

Die Sprache der Erinnerung – analysieren und verstehen

Korpuslinguistische Zugänge zu Oral-History-Daten

Gerstenberg, Annette

gerstenberg@uni-potsdam.de
Universität Potsdam, Germany

Leh, Almut

almut.leh@fernuni-hagen.de
Fernuniversität Hagen, Germany

Möbus, Dennis

dennis.moebus@fernuni-hagen.de
Fernuniversität Hagen, Germany

Pagenstecher, Cord

cord.pagenstecher@fu-berlin.de
Freie Universität Berlin, Germany

Die Motivation des Workshops ist es, die interdisziplinären Potentiale der Anwendung korpuslinguistischer Tools auszuloten, ohne die Spezifik und Sensibilität von Oral-History-Interviews aus dem Blick zu verlieren.

Beschreibung des Themas

Oral-History-Interviews sind narrative, meist lebensgeschichtliche Erinnerungsinterviews, die in der zeithistorischen Forschung, aber auch in den Sozial- und Kulturwissenschaften als Quellen bzw. Datengrundlage genutzt werden. Als Teil des kulturellen Erbes werden sie an verschiedenen Forschungs- und Gedächtniseinrichtungen gesammelt und für Sekundäranalysen aufbereitet. Interview-Archive sind Gedächtnisinstitution in zweifacher Hinsicht: einmal in der Bewahrung und Vermittlung von Wissensbeständen, die in ihrer Gesamtheit das kulturelle Gedächtnis einer Gemeinschaft bilden. Zusätzlich aber auch in einem unmittelbaren Sinn, insofern die hier archivierten Wissensbestände selbst Erinnerungen, also Gedächtnisinhalte, sind, und gehören als solche zum digitalen Gedächtnis. Und da die Audio- oder Videoaufzeichnung sowie die Transkripte vielfach in elektronischer Form vorliegen, gehören Interview-Archive zum digitalen Gedächtnis. Als maschinenlesbare Daten sind Oral-History-Interviews nicht nur für die Humanities, sondern auch für Linguistik und Informatik interessante multimodale, freilich wenig strukturierte Daten.

In den historischen Wissenschaften werden Oral-History-Interviews in intensiver hermeneutischer Arbeit analysiert und interpretiert, wobei sich die Fragestellung im Spannungsfeld persönlicher und kollektiver Relevanz situieren. Die digitale Erschließung dieser Quellen eröffnet die Möglichkeit, gerade auch überindividuelle Muster der Erinnerung und ihrer sprachlichen Verfasstheit mit Hilfe von korpuslinguistischen Tools zu erschließen.

Auf Basis von Praxisbeispielen wird im Workshop die Frage nach dem Mehrwert dieser technikgetriebenen Analysen disku-

tiert. Können solche Tools das hermeneutische Verstehen unterstützen oder gar bereichern, oder besteht die Gefahr, die Subjektivität der Erzählung und die individuelle Entstehungssituation der Quellen aus dem Blick zu verlieren? Sind die digital erkannten Muster am Ende Artefakte oder können sie Schlüssel zum tieferen Verstehen sein?

Die Motivation des Workshops ist es, die interdisziplinären Potentiale der Anwendung korpuslinguistischer Tools auszuloten, ohne die Spezifik und Sensibilität von Oral-History-Interviews aus dem Blick zu verlieren.

Im ersten Block des Workshops widmen sich drei Impulsreferate diesen Koordinaten des Themas. Zuerst wird thesenartig entwickelt, welche Implikationen die Digitalisierung für die Auswertung des Quellentyps Oral-History-Interviews hat (Almut Leh). Daran anschließend wird das Projekt Oral-History.Digital vorgestellt, das verschiedene Interviewbestände in einer webbasierten Erschließungs- und Forschungsumgebung verbindet und damit sammlungsübergreifende und vergleichende Zugänge ermöglicht (Cord Pagenstecher). Mit dem Interesse, die sprachliche Konstruktion der Erinnerung auszuloten, werden darauf aufbauend linguistische Fragestellungen der Pragmatik und Semantik entwickelt, die sich auf Oral-History-Interviews anwenden lassen (Annette Gerstenberg).

Im zweiten Block werden anhand eines vorbereiteten Arbeitskorpus ausgewählter Oral-History-Interviews drei Anwendungsszenarien vorgestellt. Sie zeigen, wie die in den Impulsreferaten entwickelten Fragestellungen konkret bearbeitet werden können.

Zur Einordnung wird zunächst das ausgewählte Arbeitskorpus vorgestellt: im Hinblick auf thematischen Schwerpunkt, Entstehungskontext und sprachliche Spezifika der enthaltenen Teiltex-te. Mit sprachstatistischen Basisdaten und automatisch ermittelten „Schlüsselwörtern“ werden Unterschiede der verwendeten Teiltex-te des Arbeitskorpus erläutert und visualisiert. Geplant sind darauf aufbauend drei Arbeitseinheiten, in denen jeweils eine in der Oral-History bisher wenig genutzte Analysemöglichkeiten im Mittelpunkt steht. In jeder Arbeitseinheit wird die Analyse demonstriert und zugleich die Möglichkeit gegeben, jeden Schritt selbst nachzuvollziehen.

(1) Auf Basis des lemmatisierten und nach Wortarten ausgezeichneten Datensatzes werden Häufigkeiten, von Wortarten und Grundwörtern untersucht. Als aufschlussreich hat sich zum Beispiel der Vergleich der Vorkommen von Pronomina (*ich* vs. *wir*) erwiesen (Knowles et al. 2021). Weiterhin werden die häufigsten Verben semantisch kategorisiert, wobei Verben des Erinnerens und des Sagens besonders berücksichtigt werden.

(2) Ausgehend von der Analyse häufig vorkommender Wortfolgen (n-grams) werden Instanzen formelhaften Sprechens ermittelt. Gerade im Vergleich verschiedener Interviews erweisen sich solche als „Floskeln“ unterschätzte Redewendungen als charakteristisch und aussagekräftig für eine distanzierende oder aktualisierende Rahmung der erzählten Erinnerungen. Häufig verwendete und in der Analyse meist übersehene Wendungen wie *das ist alles lange her*, *das werde ich nie vergessen* oder *das hat uns geprägt* helfen dabei zu beobachten, wie das Erinnernte eingeordnet wird.

(3) Die Themen des Arbeitskorpus werden mit einer vorbereiteten Topic Modeling-Analyse, bei der lexikalische Cluster herausgearbeitet und visualisiert werden, vorgestellt. Dieser statistische Zugriff auf den Wortschatz wird in Kollokations-Analysen fortgesetzt und durch Kontextanalysen ergänzt. Dabei werden statistisch relevante gemeinsame Vorkommen ermittelt, aus denen deutlich wird, wie typische Erlebnisse sprachlich kodiert werden – wenn zum Beispiel das Kollokat von *Krieg* der *Schützengraben* ist oder zu *Gefangenen* häufig die Herkunft (*französische*, *russische*) angegeben wird.

Jedes Szenario wird mit einer technischen Anleitung zum Mitmachen verbunden. Auf diese Weise können wir zielgerichtet und konkret mit den Sprachdaten arbeiten und davon ausgehend weiterdenken. In der Diskussion wird es darum gehen, welche Aussagekraft wir den Ergebnissen zumessen und ob die unterschiedlichen Sichtweisen auf den gleichen Datenbestand neue Fragestellungen aufwerfen oder bekannte Interessen neu akzentuieren.

Helfen uns die digitalen Werkzeuge, den Interviews besser zuzuhören und die Sprache der Erinnerung besser zu verstehen? Als Ergebnis des Workshops versprechen wir uns Antworten auf diese Fragen der Erkenntnismöglichkeiten von Oral-History mit dem Werkzeugkasten der Korpuslinguistik.

Format

Halbtägiger Workshop, 7.3.2022

Arbeitsform: 1. Phase: Impulsreferate mit Diskussion, 2. Phase: Anwendungsszenarien von Korpus-Tools mit vorbereiteten Datensamples und Anleitung zur eigenen Datenexploration, 3. Phase: Diskussion der Ergebnisse und ihrer Forschungsrelevanz.

Der Workshop gibt Gelegenheit zur eigenen Datenexploration, wofür ein eigener Laptop notwendig ist. Die verwendeten Tools sind einfach zu installieren. Vor dem Workshop werden Datensätze zur Verfügung gestellt und Hinweise zu Tools gegeben, die installiert werden sollten, die verwendeten Werkzeuge sind sicher und gut handhabbar. Interessensbekundungen können mit spezifischen Fragen an Oral-History-Daten eingereicht werden, ggf. auch in Verbindung mit der Beschreibung eigener Datenbestände, die dann in die Workshoparbeit einbezogen werden können. Am Workshop kann auch ohne eigene Datenarbeit teilgenommen werden.

Die Vorbereitung der Texte wird thematisiert und mit Hilfe von Screenshots nachvollziehbar gemacht (Bereinigung aus der Datenbank extrahierter Daten; Erstellen der Nur-Text-Version; Tokenisierung, POS-Tagging, Lemmatisierung mit WebLicht und TreeTagger).

Die gemeinsame Arbeit im Workshop nutzt diese vorbereiteten Dateien. Für die Analyse von Verbformen, Pronomina (1) und n-grams (2) werden robuste Werkzeuge (AntConc, TextPad) verwendet. Es werden Frequenzen ermittelt und Vorkommen im Kontext überprüft und für unterschiedliche Oral History-Quellen verglichen; in der Diskussion werden die in den Eröffnungsreferaten aufgemachten Perspektiven auf die Ergebnisse bezogen. Das Topic Modeling (3) wird mit Python auf Jupyter-Konsolen in einem Google Colab durchgeführt, Datengrundlage werden die Transkriptionen lebensgeschichtlicher Interviews sein, als Topic Modeling-Engines kommen Gensim und dessen Mallet-Implementation zum Einsatz

Angaben zum Zielpublikum, insbesondere zu notwendigem Vorwissen

Wir laden Interessierte aus allen Disziplinen ein, die mit großen Textkorpora arbeiten und daran interessiert sind, die Anwendbarkeit computerlinguistischer Verfahren auf hermeneutische Probleme zu diskutieren. Es gibt die Möglichkeit, die Analyse mitzumachen oder nur zu verfolgen und die Relevanz der Ergebnisse auszuleuchten. Für den Workshop ist ein interdisziplinär zusammengesetztes Plenum aus den Digital Humanities sowie aus

relevanten Themenfeldern wie Geschichtswissenschaften, Kulturwissenschaften, Literaturwissenschaft, Linguistik, Soziologie willkommen.

Interesse an Oral-History bzw. interviewbasierter Forschung in anderen Disziplinen und / oder an Anwendungsszenarien korpuslinguistischer Tools wird vorausgesetzt.

Zahl

10–30 Teilnehmerinnen und Teilnehmer

Technische Voraussetzung

Der Raum sollte mit Projektor und W-LAN ausgestattet sein. Eigene Laptops werden nach Verfügbarkeit mitgebracht.

Bibliographie

Apel, Linde / Almut Leh / Cord Pagenstecher (in print), "Oral History im digitalen Wandel. Interviews als Forschungsdaten", in: Linde Apel (Hrsg.): *Erinnern, erzählen, Geschichte schreiben. Oral History im 21. Jahrhundert*.

Fechner, Martin / Andreas Weiß (2017): "Einsatz von Topic Modeling in den Geschichtswissenschaften: Wissensbestände des 19. Jahrhunderts", in: *Zeitschrift für digitale Geisteswissenschaften* 2. http://doi.org/10.17175/2017_005.

Gerstenberg, Annette (2017): "A Difficult Term in Context: The Case of French STO", in: Erich Kasten / Katja Roller / Joshua Wilbur (Hrsg.): *Oral History Meets Linguistics*. Fürstenberg: Kulturstiftung Sibirien: 159-184.

Graham, Shawn / Ian Milligan / Scott Weingart (2014): *Exploring Big Historical Data: The Historian's Macroscope*. London.

Knowles, Anne Kelly / Paul B. Jaskot / Tim Cole / Alberto Giordano (2021): "Mind the Gap: Reading across the Holocaust Testimonial Archive", in: Tim Cole / Simone Gigliotti (eds.): *The Holocaust in the 21st Century: Relevance and Challenges in the Digital Age*. United States: Northwestern University Press: 216-241.

Leh, Almut / Joachim Köhler / Michael Gref / Nikolaus P. Himmelmann (2019): "Audiovisual Data in Digital Humanities", in: *VIEW Journal of European Television History and Culture* 7/14: 138.

Pagenstecher, Cord (2019): "Digital Humanities und biographische Forschung", in: *BIOS* 30/1-2/2017: 76-91.

Pagenstecher, Cord / Stefan Pfänder (2017): "Hidden dialogues. Towards an Interactional Understanding of Oral History Interviews", in: Erich Kasten / Katja Roller / Joshua Wilbur (Hrsg.): *Oral History Meets Linguistics*. Fürstenberg: Kulturstiftung Sibirien: 185-207.

Pagenstecher, Cord / Doris Tausendfreund (2013): "Das Online-Archiv 'Zwangsarbeit 1939–1945'", in: Nicolas Apostolopoulos / Cord Pagenstecher (Hrsg.): *Erinnern an Zwangsarbeit. Zeitzeugen-Interviews in der digitalen Welt*. Berlin: Metropol: 71-96.

Philipps, Axel (2018): "Text Mining-Verfahren als Herausforderung für die rekonstruktive Sozialforschung", in: *Sozialer Sinn* 19/2, S. 367-388.

Salman, Munir / Felix Engel / Almut Leh / Matthias Hemmje (2019): "Digital Humanities und biographische Forschung", in: *BIOS* 30/1-2/2017: 92-100.