

Dröge, Kai

Datenaufbereitung durch Transkription

Gras, Juliana [Hrsg.]; Schieferdecker, Ralf [Hrsg.]: Einführung in Qualitative Sozialforschung. Grundlagen für Studierende pädagogischer Studiengänge. Bad Heilbrunn : Verlag Julius Klinkhardt 2025, S. 224-234



Quellenangabe/ Reference:

Dröge, Kai: Datenaufbereitung durch Transkription - In: Gras, Juliana [Hrsg.]; Schieferdecker, Ralf [Hrsg.]: Einführung in Qualitative Sozialforschung. Grundlagen für Studierende pädagogischer Studiengänge. Bad Heilbrunn : Verlag Julius Klinkhardt 2025, S. 224-234 - URN: urn:nbn:de:0111-pedocs-338562 - DOI: 10.25656/01:33856; 10.35468/6188-15

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-338562>

<https://doi.org/10.25656/01:33856>

in Kooperation mit / in cooperation with:



<http://www.klinkhardt.de>

Nutzungsbedingungen

Dieses Dokument steht unter folgender Creative Commons-Lizenz: <http://creativecommons.org/licenses/by-nd/4.0/deed.de> - Sie dürfen das Werk bzw. den Inhalt vervielfältigen, verbreiten und öffentlich zugänglich machen, solange Sie den Namen des Autors/Rechteinhabers in der von ihm festgelegten Weise nennen und das Werk bzw. diesen Inhalt nicht bearbeiten, abwandeln oder in anderer Weise verändern.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

This document is published under following Creative Commons-License: <http://creativecommons.org/licenses/by-nd/4.0/deed.en> - You may copy, distribute and transmit, adapt or exhibit the work in the public as long as you attribute the work in the manner specified by the author or licensor. You are not allowed to alter or transform this work or its contents at all.

By using this particular document, you accept the above-stated conditions of use.



Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Datenaufbereitung durch Transkription

Abstract

Die Transkription stellt einen zentralen Schritt der Datenaufbereitung in der qualitativen Sozialforschung dar. Sie ermöglicht die Entschleunigung flüchtiger sozialer Interaktionen, die Reduktion komplexer Datenmengen und die Dokumentation der Analysegrundlagen. Der Beitrag erläutert die methodischen und technischen Anforderungen an die Transkription, einschließlich der Entscheidung für Audio- oder Videoaufzeichnungen, sowie unterschiedlicher Transkriptionssysteme mit variierendem Detaillierungsgrad. Zudem werden praxisnahe Hinweise zur Tonqualität und zur rechtlichen Absicherung von Aufzeichnungen gegeben. Abschließend diskutiert der Beitrag Möglichkeiten und Grenzen KI-gestützter, automatisierter Transkription anhand der freien Software „noScribe“.

Schlagwörter: Datenaufbereitung, Interviewanalyse, Transkription, Transkriptionssysteme

1 Warum transkribieren?

In den empirischen Wissenschaften gibt es zwischen Erhebung und Auswertung einen wichtigen Zwischenschritt: die Datenaufbereitung. Biolog:innen bspw. erstellen aus einer Pflanze zunächst ein Präparat zum Mikroskopieren. Archäolog:innen dokumentieren eine Fundstelle in einem detaillierten Lageplan. Die Qualitative Sozialforschung hingegen bereitet ihre Daten meist in Form eines schriftlichen Protokolls auf, dem Transkript. Soziale Wirklichkeit wird für die weitere Analyse in Text überführt. Je nach Forschungsziel und -methode kann sich das Transkript auf die Wiedergabe des gesprochenen Wortes beschränken oder auch weitere Aspekte einer Interaktion dokumentieren: in einer Unterrichtssituation etwa Tafelbilder sowie nonverbale Äußerungen und Gesten (Lachen, sich Melden), etc.

Generell hat diese Form der Datenaufbereitung drei wichtige Aufgaben:

- 1) *Entschleunigung*: Viele untersuchte Phänomene sind flüchtig. Sie vergehen schneller, als wir sie analysieren können. Das gilt auch und ganz besonders

für soziale Interaktionen oder Gespräche wie etwa Interviews. Um sie in Ruhe studieren zu können, müssen wir sie entschleunigen und dauerhaft festhalten. Der erste Schritt dazu ist meist eine Ton- oder Videoaufzeichnung, der zweite dann ein Transkript.

- 2) *Informationsreduktion*: Nur sehr selten werden Ton- oder Videoaufzeichnungen direkt für die Analyse verwendet. Sie enthalten zu viele Informationen, um damit sinnvoll arbeiten zu können. Ähnlich wie man beim Zeichnen eines archäologischen Lageplans entscheiden muss, was bedeutsames Artefakt und was bloß unwichtiger Staub ist, so müssen wir auch beim Transkribieren entscheiden, welche verbalen oder nonverbalen Äußerungen, welche Gesten, Geräusche, etc. für die weitere Analyse noch wichtig sein könnten und welche nicht. Transkription ist also bereits ein erster Schritt der Interpretation. Dabei ist es nicht nur die möglichst genaue und umfassende Dokumentation, die ein gutes Transkript ausmacht. Auch die Kunst des methodisch reflektierten Weglassens gehört dazu, die Konzentration auf das Wesentliche.
- 3) *Dokumentation*: Schließlich dienen Transkripte auch dazu, dass man dokumentieren und belegen kann, auf welcher Grundlage bestimmte Analysen und Schlussfolgerungen vorgenommen wurden. In einem Forschungsbericht gibt man dann Stellen aus dem Transkript als Quellen an, ähnlich wie bei einer Literaturquelle. So wird die Kette vom empirischen Ausgangsmaterial bis hin zur letztlichen Schlussfolgerung oder Theorie durchgängig transparent dokumentiert und prinzipiell auch für Andere intersubjektiv nachprüfbar.

Allerdings gibt es hier eine Einschränkung: Anders als Literaturquellen sind Transkripte aus Gründen des Datenschutzes häufig nicht öffentlich und nur auf Nachfrage auszugsweise einsehbar. Umso wichtiger ist daher, einen Forschungsbericht durch prägnante wörtliche Zitate und gute Paraphrasen anschaulich und nachvollziehbar zu gestalten (*Gütekriterien* i.d.B.). Auch dafür sind sorgfältige Transkripte wiederum eine Voraussetzung.

Bei der Planung eines Forschungsprojektes muss man beachten, dass die Transkription einen nicht unbeträchtlichen Zeitaufwand mit sich bringt. Für eine manuelle, wörtliche Transkription rechnet man mit fünf bis sechs Arbeitsstunden pro Interviewstunde. Mit KI-gestützter automatischer Spracherkennung, wie ich sie am Ende dieses Textes erläutere, lässt sich dieser Aufwand zwar deutlich reduzieren. Aber auch diese Transkripte benötigen eine sorgfältige Kontrolle und Nachbearbeitung, für die man mindestens zwei Arbeitsstunden je Interviewstunde ansetzen sollte. Sowohl bei manuellen als auch bei KI-unterstützten Transkripten können sich die genannten Zeitaufwände

stark erhöhen, wenn die Aufnahme schwer verständlich ist und/oder wenn man ein aufwändigeres Transkriptionssystem verwendet.

2 Grundlage des Transkripts: die Audio- oder Videoaufzeichnung

Noch bis in die 1970er Jahre hinein war es weit verbreitet, bei qualitativen Interviews lediglich handschriftliche Notizen zu machen (Schäffer 2022). Seitdem haben sich aber vor allem die Auswertungsmethoden stark weiterentwickelt und setzen heute in aller Regel voraus, dass man das Gesagte später wörtlich nachvollziehen und analysieren kann. Dazu ist die Aufzeichnung unumgänglich. In der Regel reicht hier eine Audioaufzeichnung. Nur wenn non-verbale Aspekte eine zentrale Rolle für die Untersuchung spielen, sollte man eine Videoaufzeichnung erwägen. Transkription und Auswertung sind dabei jedoch deutlich aufwändiger (*Videografie* i. d. B.). Ein Kompromiss kann sein, sich auf eine Audioaufnahme zu beschränken, aber wichtige Beobachtungen zu nicht-sprachlichen Ereignissen während der Untersuchungssituation zusätzlich handschriftlich festzuhalten und später in das Transkript einzufügen. Unbedingt erforderlich für jede Audio- oder Videoaufzeichnung in nichtöffentlichen Räumen ist die explizite, in der Regel schriftliche Einwilligung aller Beteiligten. Alles andere ist nicht nur forschungsethisch problematisch (*Forschungsethik* i. d. B.), sondern nach § 201 StGB sogar potenziell strafbar. Auch mit erteilter Einwilligung ist zu beachten, dass es sich bei Audio- und Videoaufzeichnungen um besonders sensible personenbezogene Daten handelt, die vor unberechtigtem Zugriff geschützt werden müssen. Dies gilt auch für die daraus erstellten Transkripte.

Die Audioqualität der Aufnahme spielt für die Transkription eine große und nicht selten unterschätzte Rolle. Jede Minute, die man hier in eine sorgfältige Planung und den Test der Technik investiert, kann sich beim späteren Transkribieren mehrfach auszahlen, weil deutlich weniger Zeit in mühsame Fehlerkorrekturen investiert werden muss. Eine gute Audioqualität ist heute kaum noch vom Aufnahmegerät abhängig. Ein normales Smartphone reicht in aller Regel aus. Viel wichtiger sind die Raumakustik sowie die Platzierung und die Qualität des Mikrofons. Dazu einige praktische Hinweise:

- Wählen Sie einen ruhigen Raum ohne Nebengeräusche.
- Schließen Sie Fenster und Türen.
- Platzieren Sie das Mikrofon nah bei der interviewten Person.
- Handelt es sich um eine Gruppendiskussion (*Gruppendiskussion* i. d. B.) oder ein anderes Setting mit mehreren Personen, kann ein externes Stereomik-

rofon sinnvoll sein. So lassen sich Redebeiträge später räumlich besser den einzelnen Sprecher:innen zuordnen, wenn Sie die Aufnahme mit Kopfhörern abhören.

- Nehmen Sie immer eine zweite Aufnahmemöglichkeit als Notfall-Backup mit, z. B. einen Laptop.

2.1 Aufzeichnung von Telefon- und Onlineinterviews

Seit der COVID-19-Pandemie werden Interviews verstärkt auch online (Zoom, Teams, etc.) oder per Telefon geführt. Was die Aufnahme betrifft, gibt es hier einige Besonderheiten zu beachten:

- Das Telefon ist für viele Menschen ein vertrautes Medium. Mitschnitte sind heute aus rechtlichen Gründen in Android und iOS aber praktisch nicht mehr möglich. Als einzige Option bleibt oft nur die Aufnahme im Lautsprechermodus mit einem zweiten Gerät, was die Tonqualität beeinträchtigen kann.
- Zoom, Teams und andere Videokonferenzdienste haben dagegen i. d. R. eine eingebaute Aufnahmefunktion. Stellen Sie in diesem Fall die Software möglichst so ein, dass die Aufnahme lokal gespeichert wird, nicht in der Cloud. Das ist datenschutzrechtlich weniger problematisch.
- Es empfiehlt sich, dass die interviewte Person ein Headset verwendet. Bei Verwendung des eingebauten Mikrofons und Lautspechers im Computer kann es zu Aussetzern in der Aufnahme kommen, wenn sonstige Geräusche übertragen werden (bspw. ein bestätigendes „Hmm“ der oder des Interviewenden).
- Auch bei Videokonferenzen müssen Sie darauf achten, dass es auf beiden Seiten möglichst wenig Nebengeräusche gibt. Instruieren Sie ihre Gesprächspartner:innen, dass sie einen ruhigen Raum mit stabiler Internetverbindung nutzen und z. B. die Fenster schließen, etc.

3 Der Prozess des Transkribierens

Je nach Forschungsfrage und Auswertungsmethode gibt es unterschiedliche Ansprüche an den Detailgrad eines Transkriptes und entsprechend auch unterschiedliche Transkriptionssysteme und -regeln. Nehmen wir als Beispiel eine Untersuchung zur Rolle von Sprachkompetenzen in Bildungskontexten. Werden hier Schüler:innen interviewt, kann es durchaus sinnvoll sein, in den Interviewtranskripten auch die dialektale Färbung, einen Akzent oder andere Sprachbesonderheiten mit zu erfassen. Werden dagegen Expert:inneninterviews geführt und es kommt vor allem auf den manifesten Inhalt des Ge-

sagten an, dann würden zusätzliche Informationen zum Dialekt der Befragten eher stören.

Grundsätzlich kann beim Detaillierungsgrad unterschieden werden zwischen:

- vollständigen oder selektiven (auszugsweisen) Transkripten sowie
- sinngemäßen, wörtlichen oder kommentierten Transkripten (Zurbriggen et al. 2018, 49f.)

Die mit Abstand häufigste Form in den Sozialwissenschaften ist das vollständige, wörtliche Transkript. Im nächsten Kapitel wird ein entsprechendes Transkriptionssystem vorgestellt. Im Unterschied zu den wörtlichen sind sinngemäße Transkripte sprachlich geglättet und fassen das Gesagte manchmal schon paraphrasierend zusammen. Kommentierte Transkripte dagegen enthalten neben dem gesprochenen Wort noch weitere Anmerkungen und spezielle Notationen zur Sprachmelodie, Lautstärke, Sprechweise, zu nonverbalen Äußerungen und vielem anderen mehr. Verbreitete Standards sind hier das „gesprächsanalytische Transkriptionssystem 2“, abgekürzt „GAT 2“ (Selting et al. 2009) oder die „Halbinterpretative Arbeitstranskription“ HIAT (Przyborski & Wohlrab-Sahr 2014, 170ff.). Sie kommen u. a. in linguistischen und konversationsanalytischen Studien zur Anwendung (*Gesprächsanalyse* i. d. B.). In der Dokumentarischen Methode der Sozialforschung (*Dokumentarische Methode* i. d. B.) wiederum ist das sogenannte „TiQ“ Transkriptionsformat (Talk in Qualitative Social Research) verbreitet, das sich insbesondere für Gruppendiskussionen (*Gruppendiskussionen* i. d. B.) sehr gut eignet (Przyborski & Wohlrab-Sahr 2014, 167ff.).

4 Die inhaltlich-semantische Transkription: Ein einfaches Regelsystem für Interviews und andere Gesprächssituationen

Im Folgenden wird mit der „inhalts-semantischen Transkription“ (Dresing & Pehl 2023) ein Regelsystem für wörtliche Transkripte vorgestellt, das für viele sozialwissenschaftliche Untersuchungszwecke und Abschlussarbeiten geeignet ist. Es orientiert sich an den über viele Jahre bewährten Empfehlungen von Thorsten Dresing und Thorsten Pehl (2023), die wiederum auf Grundlagen von Udo Kuckartz (2016) aufbauen. Das Transkriptionssystem folgt einer Reihe einfacher Regeln:

- 1) *Jedes Wort wird aufgeschrieben.* Auch grammatikalisch fehlerhafte Formulierungen, abgebrochene Sätze oder Worte werden originalgetreu transkribiert. Dialektausdrücke o. Ä. werden so gut als möglich in die Schriftsprache übertragen, ohne den semantischen Gehalt zu verfälschen.

Wie soll ich, also, so war das damals.

Wenn man etwas nicht genau versteht, wird dies in Klammern notiert und mit einem Fragezeichen gekennzeichnet. Gänzlich Unverständliches wird ebenfalls markiert: „(unverständlich)“

War eine (wilde?) Zeit.

- 2) *Füllworte wie „hmm“, „äh“, werden nicht transkribiert, außer sie transportieren eine eigene Bedeutung.* Meist haben wir es hier mit sogenannten Verlegenheitslauten zu tun, die eine kurze Denkpause überbrücken. Sie werden nicht aufgeschrieben. Das gilt auch bei reinen Bestätigungsformeln wie etwa dem typischen aufmunternden „Hhm“ vonseiten der oder des Interviewenden, das zum Weitersprechen anregen soll, aber sonst nicht in das Gespräch eingreift. Wenn ein Laut wie „hmm“ jedoch auf eine Frage antwortet (im Sinne von „Ja“) oder in anderer Weise eine eigene Bedeutung transportiert (z. B. als „hmm“ mit fragend hochgezogener Stimme am Ende), dann wird es auch transkribiert. In diesem Fall notiert man die Bedeutung in runden Klammern dahinter, z. B. „hmm (fragend)“.

I: Waren Sie da schon mit Ihrer Partnerin zusammen?

A: Hmm (bejahend).

I: Wie haben Sie sich denn kennengelernt?

- 3) *Deutliche Pausen werden festhalten.* In der Regel werden Pausen ab einer Länge von einer Sekunde markiert. In einem Gespräch ist das bereits eine auffällig lange Unterbrechung, die auch für die Analyse relevant sein kann. Pro Sekunde Stille wird ein Punkt in runden Klammern notiert, also „(...)“ für vier Sekunden Pause. Ab zehn Sekunden sollte man dies ausschreiben: „(10 Sekunden Pause)“.

Ich weiß auch nicht. (...) Vielleicht sollte ich ihn mal wieder anrufen. Ist ja immerhin mein Vater, (..) sagt er zumindest.

- 4) *Wichtige nonverbale Ausdrücke wie Lachen werden festgehalten.* Auffällige Emotionsäußerungen oder parasprachliche Elemente werden in Klammern notiert: „(lacht laut auf)“, „(schlägt auf den Tisch)“, „(flüstert)“.

Ist ja immerhin mein Vater, (..) (flüstert) sagt er zumindest.

Andere Beobachtungen und Ereignisse, die während des Gesprächs notiert werden, können ebenfalls in Klammern eingefügt werden.

(Jemand klopft an die Tür) Ja, was gibt es?

- 5) *Sprecher:innen werden gekennzeichnet.* Alle Beteiligten erhalten ein einheitliches Kürzel, bspw. die Initialen des Namens oder „I“ für Interviewende und „B“ für Befragte. Im Kopf des Transkripts steht eine Legende, die die Bedeutung der Kürzel erklärt.

I1: Interviewerin 1

I2: Interviewer 2

B: Befragte

Bei einem Sprecher:innenwechsel wird ein Absatz eingefügt und das Kürzel zu Beginn notiert.

I: Hast du noch Fragen?

6) *Transkripte bekommen automatische Zeilen- oder Absatznummern.* Zur genauen Bezeichnung einzelner Transkriptstellen im Rahmen der Analyse wird der Text mit Zeilen- oder Absatznummern versehen. Zeilennummern sind präziser und werden meist vorgezogen. Absatznummern dagegen werden vor allem in Software zur Qualitativen Datenanalyse verwendet, da sich Zeilenumbrüche (und damit auch die Zeilennummern) hier je nach Fensterbreite immer wieder ändern können. Solche Verschiebungen des Zeilenumbruchs passieren auch in einer normalen Textverarbeitung, wenn man bspw. Schriftart oder Seitenränder ändert. Deshalb sollten, wenn Zeilennummern verwendet werden, am Beginn der Analyse die Formatierung der Transkripte festgelegt und später nicht mehr geändert werden, damit die Quellenverweise gültig bleiben. Folgendermaßen kann vorgegangen werden:

- Automatische Zeilennummern in Microsoft Word: Wählen Sie im Reiter „Layout“ „Zeilenummern“ „Fortlaufend“ aus.
- Absatznummern: Markieren Sie alle Absätze und wählen Sie „Start“ „Nummerierung“ aus.

Für weitere Tipps zum Umgang mit Zeilennummern, vor allem beim Zusammenführen mehrerer Transkripte in ein Dokument, empfehle ich folgendes Videotutorial (Dröge 2022): <https://www.youtube.com/watch?v=7pJ1aUF-GUTs>.

7) *Kopf und ggf. Anhang für weitere Informationen.* Im Kopf des Transkripts sollten neben der Legende zu den Sprecher:innenkürzeln auch Ort und Dauer des Interviews bzw. der Aufzeichnung vermerkt werden, ggf. ergänzt um weitere Informationen zu den beteiligten Personen. Wenn nötig, können weitere Materialien in einen Anhang zum Transkript eingefügt werden, bspw. Fotos von Tafelbildern.

5 Die Möglichkeiten KI-basierter (halb)automatischer Transkription

Die computerbasierte Spracherkennung hat in den letzten Jahren große Fortschritte gemacht, vor allem durch den Einsatz von künstlicher Intelligenz (KI).

Heute ist eine simple automatische Spracherkennung in viele Smartphones, in Office365, Zoom oder Teams bereits eingebaut. Außerdem gibt es zahlreiche kostenlose Plattformen im Netz. Allerdings ist die Qualität dieser eingebauten und kostenlosen Spracherkennung meist nicht hinreichend für wissenschaftliche Zwecke, die Transkripte erfordern viel Nacharbeit. Außerdem ist der Datenschutz ein großes Problem, da man die kostenlose Transkription oft damit ‚bezahlt‘, dass die Anbieter:innen die sensiblen Audiodaten unbegrenzt behalten und für das Training neuer KI-Modelle nutzen dürfen. Eine Ausnahme sind hier kostenpflichtige Angebote wie f4x von audiotranskription.de aus Deutschland, töggli.ch aus der Schweiz oder die in die qualitative Analysesoftware MAXQDA integrierte Spracherkennung. Diese Programme bieten sowohl eine gute Qualität als auch Datenschutz, sind aber nicht frei verfügbar. Angesichts dieser Ausgangslage habe ich selbst im Jahr 2023 eine KI-gestützte Transkriptionssoftware namens noScribe entwickelt, die eine Qualität ähnlich den kostenpflichtigen Angeboten erreicht, aber gratis ist und komplett lokal auf dem eigenen Computer läuft (Dröge 2023). Letzteres hat für den Datenschutz große Vorteile. Voraussetzungen sind jedoch ein recht aktueller Computer (Windows/Mac/Linux) und etwas Geduld: Die automatische Transkription eines einstündigen Interviews kann durchaus drei bis fünf Stunden, auf einem älteren Rechner sogar noch deutlich länger dauern.

Die Software ist open source und viele Personen haben bereits zur Weiterentwicklung beigetragen.¹ Die eigentliche Spracherkennung basiert auf dem KI-Modell „Whisper“, das von OpenAI, der Organisation hinter ChatGPT, entwickelt und zur freien Nutzung veröffentlicht wurde (Radford et al. 2022). In einem Vergleichstest mit Interviewmaterial konnten Wollin-Giering et al. (2024) zeigen, dass diese KI aktuell die beste Transkriptionsqualität bietet und dabei auch kommerzielle Angebote teils deutlich übertrifft. Für noScribe wurde Whisper mit einem weiteren KI-Modell von Hervé Bredin (2023) kombiniert, das auf die Unterscheidung der Stimmen unterschiedlicher Sprecher:innen spezialisiert ist. Im Ergebnis ist eine speziell auf die Transkription qualitativer Interviews ausgerichtete App entstanden, die nicht nur Text erkennen, sondern ihn auch verschiedenen Sprecher:innen zuordnen kann.² Außerdem können Pausen automatisch markiert werden. NoScribe transkribiert knapp einhundert Sprachen; auch Dialekte wie etwa Schweizerdeutsch werden recht gut in die Schriftsprache übertragen. NoScribe selbst unterstützt nur einfache Transkriptionssysteme wie die oben erläuterte „inhaltlich-semantische Transkription“ oder, mit mehr Überarbeitungsaufwand, auch TiQ. Um komplexere Transkripte nach GAT 2 oder HIAT zu erstellen, lassen sich die

1 Ein besonderer Dank geht hier an Philipp Schneider für die Betreuung der macOS-Version.

2 Inzwischen sind ähnliche Programme entstanden, wie etwa „aTrain“ im Windows App-Store.

Ergebnisse jedoch in die ebenfalls freie Software EXMARaLDA übernehmen und können dort weiterbearbeitet werden.



Abb. 1: Das Programmfenster von noScribe. Das gezeigte Interview stammt aus Dröge/John 2021.

5.1 Typische Fehler KI-generierter Transkripte

Die Resultate KI-basierter Spracherkennung sind nicht fehlerfrei und müssen in jedem Fall manuell anhand der ursprünglichen Audioaufnahme geprüft und korrigiert werden. NoScribe bringt dazu einen eigenen Editor mit, in dem sich die Audioquelle passend zum Text wiedergeben lässt. F4x von audiotranskription.de und MAXQDA haben ähnliche Funktionen. Hier eine Reihe typischer Fehler, auf die man achten sollte (vgl. auch Schmidt 2024):

- Die Sprecher:innenzuordnung erfordert insbesondere bei schnellen Wechseln oder gleichzeitigem Sprechen manuelle Korrekturen.
- Eigennamen von Personen oder Organisationen, die die KI nicht kennt, werden häufig fehlerhaft transkribiert. Das gilt auch für Fachbegriffe oder andere seltene Ausdrücke, die dann teilweise fast lautsprachlich umgesetzt werden (z. B. „Stadtgolasiem“ für „Stadtgymnasien“).
- Mehrsprachigkeit in ein und derselben Audioaufnahme wird schlecht unterstützt. Wechselt eine Person während des Gesprächs in eine andere Sprache, wird der Text ggf. sogar automatisch übersetzt. Anglizismen, die

nicht sehr geläufig sind, werden ebenfalls teils fehlerhaft umgesetzt (bspw. „Buss-Wort“ statt „Buzzword“).

- Bisweilen wird der Text leicht geglättet und bspw. ein abgebrochenes Wort ausgelassen.
- Füllworte wie „hmm“ werden teils mit transkribiert, aber nicht sehr zuverlässig.
- Bisweilen kommt es zu Halluzinationen, wenn die KI Hintergrundgeräusche als Text interpretiert. Ab und an läuft die KI auch in eine Art Schleife, bei der sich dann derselbe Text endlos wiederholt. In diesem Fall ist die einzige Option, die Transkription ab dem Beginn der Schleife neu zu starten.

Insgesamt bringt KI also eine deutliche Erleichterung und Zeitersparnis, ersetzt aber nicht die sorgfältige Kontrolle und solide Kenntnis der Transkriptionsregeln und -grundsätze.

Zusammenfassung

- Zwischen Erhebung und Auswertung müssen die Daten für die Analyse aufbereitet werden. In der qualitativen Sozialforschung nutzen wir dazu ein schriftliches Protokoll, Transkript genannt.
- Der Detailgrad kann variieren. Meist werden jedoch vollständige, wörtliche Transkripte erstellt.
- Es gibt unterschiedliche Transkriptionssysteme, die festlegen, wie die Transkripte einheitlich und nachvollziehbar umgesetzt werden und welche nonverbalen Elemente miterfasst werden. Im Text finden Sie ein einfaches System erläutert. Für komplexere Systeme (GAT, HIAT, TiQ) werden entsprechende Literaturempfehlungen genannt.
- Eine gute Audioaufnahme ist für das Transkribieren unerlässlich. Eine explizite Einwilligung aller Beteiligten zur Aufnahme ist zwingend erforderlich.
- KI-gestützte Spracherkennung kann das Transkribieren erleichtern. Die Resultate sind jedoch immer vorläufig und müssen anhand der originalen Aufnahme überprüft und korrigiert werden. Im Text stelle ich meine freie Software „noScribe“ und andere Alternativen vor.

Literatur

- Bredin, H. (2023): pyannote.audio 2.1 speaker diarization pipeline: principle, benchmark, and recipe. In: 24th INTERSPEECH Conference, Dublin, Ireland, 1983-1987.
- Dresing, T. & Pehl, T. (2023): Regelsystem für die inhaltlich-semantische Transkription. Online unter: <https://www.audiotranskription.de/regeln/> (Abrufdatum: 16.05.2024).
- Dröge, K. (2022): Interviewtranskripte in Word: kleine Katastrophen, einfache Lösungen. Video. Online unter: <https://www.youtube.com/watch?v=7pj1aUFGUTs> (Abrufdatum: 16.05.2024).
- Dröge, K. (2023): noScribe: AI-powered Audio Transcription. Software. Online unter: <https://github.com/kaixxx/noScribe#readme> (Abrufdatum: 16.05.2024).
- Dröge, K. & John, M. (2021): QSF101: Ein Leitfadeninterview durchführen. Video. Online unter: <https://www.youtube.com/watch?v=fq5Utxbi35Y> (Abrufdatum: 21.05.2024).
- Kuckartz, U. (2016): Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung. Weinheim u. a.: Beltz Juventa.
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C. & Sutskever, I. (2022): Robust Speech Recognition via Large-Scale Weak Supervision. Online unter: <http://arxiv.org/pdf/2212.04356> (Abrufdatum: 12.03.2024).
- Przyborski, A. & Wohlrab-Sahr, M. (2014). Qualitative Sozialforschung. München: de Gruyter. <https://doi.org/10.1524/9783486719550>.
- Schäffer, B. (2022): „Das Medium ist die Methode“: Zur Technikgeschichte qualitativer Methoden. In: T. Fuchs, C. Demmer & C. Wiezorek (Hrsg.): Aufbrüche, Umbrüche, Abbrüche: Wegmarken qualitativer Bildungs- und Biographieforschung. Opladen: Barbara Budrich, 145-166.
- Schmidt, R. (2024): Automatische Transkriptionssoftware – ein Erfahrungsbericht. Online unter: <https://sozmethode.hypothesen.org/2315> (Abrufdatum: 03.05.2024).
- Selting, M., Auer, P., Barth-Weingarten, D., Bergmann, J., Bergmann, P., Birkner, K. et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion (10), 353-402. Online unter: <http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf> (Abrufdatum: 13.05.2024)
- Wollin-Giering, S., Hoffmann, M., Höfting, J. & Ventzke, C. (2024): Automatic Transcription of English and German Qualitative Interviews. In: Forum Qualitative Sozialforschung/Forum: Qualitative Social Research (25). <https://doi.org/10.17169/FQS-25.1.4129>.
- Zurbruggen, L., Kohler, F., Dröge, K. & Hodel, M. (2018): Qualitative Interviews meistern: Ein Fachbuch für Studierende. Berlin: Cornelsen.

Autor

Dröge, Kai, Dr.

ORCID: 0009-0004-1077-4349

Dozent und Projektleiter

Hochschule Luzern sowie Institut für Sozialforschung Frankfurt am Main

Arbeits- und Forschungsschwerpunkte: Arbeits- und Wirtschaftssoziologie, Digitalisierung und künstliche Intelligenz, Liebe und Paarsoziologie, Qualitative Forschungsmethoden

kai.droege@hslu.ch