

Wie werden Interviews mit Einzelpersonen oder Gruppen transkribiert?

Wie sieht ein einfaches Transkriptionssystem aus?

Was gibt es bei der automatischen Transkription zu beachten?

Transkriptionsregeln, Hinweise zur automatischen Transkription

Udo Kuckartz & Stefan Rädiker

Transkriptionsregeln legen fest, wie die gesprochene Sprache in die schriftliche Form übertragen wird. Informationsverluste sind mit dieser Umformung unvermeidlich verbunden. Vom Ziel und Zweck der geplanten Analyse hängt es ab, welche Verluste man für akzeptabel hält und welche nicht. Es existieren zahlreiche Transkriptionssysteme (Dittmar, 2009; Dresing & Pehl, 2018; Fuß & Karbach, 2019; Höld, 2009; Kowal & O’Connell, 2014), die sich hauptsächlich dadurch unterscheiden, ob und wie verschiedene verbale und nicht-verbale Merkmale in der Transkription berücksichtigt werden, beispielsweise Betonungen, Lautstärke, gedehntes Sprechen, Sprechpausen, Überlappungen zwischen den Äußerungen verschiedener Sprecher bzw. Sprecherinnen, Dialektfärbungen, Gestik, Mimik und paraverbale Äußerungen wie Lachen, Hüsteln, Stöhnen etc. Ferner können auch Merkmale der Interviewsituation für die Auswertung relevant sein, z.B. dass jemand den Raum betritt oder verlässt, dass ein Telefon klingelt und dergleichen mehr. Ob man all dies wirklich transkribiert, hängt nicht zuletzt von finanziellen Gesichtspunkten ab, denn wenn die Gleichzeitigkeit der Sprechenden sowie Dialektfärbung und Intonation dokumentiert werden sollen, fallen hohe Kosten für die manuelle Transkription an. Allerdings spielen nicht nur die Kosten eine Rolle, sondern entscheidend ist, welchen Grad an Genauigkeit man bei der späteren Analyse berücksichtigen kann und will. Die Maxime möglichst genau zu transkribieren kann auch die fatale Konsequenz haben, dass der Text nur mehr schwer lesbar ist – etwa im Fall von transkribierter Dialektfärbung – und die sozialwissenschaftliche Analyse hierdurch keineswegs gefördert, sondern eher behindert wird. Für die meisten Forschungsprojekte reichen relativ einfache Transkriptionssysteme völlig aus. Im Rahmen eines Evaluationsprojektes haben wir einfache und schnell erlernbare Transkriptionsregeln¹ definiert. Die Regeln sind in überarbeiteter Form (basierend auf Kuckartz & Rädiker, 2024) in der folgenden Tabelle wiedergegeben.

1 Weitere Hinweise zu komplexeren Transkriptionsregeln finden sich in Kuckartz (2010, S. 38–47).

Abb. 1: Transkriptionsregeln für die computerunterstützte Auswertung

Transkriptionsregeln
SPRECHBEITRÄGE
<p>1. Jeder Sprechbeitrag wird als eigener Absatz transkribiert. Zwischen zwei Sprechbeiträgen wird eine Leerzeile eingefügt, um die Lesbarkeit zu erhöhen.</p> <p>2. Sprechbeiträge der interviewenden oder moderierenden Person(en) werden durch „I:“ oder „M:“, die der befragten Person(en) durch eindeutige Kürzel (z. B. „B:“) eingeleitet. Zur Unterscheidung mehrerer Personen in einer Aufnahme werden die Kürzel um Nummern ergänzt („M1:“, „M2:“, „B1:“, „B2:“ etc.). Alternativ zu Kürzeln können Pseudonyme verwendet werden. Die Kennzeichnungen der Sprechenden werden zur besseren Erkennbarkeit fett gesetzt.</p>
INHALT UND GRAMMATIK
<p>3. Es wird wörtlich transkribiert, also nicht lautsprachlich oder zusammenfassend. Sprache und Interpunktion werden leicht geglättet, d. h. an das Schriftdeutsch angenähert.</p> <p>4. Die Wortstellung, bestimmte und unbestimmte Artikel etc. werden auch dann beibehalten, wenn sie grammatikalisch falsch sind. Auch Wortwiederholungen werden transkribiert.</p> <p>5. Dialekte werden möglichst genau in Hochdeutsch übersetzt, damit die Texte gut durchsucht werden können.</p> <p>6. Nicht eindeutig identifizierbare Wörter werden in Klammern mit Fragezeichen gesetzt, z. B. „(Kobold?)“. Unverständliche Wörter und Passagen werden durch „(unv.)“ kenntlich gemacht.</p>
BESTÄTIGUNGEN UND EINWÜRFE
<p>7. Zustimmungende Lautäußerungen der interviewenden oder moderierenden Person („mhm“, „aha“ etc.) werden nur transkribiert, wenn sie den Redefluss der Sprechenden deutlich unterbrechen.</p> <p>8. Kurze Einwürfe anderer Personen, wie „Ja“, „Nein“, „Genau“, werden in Klammern in den Sprechbeitrag integriert. Das Kürzel der anderen Person wird mit Gedankenstrich vorangestellt „(I – Ja)“, damit Sprechbeiträge bei der Textsuche der korrekten Person zugeordnet werden können.</p>
PAUSEN, BETONUNGEN, LAUTÄUSSERUNGEN
<p>9. Deutliche, längere Pausen werden durch in Klammern gesetzte Auslassungspunkte „(…)“ markiert. Entsprechend der Länge der Pause in Sekunden werden ein, zwei oder drei Punkte gesetzt. Bei längeren Pausen wird eine Zahl entsprechend der Dauer in Sekunden angegeben „(5)“.</p> <p>10. Besonders betonte Begriffe werden <u>unterstrichen</u>. Besonders lautes Sprechen wird durch GROßBUCHSTABEN kenntlich gemacht.</p> <p>11. Lautäußerungen werden in einfachen Klammern notiert: (lacht), (stöhnt); ebenso nonverbale Aktivitäten und Störungen: (öffnet das Fenster), (Handy klingelt).</p>
ZEITMARKEN
<p>12. Zeitmarken werden am Ende jedes Sprechbeitrags eingefügt; bei Bedarf auch bei unverständlichen Passagen in einem längeren Absatz.</p>

Dresing und Pehl führen ergänzend zu Transkriptionsregeln auch noch Regeln zur einheitlichen Schreibweise an (2018, S. 24–25), die vor allem dann unbedingt eingehalten werden sollten, wenn im Rahmen eines Forschungsprojekts viele Personen die Transkriptionen vornehmen.

Die vorgestellten Transkriptionsregeln ermöglichen eine ausreichend genaue Verschriftlichung für die Anwendung der qualitativen Inhaltsanalyse. Im Rahmen linguistischer Analyse und in der Gesprächsforschung existieren weitaus kompliziertere Transkriptionssysteme wie GAT, HIAT und CHAT (Dittmar, 2009; Rehbein et al., 2004) und mit EXMARaLDA (www.exmaralda.org) auch eine entsprechende Software.

Nicht-automatische Transkription

Zur händischen Transkription von Audio- und Videodateien am Computer stehen eine Reihe von speziellen Programmen zur Verfügung, wie z.B. EasyTranscript (www.e-werkzeug.eu) und f4/f5transkript (www.audiotranskription.de).

Auch kann man direkt mit QDA-Software wie MAXQDA, ATLAS.ti oder NVivo transkribieren. Die meisten Transkriptionsprogramme sind einfach zu handhaben und verfügen über alle Funktionen, die normalerweise für das Verschriftlichen von Interviews im Rahmen qualitativer Sozialforschung benötigt werden: Ähnlich wie bei den üblichen Mediaplayern kann man die Wiedergabe stoppen, auf Pause schalten, neu starten sowie vor- und zurückspulen. Zudem lässt sich das Rücksprungintervall einstellen, d. h. die Zeit in Sekunden, um welche die Aufnahme nach dem Anhalten zurückgesetzt wird. Es bietet sich an, beim Transkribieren einen Fußschalter zu benutzen, um die Aufzeichnung leicht stoppen und starten zu können und dadurch Zeit zu sparen.

In den meisten Fällen lässt sich die Software so einstellen, dass am Absatzbeginn oder am Absatzende automatisch eine sogenannte Zeitmarke eingefügt wird. Hierdurch wird eine Synchronisierung zwischen Text und Audioaufzeichnung ermöglicht, sodass beim Lesen einer Textstelle durch Anklicken der Zeitmarke die entsprechende Stelle der Audioaufzeichnung abgespielt wird. Die oben dargestellten Transkriptionsregeln lassen sich so ohne Probleme umsetzen.

Abb. 2 zeigt den Auszug aus einem Transkript, das unter Beachtung der Transkriptionsregeln manuell angefertigt wurde.

Abb. 2: Beispiel für ein Transkript (Auszug aus dem Interview mit B7)

B7: Ich habe, also ich habe so eine Lerngruppe mit meinem Freund. Das heißt, ich erkläre ihm alles zweimal und dann sitzt es bei mir auch. (...) Und dann noch, ja, habe ich mich noch mal mit, mit einem aus meiner Arbeitsgruppe da von Statistikgruppe getroffen. [0:09:46-6]

I: Und wie, wie fühlst du dich dabei? Also, hast du positive oder negative Einstellungen gegenüber der Statistik, oder? [0:10:02-4]

B7: Ich mag das ganz gerne. Hätte ich am Anfang auch nicht gedacht, aber ich mochte auch Mathe, und deshalb finde ich das ganz okay. [0:10:08-6]

I: Und hat sich das im Laufe des Semesters verändert? (**B7** – Ja.) Und wenn ja, wie? [0:10:15-3]

Automatische Transkription

In jüngster Zeit hat es im Bereich der automatischen Transkription von Interviews sehr große Fortschritte gegeben. Was noch vor einigen Jahren in weiter Ferne schien, ist heute Stand der Technik:

Sofern Audioaufnahmen in guter Qualität vorliegen, lassen sich diese mit Hilfe künstlicher Intelligenz automatisch verschriftlichen. Dabei werden die verschiedenen Sprechenden in Einzel- oder Gruppeninterviews erkannt, ihre Beiträge im Transkript entsprechend gekennzeichnet und mit Zeitmarken versehen.

Automatische Transkription lässt sich auf unterschiedliche Weise bewerkstelligen:

- *Webservices* erlauben es, Audio- und Videodateien hochzuladen und das fertige Transkript kann dann als Textdatei, zumeist im RTF- oder Word-Format, heruntergeladen werden. Zur Einhaltung des Datenschutzes müssen in der Regel Anbietende mit Sitz und Servern in der EU ausgewählt werden, hierzu gehören etwa AmberScript (Niederlande), f4x (Deutschland), HappyScribe (Irland) und MAXQDA Transcription (Sitz in Deutschland, Server in Niederlanden).
- *QDA-Software* wie MAXQDA, NVivo und Transana haben inzwischen eine automatische Transkription integriert. In diesem Fall kümmert sich die Software darum, die Audioaufnahme an einen Server zur Transkription zu senden, teilweise erlauben diese Programme auch die lokale Transkription auf dem eigenen Rechner.
- *Lokale Tools* ermöglichen es, die Transkription direkt auf dem eigenen Computer durchzuführen, sofern dieser leistungsstark genug ist. Beispielsweise nutzt noScribe hierfür das frei verfügbare KI-Modell Whisper von OpenAI.
- *Konferenztools* wie Microsoft Teams, Zoom und Cisco WebEx bieten Optionen für Live-Transkriptionen und im Anschluss eines Interviews kann neben der Aufzeichnung des Gesprächs gleich ein Transkript heruntergeladen werden.

Bei der Auswahl eines Tools sollte neben Aspekten des Datenschutzes berücksichtigt werden, inwieweit die oben dargestellten Transkriptionsregeln umgesetzt werden und wie detailliert das Transkript ist. Die meisten Tools erzeugen für die Sprechenden jeweils einen eigenen Absatz, aber nur einige erlauben es, Sprechpausen systematisch zu erfassen (z. B. f4x, MAXQDA, noScribe). Betonungen und Lautstärke werden unseres Wissens von keinem Tool kenntlich gemacht, ebenso werden Lautäußerungen und nonverbale Aktivitäten nicht transkribiert. In der Regel stellt der eingeschränkte Detaillierungsgrad automatischer Transkription kein Hindernis für die Durchführung einer qualitativen Inhaltsanalyse dar. Sollten die Fragestellungen jedoch einen höheren Grad an Details erfordern, kann dieser nachträglich manuell erhöht werden.

Zwar nimmt die Qualität der automatisch erzeugten Transkriptionen stetig zu, jedoch darf man nicht erwarten, dass die Technik fehlerfrei arbeitet. Automatisch erzeugte Transkripte müssen stets manuell kontrolliert und verbessert werden. Statt fünf bis zehn Stunden bei manueller Transkription ist jetzt aber deutlich weniger Zeit notwendig, um das Transkript eines einstündigen Interviews auf Richtigkeit zu prüfen und zu korrigieren. Neben dem Zeitgewinn besteht vor allem auch ein analytischer Gewinn: Erfolgt die Kontrolle eines Transkripts direkt in der QDA-Software, kann die initiierte Textarbeit bereits beim Kontrolldurchgang beginnen.

Auszug aus:

Kuckartz, Udo & Rädiker, Stefan (2024). *Qualitative Inhaltsanalyse. Methoden, Praxis, Umsetzung mit Software und künstlicher Intelligenz*. Beltz Juventa (S. 199–203)