

Projektübergreifende Richtlinien für die Transkription und Segmentierung narrativer Video-Interviews

Freie Universität Berlin, Universitätsbibliothek, Center für Digitale Systeme (CeDiS), Bereich Digitale Interview-Sammlungen

Lydia Dollmann und Dorothee Wein (unter Mitarbeit von Malte Grünkorn, Peter Kompiel, Verena Nägel, Cord Pagenstecher, Lena-Luisa Stahn, Doris Tausendfreund)

Version 0.14

26.02.2020

DOI: xx

Inhaltsverzeichnis

Einführung	3
Wortgetreu, nicht lautgetreu	3
Transkriptionszeichen	4
Namen und Zahlen.....	4
Markierung der Transkriptionszeichen.....	4
Timecodes ([hh:mm:ss.ff]).....	4
Sprecher*innen-Zuordnung (INT#, KAM#)	5
Auslassungen im Transkript (<res>).....	5
Unterbrechung der Aufnahme (<i>)	5
Pausen (<p1>)	5
Undeutliche Äußerungen (<? xyz>)	5
Unverständliche Äußerungen (<?5>).....	5
Zitate, wörtliche Rede, Buch- und sonstige Titel (“xyz”)	6
Abgebrochene Wörter oder Sätze (_).....	6
Schneller Anschluss der Wörter (<=>)	6
Fremdsprachige Wörter (<l(es)>)	6
Begriffe und Redewendungen im Dialekt (<ld>).....	6
Nonverbale Kommunikation (<v>).....	6
Sprechweise (<s(lachend) xyz>, <s(laut) xyz>, <s(leise) xyz>, <s(leise) xyz>)	6
Simultansprechen (<sim>)	7
Interjektionen, Füllwörter und Verzögerungslaute	7
Anredeformen	7
Anmerkungen (<n>)	7
Ereignisse und Geräusche (<nl>).....	7
Gestik (<g>).....	8
Mimik (<m>).....	8
Empfohlene Arbeitsweise	9
1. Arbeitsgang: Vorbereitung, Gesamteindruck	9
2. Arbeitsgang: Rohtranskript, Schwerpunkt Inhalt	9
3. Arbeitsgang: Feintranskript, Grunderschließung.....	9
4. Arbeitsgang: Feintranskript, Schwerpunkt Sprechweise	10
5. Arbeitsgang: Timecodes.....	10
6. Arbeitsgang: Rechtschreibung und Zeichensetzung	10
7. Arbeitsgang: Schlusskorrektur	10
Transkriptionszeichenliste	10

Einführung

Im Bereich der Digitalen Interview-Sammlungen des Centers für Digitale Systeme (CeDiS) der Freien Universität Berlin werden Interviews erschlossen und für Wissenschaftler*innen unterschiedlicher Disziplinen zugänglich und nutzbar gemacht. So ist bei der sozialwissenschaftlichen Interpretation narrativer Interviews „nicht nur der Wortlaut des Interviews wichtig wie bei einem journalistischen Interview, sondern auch die Art wie etwas gesagt wird.“¹ Die Transkriptionen sollen die Bearbeitung einer Vielzahl von Forschungsfragen ermöglichen.

Aufbauend auf den Erfahrungen aus früheren Projekten am CeDiS, bei denen Interview-Transkriptionen für den deutschsprachigen Bestand des Visual History Archive der „USC Shoah Foundation“², für die Interview-Archive „Zwangsarbeit 1939-1945“ und „Erinnerungen an die Okkupation in Griechenland“³ angefertigt wurden, bieten die vorliegenden Regeln eine stellenweise Vereinfachung und eine Erleichterung der Übersetzung in maschinenlesbare Formate. Um in Zukunft das Austauschformat TEI⁴ nutzen zu können, wurden die Transkriptionszeichen so gewählt, dass sie inhaltlich den oben aufgeführten Anforderungen entsprechen und leicht in ein TEI-kompatibles Format übersetzt werden können.

Mit der Transkription werden Zeitstempel (Timecodes) in den Text eingefügt, sodass der Text in Segmente geteilt wird. Diese Segmentierung dient dazu, den geschriebenen Text an die Videodatei zu koppeln. Dadurch können Transkript und Übersetzung als Untertitel im laufenden Video angezeigt werden. Außerdem kann die Nutzer*in dadurch per Volltextsuche an eine spezifische Stelle des Video-Interviews springen.

Die Transkriptionen dienen folgenden Zwecken:

- Sie dienen als Grundlage für die Übersetzungen ins Deutsche bzw. andere Sprachen.
- Sie ermöglichen im Online-Archiv die Volltextsuche in den Interviews.
- Sie können als deutsche Untertitel eingeblendet werden.
- Sie werden im Online-Archiv als mitlaufende Transkription/Übersetzung angezeigt.
- Sie werden als gesonderter Lesetext (PDF) bereitgestellt.

Auf der Interviewplattform werden unterschiedlich ausführliche Versionen des Transkripts angezeigt: Die detaillierte Auszeichnung von Wörtern und Textstellen (insbesondere die Kennzeichnung nonverbaler Phänomene) wird in der ausführlichen PDF-Version angezeigt, damit die Video-Interviews nach allem Hör- und Sehbareren durchsucht und mediengerecht interpretiert werden können.

In der Untertiteldarstellung dagegen wird die Detailauszeichnung ausgeblendet, da hier die Transkription zusammen mit dem Video angesehen wird und dieses über die auditive und die visuelle Ebene alle Informationen vermittelt. Damit diese Versionen im Online-Archiv automatisch generiert werden können, ist die genaue Anwendung der Timecodes und Transkriptionszeichen mit Spitzklammern unabdingbar.

Wortgetreu, nicht lautgetreu

Die Interviews werden wortgetreu und nicht lautgetreu in phonetischer Umschrift transkribiert. Alle Eigenheiten der gesprochenen Sprache (auch Füllwörter, Verzögerungslaute, Interjektionen, Wiederholungen, Fehler) bleiben jedoch im Transkript erhalten. Die Reihenfolge und der Wortlaut des Transkripts entsprechen exakt dem, was die interviewte Person sagt. Markiert werden Wort- und Satzabbrüche, Pausen sowie Dehnlauten und laut und leise Gesprochenes. Zudem werden in der Transkription auch Mimik und Gestik des Interviewten vermerkt, jedoch nur, wenn diese sehr auffällig ist.

Für die bessere Lesbarkeit der Transkription und ihre Darstellung als Videountertitel sollen eine normale Interpunktion und grammatische Regeln angewendet werden, d. h. Fragesätze, Ausrufesätze usw. werden mit Frage- oder Ausrufezeichen gekennzeichnet.

Als Richtlinie gilt die aktuelle Standardrechtschreibung und Standardaussprache. Es gilt die Groß- und Kleinschreibung und Interpunktion nach Duden. Maßstab für die Standardaussprache sind die Laut-

¹ Ivonne Küsters, Narrative Interviews. Grundlagen und Anwendungen, Wiesbaden 2006, S. 73.

² „Die Transkription und Übersetzung der Interviews der USC Foundation“. Vgl. Woggon, Helga: Transkription und Übersetzung. Video-Interviews als Lesetexte, in: Zeugen der Shoah. Die didaktische und wissenschaftliche Arbeit mit Video-Interviews des USC Shoah Foundation Institute, Berlin 2012, https://www.zeugendershoah.de/unterrichtsmaterialien/vha_broschuere/VHAS_Broschuere_Web.pdf, S. 24-28.

³ Transkriptionsrichtlinien für das Interview-Archiv „Erinnerungen an die Okkupation in Griechenland“.

⁴ TEI ist ein von der Text Encoding Initiative entwickeltes Austauschformat zur maschinenlesbaren Kodierung von Texten. Zur Sprachtranskription gibt es die TEI Guidelines, P5, chapter 8: Transcription of Speech (<http://www.tei-c.org/release/doc/tei-p5-doc/de/html/TS.html>). Darauf basiert der Standard ISO 24624:2016 (<https://www.iso.org/standard/37338.html>), vgl. Schmidt, Thomas/Hedeland, Hanna/Jettka, Daniel: Conversion and annotation web services for spoken language data in CLARIN. Selected papers from the CLARIN Annual Conference 2016. Linköping Electronic Conference Proceedings 136, 2017, 113-130, www.ep.liu.se/ecp/article.asp?issue=136&article=009.

Buchstaben-Zuordnungen nach der amtlichen Regelung der deutschen Rechtschreibung⁵ und das Aussprachewörterbuch, z. B. ist „wia habn wieda“ die Standardaussprache für die Schreibung „wir haben wieder“ und wird also „wir haben wieder“ geschrieben.⁶

Regionalsprachliche Eigenheiten in Grammatik, Syntax etc., z. B. lateinamerikanisches „ustedes“ für „vosotros“, werden unverändert transkribiert. Regionalspezifische Aussprachen werden nicht verschriftlicht, da die Transkription wortgetreu, aber nicht lautgetreu erfolgt. Das betrifft z. B. die argentinische Aussprache bei „calle“, „yo“ etc. Hier wird in der Standardschreibweise transkribiert.

Das Interviewtranskript sollte als Lesetext, unabhängig von dem Video, verständlich und gut zu lesen sein. Für die Lesbarkeit des Transkripts ist darauf zu achten, dass die Sätze eine überschaubare Länge haben. Falls die Sprechenden zu sehr langen Sätzen neigen, müssen Transkribierende versuchen, an geeigneter Stelle einen Punkt, ein Frage- oder ein Ausrufezeichen zu setzen. Mitunter beginnt der nächste Satz dann mit „Und“ oder „Denn“ usw. Wichtig ist, dass der Satz syntaktisch unabhängig ist.

Transkriptionszeichen

Namen und Zahlen

Alle Eigennamen (von Orten, Straßen, Personen, Organisationen) werden in korrekter Form geschrieben. Wenn diese nicht bekannt sind, bitte im Internet und ggf. mit den zur Verfügung gestellten Hilfsmitteln die korrekte Schreibweise recherchieren. Ist die Schreibweise nicht feststellbar, wird der Name nur nach Gehör transkribiert und entsprechend der Regel für unsichere Transkription markiert (<? hier steht die unsichere Transkription>). Bei unvollständig ausgesprochenen Jahreszahlen wird die korrekte Jahreszahl als Anmerkung ergänzt und mit dem Kürzel <n> für notes gekennzeichnet. Ein Beispiel ist: „Der wohnte ab '77<n (1977)> am Kudamm<n (Kurfürstendamm)>“.

Zahlen, Ordnungszahlen und Daten werden grundsätzlich nach Duden geschrieben. Im Fließtext werden Zahlen bis zwölf sowie volle (runde) Zehner, Hunderter, Tausender ausgeschrieben, wenn es sich nicht um eine Maßangabe, Datumsangabe oder andere rechnerische Angaben handelt, z. B. „Da waren vielleicht dreißig Leute vor dem Haus. Ich war damals gerade 12 geworden.“ Jahreszahlen werden beispielsweise wie gesprochen geschrieben: 1980er, 80er oder Achtzigerjahre. Werden Jahreszahlen nicht vollständig ausgesprochen, wird die vollständige Zahl mit der Anmerkungs-Kennzeichnung ergänzend hinzugefügt, z. B. „Das war so dreiun-, nee, '34<n (1934)>.“ Wenn Zahlen Teil eines Begriffs/Namens sind, müssen sie in der Standardschreibweise auftauchen, z. B.: Block 10, 196. Grundschule, Drei-Zimmer-Wohnung.

Markierung der Transkriptionszeichen

Fast alle Transkriptionszeichen werden mit den als Klammern verwendeten ASCII-Zeichen „Kleiner als“ < begonnen und mit „Größer als“ > beendet (im Folgenden als „spitze Klammern“ bezeichnet). Diese Markierung des gekennzeichneten Textes ermöglicht eine maschinenlesbare Verarbeitung der Anmerkungen. Jede geöffnete Klammer muss innerhalb eines Segments, also vor dem nächsten Timecode, geschlossen werden. Innerhalb einer Klammer können neue Klammern geöffnet und geschlossen werden. Sie dürfen sich also verschachteln, aber nicht überlappen, z. B. „[hh:mm:ss.ff]INT#<s(lachend)<s(laut) Echt!>> [hh:mm:ss.ff]MM#<s(lachend)<s(laut) Ja, echt!>>“ An diesem Beispiel ist zudem ersichtlich, dass Anmerkungen in der Transkription immer in runden Klammern „()“ stehen.

Im Folgenden sind alle Auszeichnungen detailliert erklärt, am Ende dieses Dokumentes befindet sich eine Übersichtstabelle.

Timecodes ([hh:mm:ss.ff])

Vor jedem Sprecher*innenwechsel, nach jedem Satz, jeder Sinneinheit, nach maximal 120 Zeichen muss ein *Timecode* gesetzt werden. Dabei muss die Timecode-Setzung nicht Zeichengenau erfolgen. Die Angabe von maximal 120 Zeichen ist als Schätzwert zu verstehen. [hh:mm:ss.ff] ist das empfohlene Format des Zeitstempels, das Transkriptionsprogramme wie InqScribe automatisch setzen. ff entspricht den 25 Frames pro Sekunde Video. Vor dem Timecode kann ein Leerzeichen gesetzt werden, danach nicht, z. B.: [hh:mm:ss.ff]MM#<s(laut) Ja, echt!>

⁵ https://www.rechtschreibrat.com/DOX/rfdr_Regeln_2016_redigiert_2018.pdf.

⁶ Duden, Bd.6, 6. Aufl., 2005, bes. S.34 ff.

Sprecher*innen-Zuordnung (INT#, KAM#)

Jeder Sprecher*innenwechsel wird mit dem jeweiligen Initialen und einer Raute (Hashtag) markiert. Dabei steht *INT* immer für die Interviewer*in und *KAM* immer für die Kameraperson. Für die interviewte Person wird immer der erste Buchstabe vom Vor- und Nachnamen verwendet, z. B. für Maria Malta *MM*. Für jede weitere an dem Interview beteiligte Person wird ebenso der erste Buchstabe vom Vor- und Nachnamen verwendet. Die Namen und Kürzel der Beteiligten stehen im Interviewprotokoll, das die Transkribierenden erhalten.

Die Raute wird anstatt eines Doppelpunktes eingesetzt, da dieser auch an anderen Stellen im Transkript als Satzzeichen vorkommen und daher für die Kennzeichnung nicht genutzt werden kann.

Vor jedem Sprecher*innenwechsel steht ein Timecode:

[hh:mm:ss.ff]KAM# immer für die Kameraperson

[hh:mm:ss.ff]INT# immer für die Interviewer*in

[hh:mm:ss.ff]MM# erster Buchstabe vom Vor- und Nachnamen für die interviewte Person, z. B. hier *MM* für Maria Malta.

[hh:mm:ss.ff]DT# jeweils erster Buchstabe vom Vor- und Nachnamen für jede weitere am Interview beteiligte Person, z. B. *DT* für Doris Tausendfreund.

Auslassungen im Transkript (<res>)

Auslassungen sind längere Textteile, die aus rechtlichen oder ethischen Gründen nicht veröffentlicht werden sollen. Dies gilt in erster Linie für Passagen, die die Interviewten nicht veröffentlichen möchten. Sie werden im Transkript mit einem in spitze Klammern gesetzten „res“ (*restricted*) gekennzeichnet, z. B.: „[hh:mm:ss.ff]<res hier steht der Text, der ausgelassen werden soll>[hh:mm:ss.ff]“. Vor und nach der Auslassung muss ein Timecode gesetzt werden, damit der Ton entsprechend ausgeblendet oder verrauscht werden kann. Bei dieser Kennzeichnung darf vor und nach dem Timecode kein Leerzeichen gesetzt werden. Innerhalb der Ausblendung sollen keine Timecodes gesetzt werden.

Für die Nutzer*innen wird anstatt der so markierten Textsequenz folgender Text angezeigt: „Auf Wunsch des Interviewten oder aus rechtlichen Gründen wird diese Sequenz (xy Minuten) nicht veröffentlicht.“ Angaben zu gewünschten Auslassungen werden den Transkribierenden mitgeteilt.

Anonymisierung (<an>)

Die Kennzeichnung <an> wird zur Anonymisierung (*anonymity*) von personenbezogenen Daten genutzt, wie Namen, Adressen oder je nach Kontext auch weiteren zur Identifizierung nutzbaren Angaben. Dieses Zeichen wird nur verwendet, wenn bereits feststeht, dass das Interview anonymisiert werden soll. Die z. B. mit „<an Maria Malta>“ gekennzeichneten Daten sind in der für die Nutzer*innen sichtbaren Transkription unkenntlich gemacht.

Unterbrechung der Aufnahme (<i>)

Eine Unterbrechung (*interruption*) des Interviews bzw. der Aufnahme wird mit <i> gekennzeichnet. Ist der Grund der Unterbrechung bekannt, wird dies in einer Klammer angegeben, z. B. „<i(Batteriewechsel)>“. Da die Kamera in der Zeit der Unterbrechung ausgestellt wurde bzw. das Interview in einer neuen Datei fortgeführt wird, kann die Länge der Unterbrechung von den Transkripteur*innen nicht erkannt werden. Auch der Grund für eine Unterbrechung ist nicht immer bekannt, sollte aber in dem von den Interviewenden erstellten Protokoll stehen, das neben der Transkription online gestellt wird.

Pausen (<p1>)

Kommt es zu Pausen während der Erzählung, in denen nicht gesprochen wird, werden diese folgendermaßen dargestellt: <p1> für eine kurze Pause (ca. 1 Sekunde), <p2> für eine Pause von ca. zwei Sekunden, <p3> für eine Pause von ca. drei Sekunden usw. Sprechpausen bitte nur kennzeichnen, wenn sie deutlich als Sprechpausen auszumachen sind, also in der Regel erst ab 2 Sekunden.

Undeutliche Äußerungen (<? xyz>)

Ist der Inhalt der Äußerung nur schwer zu verstehen bzw. bei unsicherer Transkription, werden in spitze Klammern ohne Leerzeichen ein Fragezeichen und der vermutete Text eingesetzt: <? hier steht die unsichere Transkription>.

Unverständliche Äußerungen (<?5>)

Ist der Inhalt der Äußerung komplett unverständlich, wird dies mit einem Fragezeichen und der Zeitangabe kenntlich gemacht. Wie bei der Pause wird die Länge der unverständlichen Äußerung mit der

Angabe von Sekunden markiert, z. B. bei einem unverständlichen Abschnitt von etwa vier Sekunden: <?4>.

Zitate, wörtliche Rede, Buch- und sonstige Titel (“xyz”)

In Anführungsstriche werden die in der Erzählung genannten Zitate, wörtliche Rede, Buch- und sonstigen Titel gemäß den in den Originalsprachen geltenden Regeln gesetzt, z. B.: Die Frau sagte: “Das war eine schlimme Zeit.” oder bei Quellenangaben: “Erinnern an Zwangsarbeit. Zeitzeugen-Interviews in der digitalen Welt“.

Abgebrochene Wörter oder Sätze (_)

Ein Unterstrich kennzeichnet ein in der Erzählung abgebrochenes Wort oder einen abgebrochenen Satz. Dabei wird direkt im Anschluss an den gesprochenen Teil des Wortes oder Satzes der Unterstrich gesetzt (z. B. Ich wollte arbeiten, aber _ . Können Sie mir bitte etwas zu trinken geben?). Das gleiche Vorgehen gilt bei einem abgebrochenen Gedankengang oder einer Formulierung.

Schneller Anschluss der Wörter (<=>)

Schnell und ohne Pause hintereinander gesprochene Wörter werden mit Gleich-Zeichen markiert: Ja<=>ja.

Fremdsprachige Wörter (<l(es)>)

Aussagen oder einzelne Wörter in einer anderen Sprache als der (Haupt-)Sprache des Interviews werden mit „l“ für *lexical phenomenon* und dem Sprachkürzel nach ISO 639-1, z. B. „l(de)“ für Deutsch, „l(es)“ für Spanisch oder „l(en)“ für Englisch, gekennzeichnet. Ein Beispiel ist „There was the <l(de) Pisspott> in the corner“. Dies betrifft nicht die durchgängig zweisprachigen Interviews.

Begriffe und Redewendungen im Dialekt (<ld>)

Bei Erzählungen, in denen größtenteils Hochsprache gesprochen wird, werden im Dialekt gesprochene Wendungen mit einem kleinen L und D (*lexical phenomenon, dialect*) gekennzeichnet, z. B. „<ld Nachtigall ick hör dir trapsn.>“ Wenn die Art des Dialektes erkannt wird, soll sie vermerkt werden, z. B.: <ld(Berliner Dialekt) Nachtigall ick hör dir trapsn.>

Nonverbale Kommunikation (<v>)

Damit die für die Auswertung meist sehr wichtige nonverbale Kommunikation im Transkript berücksichtigt wird, werden die zwar nonverbalen, aber lautlich wahrnehmbaren Phänomene (*vocal, but non-lexical phenomenon*) mit <v> markiert. Die Art der nonverbalen Äußerung wie Weinen und Lachen wird in Klammern ohne Leerzeichen direkt hinter dem v markiert, z. B.: <v(Lachen)>. Mit <v(Geräusche)> werden Geräusche wie ploppen oder pfeifen markiert, die der Interviewte macht.

Beispiele für nonverbale Phänomene (v)
<v(schweres Atmen)>
<v(Schlucken)>
<v>Weinen)>
<v(Lachen)>
<v(Räuspern)>
<v(Husten)>
<v(Seufzen)>
<v(Geräusche)>

Sprechweise (<s(lachend) xyz>, <s(laut) xyz>, <s(leise) xyz>, <s(leise) xyz>)

Lachend, weinend, räuspernd usw. gesprochene Wörter (<s(lachend) xyz>)

Wird während des nonverbalen Phänomens weitergesprochen, kann dieses als Veränderung in der Art zu sprechen charakterisiert werden (*shift in vocal quality*) und mit <s> markiert. Dann werden die Wörter oder Erzählsequenzen, die die interviewte Person beispielsweise weinend oder lachend äußert, wie folgt gekennzeichnet: <s(weinend) Ich wollte das nicht> oder <s(lachend) Ich wollte das nicht>. Die Art der Sprechweise steht in Klammern direkt hinter dem <s, der gesprochene Text folgt dann nach einem Leerzeichen.

Laut und/oder betont gesprochene Wörter (<s(laut) xyz>)

Besonders laut und/oder betont gesprochene Wörter oder Sequenzen werden ebenfalls mit <s(laut)> kenntlich gemacht (*shift or changes in vocal quality*), z. B. „<s(laut) Ich wollte das nicht>.“

Gedehnt gesprochene Wörter (<s(gedehnt) xyz>)

Auch gedehnt gesprochene Wörter werden mit <s> markiert, z. B. „<s(gedehnt) Ja>, das war so.“

Leise gesprochene Wörter (<s(leise) xyz>)

Besonders leise gesprochene Wörter oder Sequenzen werden mit <s(leise)> kenntlich gemacht (*shift or changes in vocal quality*), z. B. „<s(leise) Ich wollte das nicht>.“

Beispiele für nonverbale Phänomene, wenn dabei gesprochen wird (s)
<s(betont)>
<s(gedehnt)>
<s(leise)> <s(tief)>
<s(laut)> <s(hoch)>
<s(lachend)>
<s(lächelnd)>
<s(stockend)>
<s(stotternd)>
<s(traurig)>
<s(flüsternd)>
<s(weinend)>
<s(verärgert)>
<s(erschrocken)>
<s(zittrige Stimme)>

Simultansprechen (<sim>)

Sprechen mehrere Beteiligte gleichzeitig, wird dies mit <sim> als Kürzel für *simultan* gekennzeichnet. Vor jedem Sprecher*innenwechsel wird ein Timecode gesetzt. Dadurch entsteht ein neues Segment, das – so lange gleichzeitig gesprochen wird, ebenfalls mit <sim> gekennzeichnet wird.

Beispiel: [01:23:45.67]INT#Wo waren Sie <sim während dieser Zeit?> [01:23:45.67]MM#<sim Ich war bei Maria Malta,> als das passierte. In dem Beispiel sind „während dieser Zeit?“ und „Ich war bei Maria Malta“ simultan gesprochen.

Bei längeren Passagen, in denen wiederholt mit kurzen Unterbrechungen gleichzeitig gesprochen wird, sodass eine Transkription nach dieser Regel schwer lesbar wäre, kann vor der Passage eine entsprechende Anmerkung vermerkt werden. Erstreckt sich das gleichzeitige Sprechen über mehrere Minuten, werden Anfang und Ende notiert: <sim(Anfang)> xyz <sim(Ende)>.

Interjektionen, Füllwörter und Verzögerungslaute

Interjektionen, Füllwörter und Verzögerungslaute gehören auch in die Transkription. Grundsätzlich sollte jedes gesprochene Wort mit in die Transkription aufgenommen werden. Zuma diese Wörter beispielsweise Empfindungen zum Ausdruck bringen, den Sprachrhythmus wiedergeben oder aufzeigen, wenn die Person nachdenkt.⁷ Am häufigsten sind Verzögerungslaute wie äh, ähm, die die Überlegungspausen füllen, und Interjektionen der Aufmerksamkeitsbekundung (hm, hmhm).

Interjektionen sind im Duden verzeichnet und werden nach Duden geschrieben. Nur falls es keine festgelegte Schreibweise gibt, wird nach Gehör transkribiert.

Anredeformen

Zum besseren Verständnis sollten die Anredeformen Sie, Ihnen, Ihre, Ihres, Ihrem und Ihren großgeschrieben werden. Kleingeschrieben werden (außer am Satzanfang) – alle von „du“ und „ihr“ abgeleiteten Anredeformen („du, dir, dich, dein/e, deiner/m/n, ihr, euch, euer, eure/m/n) – alle von der 3. Person Sg. fem. und 3. Person Pl. abgeleiteten Formen des Personalpronomens „sie“.

Anmerkungen (<n>)

Anmerkungen und Ergänzungen (*notes*) wie die Korrektur offensichtlicher Irrtümer werden mit <n(Erläuterung)> gekennzeichnet, z. B.: Der wohnte ab '77 <n(1977)>am Kudamm <n(Kurfürstendamm)>! Oder: nach dem Mauerbau '51 <n(gemeint ist 1961)>. Die Anmerkung steht also in runden Klammern direkt hinter dem <n.

Ereignisse und Geräusche (<nI>)

⁷ Siehe dazu: MDR AKTUELL, WARUM „ÄHM“ EINEN SINN HAT, in: MDR Wissen, 10.04.2017: <https://www.mdr.de/wissen/faszination-technik/sinn-von-aehh-100.html>.

Ereignisse und Geräusche (*non-linguistic incidents*) wie Telefonklingeln, Straßengeräusche, Vogelgezwitscher, der Eintritt einer anderen Person oder das Herunterfallen eines Gegenstands usw. werden beispielweise mit <nl(Telefonklingeln)> kenntlich gemacht. Gießt sich die interviewte Person ein Glas Wasser ein wird dieses Ereignis und Geräusch mit <nl(gießt sich ein Glas Wasser ein)>. Wenn die Person während des Interviews trinkt, sollte in der Transkription <nl(trinkt)> vermerkt werden.

Wenn dabei gesprochen wird, wird das Gesagte innerhalb der spitzen Klammern angegeben, z. B.: <nl(Telefonklingeln) Wir haben uns immer verstanden, aber>“ oder <nl(gießt sich ein Glas Wasser ein) Das war aber ganz anders>. Erstreckt sich ein Geräusch über mehrere Minuten bzw. geschriebene Zeilen, werden der Anfang des Geräuschs und sein Ende notiert: <nl(Telefonklingeln Anfang)> transkribierte gesprochene Sprache <nl (Telefonklingeln Ende)>.

Geräusche und Ereignisse
Telefonklingeln
gießt sich ein Glas Wasser ein
trinkt
putzt sich die Nase
Türklingeln
Flugzeuggeräusche
Straßengeräusche
Kindergeräusche
Gesprächsgeräusche
Vogelgezwitscher
Miauen
Bellen
undefinierbares Geräusch

Für die Geräusche sollen die unten angegebenen Begriffe verwendet werden. Die Liste ist nicht vollständig. Geräusche, die im Interview vorkommen und nicht in der Liste aufgeführt sind, sollen von den Transkribierenden den Projektverantwortlichen besprochen werden.

Gestik (<g>)

Die Gestik (*gesture*) beim Interviewten wird nur gekennzeichnet, wenn sie sehr auffällig ist, z. B. <g(verneinende Handbewegung)>, oder, wenn währenddessen gesprochen wird: <g(verneinende Handbewegung) der dachte, dass ich das machen würde>.

auffällige Gestik des Interviewten
<g(Gestikulieren)>
<g(Achselzucken)>
<g(ablehnende Handbewegung)>
<g(Kopfschütteln)>
<g(Kopfnicken)>
<g(hebt die Hände)>
<g(hebt den Finger)>
<g(zählt mit den Fingern auf)>
<g(darstellende Handbewegung)>
<g(Anführungszeichengebärde)>

Mimik (<m>)

Die Mimik (*mimic*) beim Interviewten wird ebenfalls nur gekennzeichnet, wenn sie sehr auffällig ist. Z. B.: <m(Stirnrunzeln)>, oder, wenn sie/er währenddessen spricht: <m(stirnrunzelnd) Was soll ich dazu sagen?>.

auffällige Mimik des Interviewten
<m(Stirnrunzeln)>
<m(lächelt)>
<m(rümpft die Nase)>
<m(traurig)>
<m(kämpft mit Tränen)>
<m(verzieht das Gesicht)>
<m(schließt die Augen)>

Empfohlene Arbeitsweise

Zu jedem Interview werden Interviewprotokoll, Kurzbiografie des Interviewten und weitere Informationen zur Verfügung gestellt. Dies dient dem besseren Verständnis des Gesprochenen und der korrekten Schreibweise von Namen. Im Interviewprotokoll stehen beispielsweise auch die Namen der beteiligten Personen, mit denen die Kennzeichnung im Transkript erfolgt.

Die transkribierende Person muss sehr genau hinhören, was die im Interview aufgenommenen Menschen sagen. Gute Kopfhörer sind hilfreich.

In der gesprochenen Sprache reden wir nicht in grammatisch strukturierten, ausformulierten Sätzen, sondern formen die Gedanken nach und nach zu einem „Satz“, in dem immer wieder Denk- und Formulierungspausen auftreten und Sätze nicht eindeutig und korrekt beendet werden. Dies sollte in den Transkripten berücksichtigt werden.⁸ Dennoch gilt die Form grammatischer Sätze, d. h. jeder Satz beginnt mit einem Großbuchstaben und endet mit einem Satzschlusszeichen.

Die Transkription von der gesprochenen in die geschriebene Sprache, von Lauten in Buchstaben, ist in gewisser Weise vergleichbar einer „Übersetzung“. Wie bei Übersetzungen, kommt es auch bei der Transkription darauf an, zuerst den Gesamtzusammenhang (Inhalt und Tenor der individuellen Sprech- und Erzählweisen) zur Kenntnis zu nehmen, bevor man beginnt zu transkribieren. Zudem müssen mehrere unterschiedliche Phänomene der gesprochenen Sprache in das Transkript aufgenommen werden, die man beim Zuhören schwer alle gleichzeitig beachten kann, daher werden mehrere Arbeitsdurchgänge durch das Interview empfohlen.

Allgemein empfiehlt es sich, jeweils ein Interview möglichst in kompletten Arbeitsdurchgängen zu transkribieren, anstatt sich zu lange an einzelnen Abschnitten aufzuhalten. Dadurch behalten Sie den Gesamtzusammenhang des Interviews und ihr Arbeitstempo im Blick. Wollten Sie für jede Interviewminute alle Arbeitsgänge jeweils segmentweise bis zur Schlusskorrektur erledigen, wäre es ein frustrierend langsames Vorkommen und auch zeitaufwändiger als einzelne Gesamtdurchgänge.

Teilen Sie die Gesamtaufgabe jeweils in überschaubare Arbeitsgänge und konzentrieren Sie sich bei jedem Arbeitsgang jeweils nur auf eine genau begrenzte Aufgabe. Wir empfehlen Ihnen die folgenden sechs Arbeitsgänge. Im Laufe der Arbeit werden Sie das für Ihre individuelle Arbeitsweise effektivste Vorgehen herausfinden.

1. Arbeitsgang: Vorbereitung, Gesamteindruck

Machen Sie sich mit dem Gesamt-Interview vertraut; lesen Sie die Biografie und das Interviewprotokoll. Diese Informationen helfen Ihnen beim Verständnis und bei der Schreibweise einzelner Namen, Orte oder Begriffe aus dem Interview.

2. Arbeitsgang: Rohtranskript, Schwerpunkt Inhalt

In vielen Fällen finden Sie zu Beginn Ihrer Arbeit bereits ein automatisch durch ein Spracherkennungsprogramm erstelltes – und entsprechend fehlerhaftes – Rohtranskript vor. Diese Rohversion überarbeiten Sie bei laufendem Video von Anfang bis Ende. Wenn Sie bereits einen Eindruck von den unverfälschten Stimmen der Sprechenden haben, können Sie mit Hilfe der Transkriptions-Software die Sprechgeschwindigkeit reduzieren, z. B. auf 85%, was die Stimmen relativ unverzerrt belässt. Damit können Sie beim Mitschreiben besser dem Sprechtempo folgen. Bleiben Sie dem gesprochenen Text immer möglichst nah „auf den Fersen“.

Ist der Inhalt der Äußerung nur schwer zu verstehen bzw. bei unsicherer Transkription, wird diese als unsicher <?> gekennzeichnet. In den weiteren Arbeitsgängen können diese Stellen sukzessive mit dem bislang unverstandenen Text vervollständigt werden. Wenn der Inhalt auch bei mehrmaligen Hören nicht verstanden wird, wird die Kennzeichnung beibehalten.

Falls nicht schon Timecodes im Text vorhanden sind, sollten Sie in größeren Abständen einen Timecode setzen, etwa alle zehn Minuten oder bei Arbeitspausen, damit Sie rasch wieder an die gewünschte Stelle Ihrer Transkription zurückfinden.

3. Arbeitsgang: Feintranskript, Grunderschließung

Nun ergänzen und korrigieren Sie nach und nach die weiteren Details, die die Richtlinien vorgeben. Prüfen Sie dabei in gesamten Arbeitsdurchgängen die folgenden Arbeitsschritte:

⁸ Siehe dazu auf Seite 3 die Angaben unter „Wortgetreu, nicht lautgetreu“.

- Ergänzen Sie nun nach erneutem Anhören alle Stellen, die Sie beim ersten Durchgang als unverständlich übersprungen hatten.
- Notieren Sie Sprecher*innenwechsel und Simultansprechen von Interviewten und Interviewer*in.

4. Arbeitsgang: Feintranskript, Schwerpunkt Sprechweise

- Achten Sie auf Wortabbrüche, Satzabbrüche, Pausen etc. und notieren diese mit den vorgegebenen Zeichen (s. o.).
- Vermerken Sie starke Betonungen, Dehnungen, non-verbale Phänomene und besonders auffällige Gesten oder Mimik entsprechend den Richtlinien (s. o.).

5. Arbeitsgang: Timecodes

Die Segmentierung setzt Zeitstempel (Timecodes) in den Text. Setzen Sie bei jedem Sprecher*innenwechsel und nach jedem Satz einen Timecode. In der gesprochenen Sprache ist nicht immer eindeutig, wann ein Satz beendet ist. Achten Sie dann auf die Sprachmelodie und wählen Sie eine geeignete Sinneinheit. Setzen Sie nach maximal 120 Zeichen auf jeden Fall einen Timecode, damit die daraus erzeugten Untertitel nicht mehr als zwei Zeilen umfassen.

6. Arbeitsgang: Rechtschreibung und Zeichensetzung

Nun prüfen Sie ihr Transkript auf Rechtschreibung und Zeichensetzung.

Überprüfen Sie insbesondere die Namen von Orten, Personen und Institutionen und fügen Sie ggf. die korrekten Schreibweisen als Anmerkung hinzu. Halten Sie das Ergebnis Ihrer Recherchen und offene Fragen in einer kurzen Anmerkung an das Projektteam fest.

7. Arbeitsgang: Schlusskorrektur

Prüfen Sie ihr Transkript noch einmal im Ganzen in zwei Schritten:

- Lesetext: Lesen Sie den ausgedruckten Gesamttext des Transkripts wie einen selbständigen Text, ohne das Video dabei anzuschauen. Markieren Sie beim Lesen Wörter, Sätze oder Passagen, die Ihnen unverständlich vorkommen.
- Übereinstimmung: Vergleichen Sie den Gesamttext des Transkripts mit dem Video, prüfen Sie die exakte Übereinstimmung des Wortlauts und achten Sie dabei besonders auf die als unverständlich markierten Stellen. Es kann sein, dass ein Satz unverständlich erscheint, weil Sie vergessen haben, den Abbruch eines Wortes oder eines Satzes zu markieren, eine nonverbale Besonderheit zu kommentieren, wichtige Pausen anzugeben, oder weil andere Transkriptions- oder einfache Tippfehler vorkommen.
- Unklarheiten: Überprüfen Sie alle verbliebenen Unklarheiten und notieren Sie sie für Ihre Besprechung mit dem Qualitätsmanagement.

Transkriptionszeichenliste

Transkriptionszeichen	Beispiel	Erklärung
[hh:mm:ss.ff]	[01:23:45.67]	Vor jedem Sprecher*innenwechsel, nach jedem Satz, jeder Sinneinheit, nach maximal 120 Zeichen muss ein Timecode gesetzt werden. [hh:mm:ss.ff] ist das empfohlene Format, ff entspricht den 25 Frames pro Sekunde Video.
INT#, KAM#, MM#	[01:23:45.47]INT#Wo waren Sie während dieser Zeit? [01:23:45.67]MM#Ich war bei Maria Malta, als das passierte.	INT steht immer für die Interviewer*in, KAM immer für die Kameraperson. Für die interviewte Person wird immer der erste Buchstabe vom Vor- und Nachnamen verwendet, z. B. für Maria Malta MM. Für jede weitere an dem Interview beteiligte Person wird ebenso der erste Buchstabe vom Vor- und Nachnamen verwendet. Raute/Hashtag # wird statt Doppelpunkt gesetzt, damit Sprecher*innenmarkierungen automatisch erkannt werden können.
<res>	<res hier steht der Text, der nicht veröffentlicht werden soll>	Mit <res> (<i>restricted</i>) und Timecode davor und danach werden Sequenzen gekenn-

		zeichnet, die nicht veröffentlicht werden sollen. Für die Nutzer*innen wird anstatt der so markierten Textsequenz folgender Text angezeigt: „Auf Wunsch des Interviewten oder aus rechtlichen Gründen wird diese Sequenz (xy Minuten) nicht veröffentlicht.“ Angaben zu gewünschten Auslassungen sind im Interviewprotokoll vermerkt.
<an>	Ich war bei <an Maria Malta>, als das passierte.	Mit <an> (<i>anonymous</i>) werden anonymisierte personenbezogene Daten markiert, die den Nutzer*innen mit XXX oder geschwärzt angezeigt werden.
<n>	Der wohnte ab '77 <n(1977)>am Kuddamm <n(Kurfürstendamm)>.	Anmerkungen und Ergänzungen (<i>notes</i>) wie die Korrektur offensichtlicher Irrtümer werden mit <n(Erläuterung)> gekennzeichnet.
<i>	Wo waren Sie während dieser Zeit? [01:23:45.67]<i(Batteriewechsel)> [01:23:45.67]Ich war bei Maria Malta.	Eine Unterbrechung (<i>interruption</i>) des Interviews bzw. der Aufnahme wird mit <i> gekennzeichnet. Ist der Grund der Unterbrechung bekannt, wird dies in einer Klammer angegeben, z. B. „<i(Batteriewechsel)>“.
<p1>, <p2>, <p3> usw.	Ich war <p3> bei Maria Malta, als das passierte.	Sprechpausen werden mit der Länge in Sekunden gekennzeichnet, <p1> eine Sek., <p2> zwei Sek., <p3> drei Sek. usw.
<? xyz>	Ich war bei Maria <? Meier>, als das passierte.	Eine undeutliche, schwer verständliche Passage wird mit <? hier steht die unsichere Transkription> transkribiert.
<?1>, <?2>, <?3> usw.	Ich war bei Maria <?4>, als das passierte. Ich war <?3>.	Unverständliche Passagen werden mit <? und ihrer Dauer in Sekunden markiert, <?1>, eine Sek., <?2> zwei Sek., <?3> drei Sek. usw.
“xyz“	Die Frau sagte: „Das war eine schlimme Zeit.“ Oder bei Quellenangaben: „Erinnern an Zwangsarbeit. Zeitzeugen-Interviews in der digitalen Welt“.	Zitate, wörtliche Rede, Buch- und sonstige Titel werden in Anführungszeichen gesetzt (in Deutsch „unten und oben“).
–	Ich wollte arbeiten, aber _ . Können Sie mir bitte etwas zu trinken geben?	Ein Unterstrich kennzeichnet abgebrochene Wörter oder Sätze.
<=>	Ja<=>nein<=>vielleicht.	Der schnelle Anschluss von Wörtern wird mit <=> gekennzeichnet.
<l(es) xyz>	Wo waren Sie <l(es) en este tiempo>?	Fremdsprachige Begriffe werden mit „l“ für <i>lexical phenomen</i> und dem Sprachkürzel nach ISO 639-1, z. B. „l(de)“ für Deutsch, „l(es)“ für Spanisch oder „l(en)“ für Englisch, gekennzeichnet. Ein Beispiel ist „There was the <l(de) Pisspott> in the corner“.
<ld xyz>	<ld(Berliner Dialekt) Ick kanns nich glob'n>, aber ich war bei Maria Malta.	Bei Erzählungen, in denen größtenteils Hochsprache gesprochen wird, werden im Dialekt gesprochene Wendungen mit einem kleinen L und D (<i>lexical phenomen, dialect</i>) gekennzeichnet. Wenn die Art des Dialektes erkannt wird, soll sie vermerkt werden.
<v>	<v(Lachen)>	Nonverbale, aber lautliche Äußerungen wie Lachen, Gähnen etc. (<i>vocal</i>) werden mit <v(Art der Äußerung)> gekennzeichnet. Für deren Spezifizierung gibt es eine Liste möglicher Einträge (Lachen, Gähnen etc.), die in Absprache erweitert werden kann.
<s(Sprechweise) xyz>	Wo waren Sie <s(lachend) während dieser Zeit?>	Sprechweisen wie Lautstärke, Tempo (<i>shift in vocal quality</i>) etc. werden mit <s(Art der Sprechweise) xyz> gekennzeichnet. Diese Kennzeichnung wird vorgenommen, wenn

		laut, leise, langsam, gedehnt, lachend, weinend usw. gesprochen wird.
<sim>	[01:23:45.47]INT#Wo waren Sie<sim während dieser Zeit?> [01:23:45.67]MM#<sim Ich war bei Maria Malta,> als das passierte.	Beim Simultansprechen, wenn Interviewer*in (INT) und Interviewte (hier MM) gleichzeitig sprechen, wird beides mit <sim> (<i>simultan</i>) gekennzeichnet. Vor der Sprecher*inneninitiale muss ein Timecode stehen. Erstreckt sich das gleichzeitige Sprechen über mehrere Minuten, werden Anfang und Ende notiert: <sim(Anfang)> xyz <sim(Ende)>.
<nl(Geräusch) xyz>	Wo waren <nl(Telefonklingeln) Sie während dieser> Zeit?	Geräusche wie Telefonklingeln und Straßengeräusche oder Ereignisse wie ein ins Bild laufender Mensch usw. (<i>nonlinguistic incidents</i>) werden mit <nl> kenntlich gemacht. Erstreckt sich ein Geräusch über mehrere Minuten, werden Anfang und Ende des Geräuschs notiert: <nl(Telefonklingeln Anfang)> xyz <nl(Telefonklingeln Ende)>.
<g(Gestikart) xyz>	Wo waren <g(Fingerzeigen) Sie> während dieser Zeit?	Auffällige Gestik (<i>gesture</i>) wird mit <g(Art der Geste)> gekennzeichnet. Für die eingeklammerte Art gibt es eine Liste möglicher Einträge (Kopfbewegung, Handbewegung etc.), die in Absprache erweitert werden kann.
<m(Mimikart) xyz>	<m(stirnrunzelnd) Was soll ich dazu sagen?>	Auffällige Mimik (<i>mimics</i>) wird mit <m (Art der Mimik)> gekennzeichnet. Für die eingeklammerte Art gibt es eine Liste möglicher Einträge (Stirnrunzeln etc.), die in Absprache erweitert werden kann.