

Wie werden Interviews und Gruppendiskussion/Fokusgruppen transkribiert?

Wie sieht ein Muster für ein einfaches Transkriptionssystem/ einfache Transkriptionsregel aus?

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; 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 Interviews oder Fokusgruppen/Gruppendiskussionen zu transkribieren ist zeitaufwändig und verursacht erhebliche Kosten, selbst für einfache Transkriptionen wird etwa das fünf- bis zehnfache der Interviewzeit benötigt. Wenn bei mehreren Interviewten oder bei einer Gruppendiskussion die Gleichzeitigkeit der Sprechenden sowie Dialektfärbung und Intonation dokumentiert werden sollen, können sich die Kosten weiter vervielfachen. 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 Transkriptionsregeln definiert. Diese Regeln sind ergänzt um zwischenzeitliche eigene Erfahrungen und um erweiternde Vorschläge von Dresing und Pehl (2018) in der folgenden Tabelle wiedergegeben. In der zweiten Spalte der Tabelle finden sich Hinweise dazu, inwieweit die Regeln von Softwaretools für die automatische Transkription umgesetzt werden: Ein „+“ bedeutet, dass die Regel üblicherweise umgesetzt wird, ein „-“ steht für „keine oder nur bedingte Umsetzung“, ein „+/-“ steht für „teilweise Umsetzung“. Auf die automatische Transkription gehen wir weiter unten in einem gesonderten Abschnitt ein.

Abb. 1: Transkriptionsregeln für die computerunterstützte Auswertung und Hinweise zu deren Umsetzung durch Software für automatische Transkription

Transkriptionsregeln	autom. Transkription
1. Jeder Sprechbeitrag wird als eigener Absatz transkribiert, auch kurze Einwüfe anderer Personen wie „Ja“, „Nein“, „Genau“. Zwischen den Sprechbeiträgen wird eine Leerzeile eingefügt, um die Lesbarkeit zu erhöhen.	+/- (kurze Einwüfe werden meist nicht erkannt)
2. Absätze 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 Namen oder Pseudonyme verwendet werden. Die Kennzeichnungen der Sprechenden werden zur besseren Erkennbarkeit fett gesetzt.	+
3. Es wird wörtlich transkribiert, also nicht lautsprachlich oder zusammenfassend. Vorhandene Dialekte werden nicht mit transkribiert, sondern möglichst genau in Hochdeutsch übersetzt, damit die Texte gut durchsucht werden können.	+/- (keine automatische Übersetzung ins Hochdeutsche)
4. Die Sprache wird leicht geglättet, das heißt an das Schriftdeutsch angenähert. Zum Beispiel wird aus „Er hatte noch so'n Buch genannt“ → „Er hatte noch so ein Buch genannt“. Die Wortstellung, bestimmte und unbestimmte Artikel etc. werden auch dann beibehalten, wenn sie Fehler enthalten.	+/- (keine Annäherung an das Schriftdeutsch)
5. Deutliche, längere Pausen werden durch in Klammern gesetzte Auslassungspunkte „(...)“ markiert.	- (ggf. erkennbar an Zeitmarken)
6. Besonders betonte Begriffe werden durch Unterstreichungen gekennzeichnet.	-
7. Sehr lautes Sprechen wird durch Schreiben in Großschrift kenntlich gemacht.	-
8. Zustimmungende bzw. bestätigende Lautäußerungen („mhm“, „aha“ etc.) werden nicht mit transkribiert, sofern sie den Redefluss der Sprechenden Person nicht unterbrechen oder als direkte Antwort auf eine Frage zu verstehen sind.	-
9. Fülllaute wie „ähm“ werden nur transkribiert, wenn ihnen eine inhaltliche Bedeutung zugemessen wird.	-
10. Störungen von außen werden unter Angabe der Ursache in Doppelklammern notiert, z.B. „((Handy klingelt))“.	-
11. Lautäußerungen werden in einfachen Klammern notiert, z.B. „(lacht)“, „(stöhnt)“ und Ähnliches.	-
12. Bei Videoaufzeichnungen von Interviews und Fokusgruppen: Nonverbale Aktivitäten werden wie Lautäußerungen in einfache Klammern gesetzt, z.B. „(öffnet das Fenster)“, „(wendet sich ab)“ und Ähnliches.	-
13. Unverständliche Wörter und Passagen werden durch „(unv.)“ kenntlich gemacht. Wörter und Passagen, bei denen der Wortlaut nur vermutet wird, werden eingeklammert und am Ende mit einem Fragezeichen versehen, z.B. „(Kobold?)“.	+/- (nur bedingt)
14. Zeitmarken werden am Ende jedes Sprechbeitrags eingefügt; bei Bedarf auch bei unverständlichen Passagen in einem längeren Absatz.	+/- (selten Zeitmarken an unverständlichen Stellen)

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. die folgenden, teilweise sogar kostenlosen Softwaretools:

- EasyTranscript (www.e-werkzeug.eu)
- ExpressScribe (www.nch.com.au/scribe/de)
- f4/f5transkript (www.audiotranskription.de)
- HyperTRANSCRIBE (www.researchware.com)
- InqScribe (www.inqscribe.com)
- Transcriber (sourceforge.net/projects/trans)

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 (Return-Taste) jeweils eine sogenannte Zeitmarke eingefügt wird. Hierdurch wird eine Synchronisierung zwischen Text und Audioaufzeichnung ermöglicht, so dass 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 oben dargestellten Transkriptionsregeln (ohne Zeitmarken) 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.

I: Und wie, wie fühlst du dich dabei? Also, hast du positive oder negative Einstellungen gegenüber der Statistik, oder?

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.

I: Und hat sich das im Laufe des Semesters verändert?

B7: Ja.

I: Und wenn ja, wie?

Automatische Transkription. In der jüngsten Zeit hat es im Bereich der automatischen Transkription von Interviews sehr große Fortschritte gegeben. Was noch vor einigen Jahren in weiter Ferne schien, ist inzwischen State-of-the-Art: Sofern Audio- und Videodateien in guter Aufnahmequalität vorliegen

und die aufgenommenen Stimmen möglichst dialektfrei sind, lassen sich diese mithilfe von künstlicher Intelligenz automatisch in analysierbare Transkripte umwandeln. Dabei werden auch Zeitmarken in den Text eingefügt, die Sprechenden identifiziert und ihre Sprechbeiträge im Transkript entsprechend gekennzeichnet. Auf diese Weise können sogar Transkripte von Fokusgruppen mit mehreren Sprechenden automatisch erzeugt werden.

Auch wenn der Markt der Anbieter und die angebotene Funktionalität einem starken Wandel unterliegt, wollen wir an dieser Stelle einige der derzeit bekannteren Anbieter nennen und haben als ersten Anhaltspunkt bezüglich des Datenschutzes auch den Sitz der Firma ergänzt: AmberScript (Niederlande), f4x (Deutschland), HappyScribe (Irland), Otter.ai (USA), TEMI (USA), Trint (UK), Sonix (USA). Einige dieser Softwarelösungen ermöglichen es, ein eigenes Vokabular zu verwalten, um die Erkennungsrate zu steigern. Der größte Unterschied zwischen den Anbietern besteht aber bezüglich der unterstützten Sprachen. Während einige Firmen nur Englisch offerieren, erlauben es andere, aus mehr als zehn verschiedenen Sprachen zu wählen. Inzwischen bieten auch Microsoft Teams, Zoom.us, Cisco WebEx und andere Tools für die Durchführung von Videokonferenzen die Möglichkeit, Gespräche live zu transkribieren und Untertitel einzublenden. Im Nachgang eines Interviews, das mit solchen Tools durchgeführt wird, kann neben der Audio- und Videoaufzeichnung ein Transkript heruntergeladen werden, das sich in QDA-Software importieren lässt.

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 5–10 Stunden bei manueller Transkription sind jetzt aber nur noch etwa 3–6 Stunden Zeitaufwand 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.

Bei all den Möglichkeiten, die sich durch automatische Transkription ergeben, dürfen zwei Punkte jedoch nicht unberücksichtigt bleiben: Erstens gilt es auf die Einhaltung des Datenschutzes zu achten, denn die typischerweise vertraulichen Audio- und Videoaufnahmen müssen in der Regel auf die Server der Anbieter hochgeladen werden. Deshalb müssen insbesondere der Serverstandort als auch die Möglichkeit, Verträge zur Auftragsdatenverarbeitung abschließen zu können, vor der Auswahl eines Transkriptions-Dienstleisters geprüft werden. Zweitens sollte berücksichtigt werden, dass die meisten Anbieter nicht auf sozialwissenschaftliche Transkripte spezialisiert sind und die automatischen Techniken nur einen Ausschnitt der analytisch relevanten Informationen erfassen. Selbst einfachste Informationen, wie Pausen im Redefluss, werden selten dargestellt, weshalb wir oben in Abb. 1 in den von uns vorgeschlagenen Transkriptionsregeln auf entsprechende Einschränkungen bei automatischer Transkription hingewiesen haben. Es muss auf Basis der Forschungsfragen entschieden werden, ob die eingeschränkte Genauigkeit automatischer Transkription für ein Forschungsprojekt akzeptabel ist.

Auszug aus:

Kuckartz, Udo & Rädiker, Stefan (2022). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Beltz Juventa (S. 199-203)