpus der Krisenpassagen sieben Gruppen von Wörtern, die als latente Themen behandelt werden, die sich im Rahmen eines Korpus voneinander unterscheiden lassen, wobei dann pro Topic für jedes Wort des Topics eine Wahrscheinlichkeit berechnet wird, in diesem Topic zu erscheinen. Hierbei tritt wiederum das Problem auf, dass nun anhand der wahrscheinlichsten Wörter pro Topic zu interpretieren ist, worin das latente Thema des Topics zu vermuten ist.

Wenn wir dazu wieder einzelne Topics beispielhaft herausgreifen, etwa Topic 1, dann wäre spekulativ zu erörtern, inwiefern Termini wie „ausschließen“ oder „kämpfen“ Momente von Krisen repräsentieren mögen, die dann mit Aktivitätsfeldern wie „Engagement“ oder „Kirche“ in Verbindung stehen. Oder dass womöglich Krisen im „Beruf“ oder bezogen auf den „Arbeitsplatz“ wie in Topic 0 mit mental-psychischen Prozessen wie „Erleben“ oder „Empfinden“ in Zusammenhang gebracht werden. Allerdings gibt es jenseits der statistischen Wahrscheinlichkeit eines ko-okkurrenten Auftretens keine weiteren Validierungsmöglichkeiten, derartige semantische oder narrative Zusammenhänge zu validieren. Es sei denn, man vergleicht wieder fallweise (also qualitativ/interpretativ) das Auftreten der jeweiligen Wörter mittels einer Kombination von „close“ und „distant reading“ im Rahmen einer „Blended-Reading-Strategie“ (Wiedemann und Lemke 2015; Biemann et al. 2022).

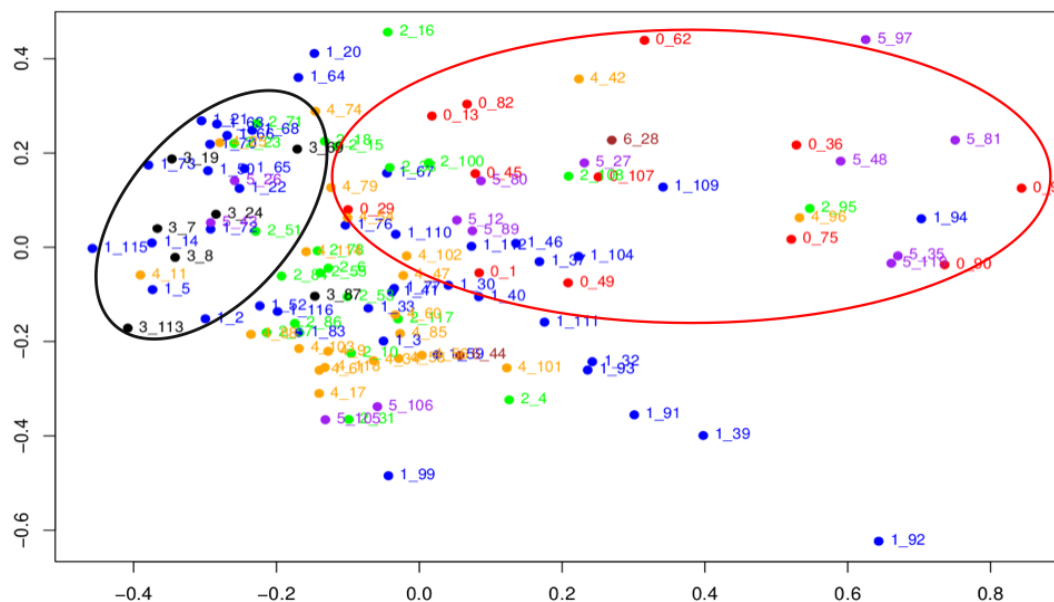
Wir halten als Zwischenbilanz fest, dass die Vorverarbeitung und Eingrenzung des untersuchten Korpus zur Herauspräparierung eines spezifischeren Vokabulars für Krisen und damit verbundenen Themensträngen führen kann. Und ein solches Verfahren könnte nun durch die Hinzunahme weiterer Textkorpora (von Einstiegserzählungen) *überprüft* werden. Zudem wäre der *Aufbau eines Diktionärs* für die Detektion von Krisenpassagen in Interviews daraus ableitbar. Gleichwohl entsprechen die über Topic Modeling gefundenen „Krisen“-Segmente nur unzureichend den „Krisen“-Segmenten aus dem Goldstandard. Daher erproben wir als nächsten Ansatz die Möglichkeiten der quantitativen Stilometrie (z. B. Pennebaker 2011).

Stilometrie

Dafür, dass Muster des Erzählens als ein *stilometrisches* Problem sensu Pennebaker (2011) aufgefasst werden können, sprechen einige der Arbeiten, die sich im Kontext der maschinellen Sprachverarbeitung mit Narrationen befassen. So haben Ryan L. Boyd et al. (2020) eine stilometrische Analyse der dramati-

schen Organisation von Erzählungen vorgelegt, die sie am Erzählmodell von Gustav Freytag entwickeln. Dabei legen sie eine fünfphasige Sequenz des Erzählens („exposition, rising action, climax, falling action, and denouement“, Boyd et al. 2020, S. 3) zugrunde, die mit kleineren Abstrichen den oben zitierten Phasenmodellen im Anschluss an Labov und Waletzky (s. Abbildung 2) entspricht.

Hierbei wird als ein gängiges Verfahren die multidimensionale Skalierung genutzt, um Ähnlichkeiten über Distanzen zwischen kategorisierten Texten in einem zweidimensionalen Raum abzubilden. Die beiden Hauptachsen der Unterscheidung werden durch die Schätzung der Dimensionen gebildet, die die größte Varianz erreichen. Die multidimensionale Skalierung (MDS) erstellt ein Streudiagramm, das Texte in einem zweidimensionalen Raum abbildet, wobei ihre Positionen die stilistische Nähe widerspiegeln (siehe Abbildung 7). Die Berechnung besteht in einer Burrows-Delta-Matrix, um paarweise stilistische Abstände zu erfassen. MDS projiziert diese hochdimensionalen Beziehungen in einen niedrigdimensionalen Raum, was eine intuitive Visualisierung ermöglicht. Das Streudiagramm hebt Cluster ähnlicher Textpassagen hervor und bietet eine räumliche Darstellung von Ausreißern. Jedes klassifizierte Segment wird als Punkt dargestellt, mit einem Kürzel beschriftet und anhand derselben Phasenklassifikationen, die bei der manuellen Annotation verwendet wurden, farblich gekennzeichnet. Der Textskalierung gingen als vorbereitende Schritte eine Überführung in Standardsprache und eine Lemmatisierung voraus.



Legende: 0: Präambel, 1: Orientierung, 2: Komplikation, 3: Krise, 4: Auflösung, 5: Koda

Abbildung 7: Multidimensionale Skalierung der kategorisierten Einstiegserzählungen

Als zentrale Ergebnisse können festgehalten werden: Die als Orientierung, Komplikation und Auflösung klassifizierten Textpassagen streuen sehr breit in dem statistisch konstruierten Raum, während sich die Präambeln und die Krisenerzählungen auf spezifischere Bereiche verteilen (Hervorhebung durch Umrandung). Insgesamt scheint dies gemessen an der Spezifik des Vokabulars für wenig stilistische Differenz zwischen den nach ihren Funktionen kategorisierten Segmenten zu sprechen. Dennoch bestätigt es den im Topic Modeling gewonnenen Eindruck, dass Krisenerzählungen mit einem begrenzteren Vokabular einhergehen. Ähnliches gilt für die Präambeln.

Rückschlüsse auf qualitative Klassifikationstechniken

Die Ergebnisse des Topic Modelings und der Stilometrieanalyse zeigen jedoch einen Bedarf, weitere Analysen an die noch explorativen Befunde anzuschließen. Dies könnten regelbasierte statistische Verfahren, Verfahren unter Einbezug generativer KI oder auch interpretative Methoden sein, die den Auffälligkeiten der explorativen Befunde präziser nachgehen können und sie im Sinne einer höheren Inhaltsvalidität aufzuklären vermögen.

Fazit

Unser Ausgangsproblem lautete: Kann *Topic Modeling* (als NLP-Anwendung auf Grundlage distributioneller Semantik) nutzbringend in qualitative Analysen eingebracht werden, die insbesondere auf Formelemente der Pragmatik von Texten (als protokollierter sprachlicher Interaktion/Tätigkeit) abzielen?

Der Einsatz von BERTopic (teils in Verbindung mit generativer KI) zeigte, dass nach Regeln einer Pragmatik des Erzählens klassifizierte Texte auch semantisch systematisch differieren, dass sich also speziell Korpora (Textpassagen), in denen formalpragmatisch *Krisen* thematisiert werden, ebenfalls im Hinblick auf die genannten Wortfelder im Sinne von ‚bags of words‘ voneinander unterscheiden. Es kann somit als sinnvoll erachtet werden, Verfahren für Topic Modeling zu nutzen, um *erstens Befunde manueller Klassifikation (indirekt) zu prüfen* oder *zweitens Diktionäre (oder Seeds) zu entwickeln*, um automatisiert Krisenpassagen in Erzähltexten (hier: Einstiegserzählungen in Interviews) zu detektieren.

Gleichwohl konnten wir die Ergebnisse von Topic Modeling (bisher) *nicht direkt an die von uns gesuchten narratologischen Strukturen* rückbinden. Die stilometrische Analyse zeigte: Funktionswörter sind stärkste Diskriminatoren, aber nicht nur einzelne narrative Phasen unterscheiden sich trennscharf in einem multiskalierten Raum, während die meisten Erzählphasen zumindest bezüglich des Vokabulars ein ähnliches Verteilungsmuster aufwiesen.

Auch weisen die bisherigen Experimente Limitationen auf, weil sie auf einem kleinen Sample beruhen, was jedoch wiederum insofern methodisch wünschenswert ist, als qualitative Forschung in der Regel keine großen Textmengen untersucht.

Literatur

- Austin, John L. 1962. *How to do Things with Words*. Harvard: University Press
- Biemann, Chris, Gerhard Heyer und Uwe Quasthoff. 2022. *Wissensrohstoff Text*. Wiesbaden: Springer.
- Blei, David M., Andrew Y. Ng und Michael I. Jordan. 2003. Latent Dirichlet Allocation. *The Journal of Machine Learning Research* 3:993-1022
- Boyd, Ryan L., Kate G. Blackburn und James W. Pennebaker. 2020. The narrative arc: Revealing core narrative structures through text analysis. *Science Advances* 6(32):eaba2196
- Bourdieu, Pierre. 1982. *Qu'est-ce que parler veut dire?* Paris: Fayard.
- Corsten, Michael, und Laura Maleyka. 2024. Die Präsentation der biographischen Visitenkarte: Erzählstimulus und Selbsteinführung am Interviewbeginn. *BIOS – Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen* 37:157–181
- De Saussure, Ferdinand. 1995 [1916]. *Cours de linguistique générale*. Paris: Payot, coll.
- Eder, Maciej, Jan Rybicki, J. und Mike Kestemont. 2016. Stylometry with R: a package for computational text analysis. *R Journal* 8:107–121.

- Fischer-Rosenthal, Wolfram, und Gabriele Rosenthal. 1997. Narrationsanalyse biographischer Selbstpräsentation. In *Sozialwissenschaftliche Hermeneutik*, Hrsg. Ronald Hitzler und Anne Honer, 133–164. Opladen: Leske + Budrich.
- Foucault, Michel. 1969. *L'archéologie du savoir*. Paris: Éditions Gallimard.
- Gao, Xin, und Cem Sazara. 2023. Discovering mental health research topics with Topic Modeling. arXiv:2308.13569
- Greimas, Algirdas J. 1966. *Sémantique structurale. Recherche de méthode*. Paris: Larousse.
- Grootendorst, Marten. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794
- Habermas, Tilmann, und Elaine Reese. 2015. Getting a life takes time: The development of the life story in adolescence, its precursors and consequences. *Human Development* 58:172–201
- Harris, Zellig. 1954. Distributional Structure. *Word* 10(2/3):146–162.
- Jafke, Larissa, Laura Maleyka, Holger Herma, Karsten Spindler, Michael Corsten und Kathrin Audehm. 2022. Sich als Subjekt des Sprechens über das eigene Leben einführen. In *Positioning the Subject*, Hrsg. Saša Bosančić, Folke Brodersen, Lisa Pfahl, Lena Schürmann, Tina Spies und Boris Traue, 115–138. Wiesbaden: Springer VS.
- Kahle, Patrick, und Kliche, Fritz. 2022. Learning inductive and deductive topics in parallel using seeded topic modeling. In *Diversity of Methods and Materials in Digital Human Sciences: Proceedings of the Digital Research Data and Human Sciences DRDHum Conference 2022, December 1–3*, Hrsg. Jarmo H. Jantunen, Johanna Kalja-Voima, Matti Laukkarinen, Anna Puupponen, Margareta Salonen, Tuija Saresma, Jenny Tarvainen und Sabine Ylönen, 28–42. Jyväskylä, Finland.
- Kallmeyer, Werner, und Fritz Schütze. 1977. Zur Konstitution von Kommunikationsschemata der Sachverhaltsdarstellung. In *Gesprächsanalysen. Vorträge, gehalten anlässlich des 5. Kolloquiums des Instituts für Kommunikationsforschung und Phonetik*, Hrsg. Dirk Wegner, 159–274. Bonn.
- Küsters, Ivonne. 2021. *Narrative Interviews*. Wiesbaden: Springer VS.
- Labov, William, und Waletzky, Joshua. 1967. Narrative analysis: Oral versions of personal experience. In *Essays on the Verbal and the Visual Arts*, Hrsg. June Helm, 3–38. Seattle und London: University of Washington Press.
- Lämmert, Eberhard. 1955. *Bauformen des Erzählens*. Stuttgart: Metzler.
- Pennebaker, James W. 2011. *The secret life of pronouns: What our words say about us*. New York: Bloomsbury.
- Schütze, Fritz. 1984. Kognitive Figuren des autobiographischen Stegreiferzählens. In *Biographie und soziale Wirklichkeit. Neue Beiträge und Forschungsperspektiven*, Hrsg. Martin Kohli und Günter Robert, 70–117. Stuttgart: Metzler.
- Stanzel, Franz K. 1982. *Theorie des Erzählens*. Göttingen: Vandenhoeck und Ruprecht.
- Weber, Dietrich. 1998. *Erzählliteratur. Schriftwerk – Kunstwerk – Erzählwerk*. Göttingen: Vandenhoeck und Ruprecht.
- Wiedemann, Gregor, und Matthias Lemke. 2016. Text Mining für die Analyse qualitativer Daten. Auf dem Weg zu einer Best Practice? In *Text Mining in den Sozialwissenschaften*, Hrsg. Matthias Lemke und Gregor Wiedemann, 397–419. Wiesbaden: Springer.
- Ziems, Caleb, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, Diyi Yang. 2024. Can Large Language Models Transform Computational Social Science? *Computational Linguistics* 50:237–291.