Sarah Hitzler*, Daniela Böhringer

“Conversation is simply something to begin with”: Methodologische Herausforderungen durch Videodaten in der qualitativen Sozialforschung am Beispiel der Konversationsanalyse

“Conversation is simply something to begin with”: Methodological Challenges of Video Data in Qualitative Research – the Case of Conversation Analysis

<https://doi.org/10.1515/zfsoz-2021-0007>

Zusammenfassung: Neue Datenformen fordern bestehende Forschungsansätze heraus, nicht nur technisch, sondern auch theoretisch-methodologisch. Eine der technischen Neuerungen, die in jüngerer Zeit starke Bewegung in die qualitative Forschung gebracht haben, ist die zunehmende Vereinfachung bei der Erstellung, Speicherung und Bearbeitung von Videomaterial. Der Artikel stellt zentrale Herausforderungen im Umgang mit solchen „bewegten Bildern“ und ihrer Transkription und Analyse dar. Diese spezifischen Herausforderungen (Detaillierungssog, Allwissenheit und zu starke bzw. mangelnde Befremdung im Umgang mit Datenmaterial) werden am Beispiel der Konversationsanalyse herausgearbeitet. Die Konversationsanalyse stellt eines der ältesten Forschungsprogramme innerhalb des qualitativen Methodenspektrums dar. An ihrem Beispiel lässt sich zeigen, wie stark „neue“ Datenformen auf ursprüngliche methodologische Grundannahmen rückwirken und sie en passant in Frage stellen. Im Artikel wird dafür plädiert, solche Veränderungen aktiv methodologisch zu diskutieren und sie nicht schleichend geschehen zu lassen.

Schlüsselwörter: Konversationsanalyse; qualitative Sozialforschung; Methodologie; Videodaten.

Abstract: Next to technical challenges, new forms of data also present theoretical and methodological challenges to existing qualitative approaches. The development of video data and the increasing ease with which it can be produced, stored, and edited has caused significant changes in qualitative research. This article discusses central challenges which arise from transcribing and analyzing these ‘moving images’. These specific challenges (the pull of detail, the temptations of omniscience, and too high or too low degrees of distancing from the data) are demonstrated by drawing on recent developments in conversation analysis. Conversation analysis presents one of the oldest research programs within the spectrum of qualitative methods. It serves well to demonstrate how ‘new’ forms of data affect primary methodological assumptions and call them into question only in passing. The article argues that such profound challenges to methodology ought to be discussed in an active way rather than gradually letting them happen.

Keywords: Conversation Analysis; Qualitative Social Research; Methodology; Video Data.

Anmerkung: Sacks 1985: 26

*Korrespondenzautorin: Dr. Sarah Hitzler, Universität Bielefeld, Fakultät für Soziologie, Postfach 10 01 31, 33501 Bielefeld, E-Mail: sarah.hitzler@uni-bielefeld.de

Dr. Daniela Böhringer, Universität Duisburg-Essen, Fakultät für Gesellschaftswissenschaften, Forsthausweg 2, LE 528, 47057 Duisburg, E-Mail: daniela.boehringer@uni-due.de

1 Einleitung¹

Die qualitative Sozialforschung hat sich schon immer auf kreative Weise mit der Frage auseinandergesetzt, wie und mit Hilfe welcher Daten sich soziale Wirklichkeit einfangen, beschreiben, dokumentieren und analysieren lässt. Immer wieder geben dabei technische Entwicklungen Impulse von außen und irritieren damit methodologische Vorannahmen ebenso wie praktische Abläufe, wodurch sie zu Veränderungen und Weiterentwicklungen von Methoden und ihren Feldern führen. Eine der technischen Neuerungen, die in jüngerer Zeit starke Bewegung in die qualitative Forschung gebracht haben, ist die zunehmende Vereinfachung bei der Erstellung, Speicherung und Bearbeitung von Videomaterial. Vermehrter Rückgriff auf Videoaufnahmen geht mit neuen Anforderungen an die jeweiligen Forschungsmethoden einher. Am Beispiel der Konversationsanalyse, die sehr früh und sehr umfassend eine Wende zu Videodaten hin vollzogen hat, sollen in diesem Text diese methodologischen Implikationen diskutiert und sich hieraus ergebende Fragen zur Debatte gestellt werden.

Die Entwicklung von Forschungsmethoden lässt sich als ein Prozess begreifen, in welchem sich methodische, theoretisch-inhaltliche und methodologische Überlegungen mit der ganz pragmatischen Frage verflechten, welches Datenmaterial hierfür eigentlich zur Verfügung steht. In der deutschsprachigen Soziologie gibt es bekannte Beispiele dafür, wie sich Forschungsmethoden im Wechselspiel von theoretischen Überlegungen, Forschungsfragen und Datenerhebungsmöglichkeiten entwickelt haben. So beschrieb beispielsweise Fritz Schütze in den 70er Jahren das Unbehagen über die damalige Gemeindeforschung und ihre methodischen Zugänge bei der Datenerhebung und -analyse. Dieses Unbehagen habe ihn und das Forschungsteam dazu bewogen, Gemeindepolitiker nicht im engeren Sinne mittels Interviewleitfaden zu befragen, sondern sie zu sogenannten Stegreiferzählungen zu animieren (Schütze 1976: 163). In der Folge entwickelte er eine

methodologische Begründung für die sozialwissenschaftliche Relevanz von Erzählungen und einen Vorschlag, wie narrative Interviews durchzuführen und wie mit ihnen analytisch zu verfahren sei. Ein weiteres bedeutsames Beispiel für die Verschränkung von (neuen) Daten und methodischen Entwicklungen ist die Objektive Hermeneutik. So verdankt die Sequenzanalyse der Objektiven Hermeneutik ihre Entstehung in den 70er Jahren in gewisser Hinsicht der Problematik, ihren Umgang mit Daten begründen zu müssen, die bis zu jenem Zeitpunkt in der deutschsprachigen Soziologie eher unüblich waren: nämlich Audio-Aufnahmen von natürlichen Interaktionen (in Familien). Im Forschungsprojekt „Elternhaus und Schule“, das Ulrich Oevermann Anfang der 70er Jahre leitete, standen solche Aufnahmen (Gespräche zwischen Eltern und Kindern im familiären Umfeld) zur Verfügung. In einem ersten wichtigen Text zum Projekt schreiben die Autor_innen, dass „latente Sinnstrukturen der sozialisatorischen Interaktion nur durch extensive Strukturinterpretation und Sinnauslegungen beobachtbarer Interaktionen sichtbar gemacht werden können.“ (Oevermann et al. 1976: 275)

Somit gibt es in dieser Hinsicht gelegentlich eine beeindruckende Koinzidenz von theoretischem Argument und verfügbarem neuartigem Datenmaterial. Dies war auch in der Entstehung der Konversationsanalyse der Fall. Zum Zeitpunkt ihrer Entwicklung in den 60er und 70er Jahren verknüpften sich theoretische Überlegungen aus der Ethnomethodologie zur prozessualen Etablierung von sozialer Wirklichkeit und der zufällige Zugriff auf vorliegende Audio-Aufzeichnungen von Telefonaten, anhand derer diese Überlegungen sich erhärten und zu einem stabilen Forschungsprogramm präzisieren ließen (Schegloff 1992). Audio-Aufnahmen von natürlichen Gesprächen ermöglichten es, der inhärenten Methodizität von Interaktionen, wie sie sich in deren Verlauf zeigt, auf die Spur zu kommen:

I started to work with tape-recorded conversations. Such materials had a single virtue, that I could replay them. (...) It was not from any large interest in language or from some theoretical formulation of what should be studied that I started with tape-recorded conversations, but simply because I could get my hands on it. (Sacks 1985: 26)

Ausgehend von Aufnahmen von Anrufen bei einer Selbstmord-Hotline und der Audioaufnahme einer Gruppentherapiesitzung entstanden so in der Folge die ersten Arbeiten zur situativen Ordnungsleistung von Mitgliedern (zum Sprechenden-Wechsel oder zur Beendigung von Gesprächen), also zur Interaktion im Gespräch².

¹ Der Text basiert auf Vorträgen, die im Rahmen des Kongresses der Amerikanischen Gesellschaft für Soziologie (New York 2013), auf dem Kongress des International Institute for Ethnomethodology and Conversation Analysis (Mannheim 2019) sowie der Ad-hoc-Gruppe „Videoanalyse als Gesellschaftsanalyse?“ auf dem 40. DGS-Kongress (Berlin/online 2020) gehalten wurden. Wir haben außerdem von produktiven Diskussionen mit den Mitgliedern des Kolloquiums der AG Qualitative Methoden an der Universität Bielefeld und des Forschungskolloquiums Qualitative Methoden am Institut für Soziologie der TU Dresden sehr profitiert. Den Gutachter_innen und Herausgeber_innen der ZfS danken wir für weitere konstruktive und hilfreiche Hinweise.

² Jenseits der offiziell publizierten Arbeiten von Sacks, Jefferson und

2 Video-Daten in einzelnen methodischen Ansätzen der qualitativen Sozialforschung

Neue Datenformen sind aber nicht nur ein wichtiges Element in der Gründungsgeschichte von neuen Forschungsmethoden. Sie sind auch ein Impulsgeber für die Veränderung von etablierten Forschungsansätzen. Neben der Entwicklung des Internets hatte in den letzten Jahren insbesondere die zunehmende Vereinfachung bei der Erstellung, Speicherung und Bearbeitung von Videomaterial bedeutenden Einfluss auf die qualitativen Methoden. Der Einsatz von Videokameras in der Datengenerierung wird zunehmend zum Regelfall in der qualitativen Sozialforschung, sei es zur Erweiterung des Phänomenbereichs, sei es aus Überlegungen der Erhebung und Archivierung möglichst umfassender „Kontext“-Informationen. Diese Orientierung an Videomaterial erfordert in den jeweiligen Methoden Anpassungs- und Reflexionsleistungen, die einerseits innerhalb der Methodendiskurse selbst geschehen, andererseits auch methodenübergreifend debattiert werden³.

Davon berührt werden zunächst insbesondere praktische Fragen, so beispielsweise im Feld Probleme der Positionierung und Anzahl der Kameras oder des Einstellungswinkels, am Schreibtisch Probleme der Bearbeitung, Archivierung und Transkription. Dass mit diesen vordergründig praktischen Entscheidungen im Forschungsprozess auch erkenntnistheoretische Annahmen über die Rolle der Forscher_in verbunden sind, zeigt beispielhaft die Diskussion zwischen Margret Mead und Gregory Bateson (dokumentiert in Brand 1976). Die beiden waren, was den Einsatz der Kamera bei ethnographischen Feldstudien angeht, sehr unterschiedlicher Ansicht – sollte die Kamera als Handkamera „künstlerisch-blickend“ oder auf einem Stativ „neutral-registrierend“ verwendet werden?

Schegloff gab es aber auch schon Ansätze, sich mit Video-Daten zu beschäftigen. So berichtet Lynch (2019) von Seminaren zu Video-Analyse, die Sacks zu Beginn der 70er Jahre an der Universität von Kalifornien in Irvine angeboten hat. 1975 hielten Sacks und Schegloff bei der American Anthropological Association einen dann später in *Gesture* veröffentlichten Vortrag zur Home Position, bei dem sie auch Videomaterial vorspielten (Sacks & Schegloff 2002).

3 So hat sich in der DGS-Sektion „Wissenssoziologie“ eine Arbeitsgruppe zur „Visuellen Soziologie“ gebildet; am KWI in Essen ist eine „Forschergemeinschaft Visuelle Soziologie“ angesiedelt. Der Umgang mit visuellen (nicht ausschließlich Video-) Daten wurde in den letzten Jahren exemplarisch und methodologisch in Sammelpublikationen verhandelt, etwa Knoblauch et al. 2006, Schnettler und Baer 2013, Moritz und Corsten 2018.

Bateson: If you put the damn thing on a tripod, you don't get any relevance.

Mead: No, you get what happened.

Bateson: It isn't what happened.

(Brand 1976: 9)

Die Art und Weise, wie Feldforschung durch Mead oder andere Forscher_innen aus dem Kreis um Franz Boas betrieben wurde, wurde durch technische Möglichkeiten herausgefordert: Welcher Dokumentationsstatus sollte der Kamera zuerkannt werden? Hat sie einen ähnlichen Status wie die vielen Feldnotizen, Skizzen und Verwandtschaftsschemata, die Feldforscher_innen gewöhnlich aus dem Feld mitbringen? *Blickt* man mit der Kamera, wie das beispielsweise Mohn (2013) für eine so verstandene Kameraethnographie sehr pointiert herausgearbeitet hat, und richtet sie gezielt, die Situation interpretierend, auf bestimmte interessante Teilaspekte einer Situation? Oder nimmt man an, dass die Kamera lediglich ein relativ neutrales Aufzeichnungsgerät ist, das möglichst viel mitnehmen soll, ohne dass der/die Forscher_in bereits bei der Aufzeichnung eine Vorauswahl trifft?

In allen gängigen qualitativen Forschungsmethoden, nicht nur in der Ethnographie, sind Suchbewegungen feststellbar, was den Umgang mit Videodaten angeht: In der Objektiven Hermeneutik wird die Frage relevant, inwieweit intrinsisch nicht-sprachliches Material so verschriftlicht werden kann, dass es einer sequenzanalytischen Herangehensweise zugänglich wird (Corsten 2018, S. 805), oder ob es Begründungen dafür geben kann, die Datenbasis auf visuelle Elemente zu erweitern, und wie dies aussehen könnte (Herrle 2007). Die Grounded Theory Methodology diskutiert, welchen Stellenwert visuelle Elemente in der Analyse einnehmen sollen – als elaborierende Beigabe zu Interaktionen (Konecki 2011) oder als sichtbar gewordene „Resultate des Sozialen“ (Kautt 2017). Dietrich und Mey (2018) konstatieren für die Grounded Theory Methodology zwar eine „verspätete Erschließung“ audiovisueller Daten (S. 148), gehen aber davon aus, dass die Analyse von Videodaten in Analogie zur Textuntersuchung, wie sie in der GTM überwiegend praktiziert wird, vollzogen werden kann. Auch die Segmentierung relevanter Abschnitte aus dem Material, für das Kodieren zentral, erscheint ihnen nicht schwieriger als bei textförmigem Material, wenn man nur das Fokussphänomen konzentriert im Blick behält. Ungeachtet der konstatierten unproblematisch möglichen Behandlung von Videodaten mit dem Instrumentarium der Grounded Theory Methodology, geben Dietrich und Mey (2018) einem solchen empirischen Vorgehen jedoch einen eigenen Namen („Grounding Visuals“). Ähnliche Neu-Benennungen finden sich im Bereich der Ethnographie, wie die schon angesprochene „Kameraethnogra-

phie" (Mohn 2013) oder die soziologische „Videographie“ (Tuma et al. 2013).

Videoethnographische Ansätze, die oftmals Bezüge zur Ethnomethodologie herstellen, (etwa Schindler 2012; Meyer 2018; Meier zu Verl 2018) verwenden Videodaten üblicherweise zur ergänzenden Dokumentation und setzen das im Feld erworbene Teilnehmerwissen zentral für eine angemessene Analyse. Videodaten bilden dabei das Korrektiv für eine bloße rekonstruierende Registrierung. In Übereinstimmung mit Hirschauers (2001) Argument, das Soziale sei zunächst „schweigsam“ und die Leistung der Forschenden liege wesentlich darin, es zum Sprechen zu bringen, verlässt sich die Ethnographie nicht auf die Aussagekraft von Videos, sondern setzt für deren Analyse und Interpretation die kontextualisierte Verstehensleistung des/der Forschenden voraus.

Klare methodologische Positionierungen, die die Affordanzen von Videodaten kritisch unter die Lupe nehmen, kommen jedoch insgesamt eher selten vor, während die Fülle von Studien auf der Basis von Videomaterial stetig zunimmt. Die Methodik dieser Studien scheint sich oftmals an der technischen Bewältigbarkeit des Materials zu orientieren.

Das qualitative Methodenspektrum ist vielfältig und die theoretischen Vorannahmen unterschiedlich, der Versuch einer Bestimmung der konkreten Herausforderung von Videodaten für die qualitative Sozialforschung erscheint uns daher wenig zielführend. Wir möchten die besonderen Herausforderungen, die sich durch diesen „neuen“ Datentyp für ein bestehendes Forschungsprogramm ergeben, vielmehr für einen *bestimmten* Forschungsansatz erläutern. Die Konversationsanalyse erscheint uns dafür besonders instruktiv, weil sie, wie weiter unten zu sehen sein wird, schon sehr früh und sehr umfassend eine Wende von reinen Audio-Aufnahmen von Gesprächen hin zu Videodaten vollzogen hat. In diesem Forschungsbereich gibt es außerdem wie in kaum einem anderen qualitativen Ansatz eine Vielfalt von Forschungsergebnissen, die auf Video-Aufzeichnungen beruhen, technologische Entwicklungen und Transkriptionskonventionen, die bei der Handhabung dieser spezifischen Datenform helfen sollen und eben jahrzehntelange Erfahrung im analytischen Umgang mit „bewegten“ Bildern.

Wir möchten zunächst einige methodologische Grundannahmen der Konversationsanalyse skizzieren, die sich aus ihrer engen Verbundenheit mit der Ethnomethodologie ergeben, welche sich auch in der gemeinsamen Publikationstätigkeit ihrer beiden Protagonisten Harold Garfinkel und Harvey Sacks (für einen historischen Überblick: Lynch 2019) ausgedrückt hat. Videodaten stellen eine Herausforderung für diese Grund-

annahmen dar. Wie wir im Weiteren zeigen werden, zielen Videotranskripte auf denselben Wahrnehmungskanal wie Videoaufnahmen und treten damit in Konkurrenz zum Originaldatum, statt seine Analyse zu komplementieren. Darüber hinaus sind Videotranskripte notwendigerweise sowohl selektiver als auch interpretativer als Audiotranskripte, während sie gleichzeitig eine analytische Befremdung aber erschweren. In der Konsequenz ändern sich, wie zu sehen sein wird, der Umgang mit den Daten, die Identifikation von Fokusphänomenen und deren Analyse grundlegend.

3 Der ‚Video Turn‘ der Konversationsanalyse

Die Entwicklung der Konversationsanalyse (KA) in den 60er und 70er Jahren wurde als Forschungsansatz durch zentrale theoretische und soziologiekritische Argumente der Ethnomethodologie geprägt (Garfinkel 1967). Außerdem profitierte sie von den empirischen und programmatischen Arbeiten Goffmans, die die Interaktionsordnung damals als einen eigenen Gegenstand auf der soziologischen Forschungsagenda etablierten. Vor allem in den ersten konversationsanalytischen Studien kann man die Goffmansche Frage hinter der technischen Sprache, die darin vorherrscht, noch deutlich erkennen: Welche Probleme haben Mitglieder aufgrund der Tatsache, dass sie sich begegnen (z. B. die Organisation des Sprechendenwechsels) und wie lösen sie diese (Bergmann 1981)?

In ihrer Beantwortung dieser Frage schlugen Konversationsanalytiker_innen allerdings einen anderen Weg ein als Goffman. Es ging und geht ihnen weniger um das Selbst und seine Etablierung beziehungsweise Bedrohung in der direkten Begegnung mit anderen. Sie folgten eher der ethnomethodologischen Argumentation Garfinkels und konzentrierten sich auf soziale Wirklichkeit als ein reziprok und im zeitlichen Handeln situiert hergestelltes Produkt von Gesellschaftsmitgliedern. Für die KA stehen weder das forschende noch das beforschte Subjekt und deren Intentionen im Mittelpunkt, sondern etwas Drittes, eine Struktur auf Zeit, die sich in der Interaktion herausbildet und nur dort besteht. Die Interaktion zwischen Mitgliedern erfolgt methodisch und ist daher für andere – auch für Forschende – erkennbar und deutbar. Diese Methodizität der Interaktion trägt weit; sie ist ebenso kontextfrei wie kontextadaptiv (Sacks, Schegloff & Jefferson 1974: 699) und versetzt Mitglieder in die Lage, ihre alltäglichen Angelegenheiten in Begegnungen zu regeln. Der zentrale Forschungsgegenstand der KA liegt also in der Frage, wie

Mitglieder die Wirklichkeit, von der sie annehmen, dass sie Gültigkeit hat, im zeitlichen Verlauf herstellen, mit welchen Problemen sie dabei konfrontiert werden und welche methodischen Lösungen sich zeigen.

Die gemeinschaftliche Herstellung von sozialem Sinn und Ordnung wird als in situ und über den zeitlichen Verlauf von in sich geordneten Sequenzen (Rawls 2008) begriffen. Mitglieder nutzen die sequenzielle Positionierung von eigenen oder fremden Handlungszügen als zentrale Ressource, um Bedeutung und Verständigung fortlaufend herzustellen. Nur ein Beispiel hierfür bildet die sogenannte „next-turn proof procedure“ (Hutchby & Wooffitt 2008), anhand derer die Bedeutung einer Äußerung über die Reaktion darauf, nämlich den darauffolgenden Zug, abgelesen und überprüft wird. Um diese Sequenzialität in der Analyse nachvollziehen zu können, muss sie im Prozess der Datenerhebung, also bei der Aufnahme von Interaktionen und Situationen, aufrechterhalten werden.

Eben weil das Soziale in sich bereits geordnet ist, braucht es auch keine „besonderen“ Arrangements für Aufnahmen, wie das beispielsweise in Laborstudien der Fall ist. Die verwendeten Daten sind vielmehr Video- oder Audio-Aufnahmen von Situationen, die sich so auch ereignet haben könnten, ohne dass Forscher_innen anwesend gewesen wären oder daran Interesse gezeigt hätten (Potter 2002). Im Zuge der Datenerhebung und in der weiteren Behandlung (Transkription und Analyse) der Daten wird davon ausgegangen, dass auch scheinbar unbedeutende Details im Gesamtgeschehen ihren Platz haben und von Mitgliedern methodisch eingesetzt und als methodisch wahrgenommen werden.

Die ersten konversationsanalytischen Arbeiten zur Interaktion im Gespräch entstanden auf der Basis von Audio-Aufnahmen, und es bildeten sich präzise methodische Vorgehensweisen im Umgang mit solchen dynamischen und vielschichtigen Daten (insbesondere setzte Gail Jefferson Standards bei der Transkription, die bis heute gelten). Hiermit ging eine langjährige Fokussierung der Konversationsanalyse auf sprachliche Phänomene einher, was auch zu einer intensiven Rezeption ihrer Einsichten in Teilen der Linguistik führte. Dennoch war von Anfang an klar, dass sich Mitglieder nicht nur sprachlich verständlich machen, sondern sich selbst in Begegnungen ganz und gar und in strukturierter, nachvollziehbarer Weise zur Verfügung stellen.

Deshalb erschien die Hinzunahme einer weiteren Datenebene – des Visuellen – zunächst nur folgerichtig. Untermauert wurde der Wunsch, möglichst reichhaltige Daten zur Verfügung zu haben, durch einige methodologische Äußerungen von Harvey Sacks (1985), der sich die Soziologie dezidiert als eine *beobachtende* Wissen-

schaft wünschte (obwohl die damalige Analysepraxis wohl eher darin bestand, Tonbänder abzuhören). Das Besondere an den Audioaufnahmen, mit denen er arbeitete, war für Sacks, dass er ein empirisches Datum zur Hand hatte, das sich wiederholt, in immer der gleichen Form abspielen lässt und das man anderen vorlegen kann. Dass diese Daten keine vollständigen Abbilder von Situationen boten, erschien Sacks für seine analytischen Zwecke unproblematisch: „The tape-recorded materials constituted a ‘good enough’ record of what happened. Other things, to be sure, happened, but at least what was on the tape has happened“ (Sacks 1985: 26). Neben dem Hörbaren wenn möglich auch das Sichtbare auf Band zu konservieren entsprach wiederum völlig dem Anspruch, möglichst unverfälschte Abbildungen sozialer Analysen zur Grundlage sozialer Analysen zu machen⁴. Zunehmend jedoch lässt sich an der Entwicklung des konversationsanalytischen Diskurses beobachten, dass die Entscheidung für Audio- oder Video-Daten methodologische Implikationen mit einiger Reichweite mit sich bringt.

Die aktuelle Diskussion um den Umgang mit Video-daten in der Konversationsanalyse, die gegenwärtig stark durch zwei ihr eng verwandte Methoden mitgeprägt wird, geht darauf noch zu wenig ein. So versteht sich die multimodale Gesprächsforschung als konsequente Fortführung der Konversationsanalyse (Mondada & Schmitt 2010), welche bislang (auch aus praktischen Gründen) vernachlässigte Modalitäten der Bedeutungsproduktion egalitär berücksichtigen möchte und damit das Primat des Sprachlichen als durch technische Affordanzen verzerrte Schwerpunktsetzung zurückweist. In sehr präzisen und kleinteiligen Analysen arbeitet sie heraus, wie soziale Phänomene im Zusammenspiel verschiedener Modalitäten hergestellt werden. Dabei verfolgt sie unter anderem das Ziel, „die Reflexion und Modifikation konversationsanalytischer Konzepte anzuleiten“ (Schmitt 2005: 26), also insbesondere bereits etablierte Konzepte der Konversationsanalyse auf ihre multimodale Tragfähigkeit hin zu untersuchen und gegebenenfalls auszudifferenzieren. In den letzten Jahren hat dieses Forschungsprogramm sehr stark an Einfluss gewonnen und seine Interessensgebiete zunehmend auf

⁴ Ein wichtiger Hinweis dazu, der sich direkt an die Protagonist_innen der Konversationsanalyse richtete, kam schon von Erving Goffman (1981: 143). Goffman betont dort am Beispiel von zwei Automechanikern, die gemeinsam an einem Auto arbeiten, dass der Kontext für eine Äußerung nicht immer eine vorangegangene Äußerung ist, sondern möglicherweise auch ein angereichertes Werkzeug, dessen Erhalt mit „danke“ quittiert wird. Wie man mit Dingen Worte macht, kann also genauso wichtig sein und wird durch Audio-Aufnahmen allein natürlich nicht aufgezeichnet.

durch die klassische Konversationsanalyse weitgehend unbeachtete Phänomene wie etwa Nutzung von räumlichen Gegebenheiten oder die Verfügbarmachung von sensorischen Wahrnehmungen ausgeweitet. Methodologisch befasst es sich vornehmlich mit Fragen der angemessenen Transkription (Mondada 2016; Greco et al. 2019). Wie Schmitt in seinem programmatischen Text 2006 bereits vorhersieht, führt die starke Relevantsetzung des Zusammenspiels aller Modalitäten vornehmlich zur Analyse sehr kleiner Phänomene, wohingegen Fragen nach der kontextualisierten, sozialen Bedeutung interaktiver Ordnungsleistungen in den Hintergrund rücken.

Die ethnomethodologische Interaktionsanalyse, wie sie vom Lehn (2018a) in der Tradition von Heath, Hindmarsh und Luff (Luff & Heath 2012; Heath & Hindmarsh 2002; Heath & Luff 1992) und Goodwin (Goodwin 1994, 2018) ausgerufen hat, fokussiert mit dem Einsatz von Videodaten auf die interaktive Ordnungsleistung von Mitgliedern. Auch sie begreift sich als Ergänzung oder auch umfassendere Alternative zur Konversationsanalyse mit Daten, die es ermöglichen, „visuelle, körperliche und materiale Handlungen mit der Analyse von Gesprächen zu verknüpfen.“ (vom Lehn 2018b: 190) Der Fokus liegt hier allerdings darauf, solche Interaktionen analysierbar zu machen, die nicht vornehmlich sprachlich verfasst sind, sondern eine geteilte Bedeutung etwa über die Manipulation von Dingen herstellen. So entstehen zunehmend Studien, die sich mit sozialen Situationen befassen, in welchen Sprechen eine untergeordnete Rolle spielt (bspw. Meyer & Wedelstaedt 2013 zur Koordination von Boxer und Trainer im Ring oder Marstrand & Svennevig 2018 zu Berührungen im Pflegekontext). Während vom Lehn in Anlehnung an die wegweisenden Arbeiten von Heath eher die Vorzüge von Videodaten und ihre Reichhaltigkeit herausstellt, weil sie eben auch verkörperte Details von Interaktion zur Verfügung stellen, weisen beispielsweise Smith (2020) oder Mair et al. (2012), die einen stärker ethnographischen Hintergrund besitzen und Videodaten als Beobachtungen ergänzendes Material betrachten, darauf hin, dass gerade die Detailliertheit von Videodaten eine Herausforderung für die ethnomethodologische Forschung darstellt⁵.

⁵ Zunehmend gerät auch der Körper als Forschungsgegenstand und Forschungsinstrument in den Fokus (Neville 2015 spricht vom „embodied turn“), was einerseits die Notwendigkeit von Videodaten unterstreicht, andererseits auch deren Begrenztheit hervorhebt, da sie körperliche Erfahrungen nicht angemessen dokumentieren können, und so die Diskussion von ethnographischem Wissen zusätzlich relevant macht.

Dass man es bei Videodaten mit einem besonderen Datentyp zu tun hat, wird dabei in beiden Ansätzen durchaus diskutiert. Heath, Hindmarsh und Luff (2010) geben etwa in ihrem Einführungswerk in einem eigenen Kapitel Hinweise dazu, wie man Videodaten für die Veröffentlichung und die Diskussion mit Fachkolleg_innen präparieren, aufbereiten und damit zugänglich machen sollte. Auch Mondada diskutiert die vielen situierten Entscheidungen, die bei der Generierung von Videodaten getroffen werden müssen, und schließt, dass diese Daten demnach reflexive und praktische Hervorbringungsleistungen sind, mithin vielleicht eher rekonstruierende als registrierende Konserven, und dass dieser Einfluss in das Material eingeschrieben ist und analytisch berücksichtigt werden muss (Mondada 2006).⁶

Es scheint zwischen diesen Ansätzen aber Einigkeit zu bestehen, dass eine erkenntnisreiche Analyse von Interaktionen am Einsatz von Videodaten nicht mehr vorbeikommt. Das Verhältnis dieser Daten zum ursprünglichen Erkenntnisinteresse der Konversationsanalyse gerät dabei ebenso aus dem Blick wie die Reflexion über die Auswirkungen einer solchen Reorientierung auf die Analysepraxis. Im Folgenden werden wir erörtern, welche alten und neuen methodologischen Probleme demgegenüber durch den Wunsch und die technischen Möglichkeiten, soziale Situationen möglichst umfassend „einzufangen“, quasi en passant entstehen. Denn mit der zunehmenden Nutzung von Video-Aufzeichnungen und -Transkriptionen stehen unserer Auffassung nach nicht nur einfach reichhaltigere Daten („mehr vom Selben“) zur Verfügung, sondern auch ein grundsätzlich anderer Datentypus. Dies hat grundlegende Auswirkungen auf alle Phasen des konversationsanalytischen Forschungsprozesses.

4 Erhebung und Aufbereitung von Videodaten

Dass in der KA lange Zeit davon ausgegangen wurde, dass Interaktion weitgehend sprachlich verfasst ist, und dass diese sprachliche Ebene zum Verständnis der Herstellung der sozialen Situation in der Mehrzahl der Situationen völlig ausreichend sei, schlug sich in der Logik der verwendeten, sprachbasierten Transkripte nieder. Gelegent-

⁶ Die zahlreichen digitalen Hilfsmittel zur Transkription und Aufbereitung von Videodaten geben ebenfalls Hinweise darauf, dass der analytische Umgang mit Videodaten eine ganz praktische Herausforderung darstellt.

lich wurde dennoch auch die sogenannte nonverbale Kommunikation mit einbezogen, welche Phänomenbereiche wie Mimik, Gestik, Proxemik und Blickkontakt umfasste (frühe Überlegungen etwa durch Philips 1976 oder Burgoon 1980). Ob und auf welche Weise derartige Phänomene in die Transkription und Analyse einfließen, folgte in der KA im Groben Schegloffs Prinzip der „procedural consequentiality“ (Schegloff 1992), also der Maxime, dass soziale Tatsachen nur dann von Relevanz für die lokale Situation sind, wenn sich analytisch nachzeichnen lässt, dass sie von den Beteiligten relevant gemacht werden. Hatte man also den Eindruck, dass ein bestimmter Blick oder eine bestimmte Geste für das analysierte Gespräch von Bedeutung waren, wurden diese mit in die Analyse aufgenommen. Die ersten Transkripte, die nonverbale Kommunikation mit aufführten, waren somit tendenziell eklektisch. Die Anführung nonverbaler Elemente diente im Allgemeinen vor allem der Unterstützung oder der Desambiguierung der verbalen Elemente und folgten demnach, wie Ayaß (2015) formuliert, einer „additiven“ Logik.

Das Gesprächsanalytische Transkriptionssystem GAT, das 1998 durch Gesprächsforscher_innen aus der Linguistik und Soziologie entwickelt wurde, um ein auf die Eigenheiten der deutschen Sprache angepasstes Äquivalent zum im englischen Sprachraum vorherrschenden ‚Jeffersonian system‘ bereitzustellen und so die unübersichtliche Diversität von Darstellungskonventionen etwas einzuhegen, weist, ganz in diesem Geist, für die Transkription von „nicht-lautlichen, sichtbaren Anteilen der Kommunikation“ an:

Sichtbares Handeln wird immer dann transkribiert, wenn es in bedeutsamer Weise zur Eindeutigkeit – oder auch zur Uneindeutigkeit – anderer Aktivitätsebenen (wie etwa gesprochener Äußerungen) beiträgt. So sind beispielsweise Blickrichtung und Körperorientierung vor allem dort von Bedeutung, wo die Koordination der Beteiligten und die Koordinierung ihres Handelns für die gemeinsame Aktivität erforderlich ist. (Selting et al. 1998: 25–26)

Ayaß (2015: 512) spricht bei diesen Transkriptionsformen von einem *Logozentrismus*, der unhinterfragt davon ausgeht, dass soziale Situationen zuallererst sprachlich strukturiert sind, und dass sämtliche weiteren Informationen dieser Ebene höchstens sinnhaft beigeordnet sind.

Der Zugriff auf Videoaufnahmen ermöglichte es jedoch, zunehmend auch solche Phänomene in den Blick zu nehmen, die eben nicht vornehmlich sprachlich verfasst sind. Charles Goodwin gehörte zu den ersten, die in sozialen Situationen Körpern, Dingen und räumlicher Organisation ebenso viel Aufmerksamkeit schenkte wie

sprachlichen Beiträgen (etwa Goodwin 1994). Die komplexen Analysen, die solche Studien präsentierten, schärfte das Bewusstsein dafür, dass eine präanalytische Aussonderung von nicht-relevanten ‚sichtbaren‘ Phänomenen nicht verlustlos möglich ist, sondern dass „there is order at all points“ (Sacks 1992: 22) eben auch und ebenso sehr für die nicht-verbalen Handlungen von Interagierenden gilt.

Wie Ayaß (2015) nachzeichnet, haben sich diese Entwicklungen methodologisch vor allem in Debatten über angemessene Transkription niedergeschlagen. Die dabei diskutierten Varianten reichen vom additiven Transkript, in welchem einzelne Beobachtungen der zeitlichen Struktur beigeordnet werden, die sich durch die Verschriftlichung des Verbalen ergibt, bis hin zum multimodalen Transkript, in welchem verschiedene Ebenen sozialen Handelns in Partiturschreibweise aufeinander bezogen werden. Aber auch mit Formen der Visualisierung wird experimentiert, wie etwa fokussierten Zeichnungen mit Pfeilverweisen, Comics mit Sprechblasen oder komplexen überlappenden Bildfolgen, die es ermöglichen, Bewegungsabläufe darzustellen. An dieser Vielfalt und den mit ihr einhergehenden methodologischen Entscheidungen lässt sich deutlich erkennen, so Ayaß, dass Transkripte in der Konversationsanalyse keine objektiven Ausgangsdaten darstellen, sondern dass schon die Erstellung eines Transkripts durch das eigene Erkenntnisinteresse und die eigenen Hypothesen geleitet ist; Transkripte sind demnach immer bereits epistemische Objekte.

Mit der zunehmenden Selbstverständlichkeit von Videomaterial und damit zunehmenden Erfahrungen mit der materialen Technik und ihren Eigenheiten gerieten neben Fragen nach einer angemessenen Transkription insbesondere Fragen nach der angemessenen Erhebung in den Vordergrund. Die Entdeckung der Reichhaltigkeit visueller Interaktionsphänomene ging gleichzeitig mit der frustrierenden Erkenntnis einher, dass auch Videoaufnahmen immer nur Ausschnitte einer sozialen Situation einfangen können – dass Personen aus dem Bild laufen können, Kameras gegebenenfalls die Inhalte von durch die Beteiligten bearbeiteten Dokumenten nicht erfassen oder Blicke sich auf Dinge oder Gegebenheiten richten, die vom Kameraausschnitt nicht erfasst sind. Methodologische Debatten bezogen und beziehen sich demnach ausführlich auf konkrete Fragen nach dem passenden Fokus, der Positionierung der Kamera im Raum oder der Anzahl von Kameras (Luff & Heath 2012; Heath, Hindmarsh & Luff 2010). Zunehmend wird die Frage nach „Big Video“ gestellt, die der Vision folgt, eine weitgehend perfekte Verdoppelung der zu untersuchenden Situation herzustellen, die aber, ganz den konversationsanalytischen Prämissen

der Aufhebung sozialer Flüchtigkeit (Bergmann 1985) folgend, verlangsamt, gedreht und gewendet werden kann. Hierzu wird in sehr kreativer und technisch höchst beeindruckender Form mit dem Zusammenspiel verschiedener Kameratypen (360°-Kameras, Drohnen) sowie der Aufstellung mehrerer Kameras und Mikrophone experimentiert, was seinen vorläufigen Kulminationspunkt in einem virtuellen dreidimensionalen Raum gefunden hat, in dem sich die Forscherin nicht nur immersiv bewegen, sondern auch direkte Annotationen an den Phänomenen selbst vornehmen kann (McIlvenny 2019). Die Lust am Visuellen ist ungebrochen, und zunehmend herrscht in der Community die Ansicht vor, dass Daten, die die visuelle Ebene nicht berücksichtigen, im Grunde für eine aussagekräftige Analyse nicht mehr ausreichen.

5 Analytischer Umgang mit Video-Daten: Alte Probleme in neuer Prägnanz

5.1 Detaillierungssog

Schon 1985 diskutiert Bergmann Risiken, die sich aus zwei der oben skizzierten zentralen konversationsanalytischen Vorannahmen ergeben. In der Befolgung strenger Annahmen von Sequenzialität und ubiquitärer immanenter Ordnung (Sacks 1992) entsteht für Konversationsanalytiker_innen ein gewisser Sog, der geordnete Zusammenhänge im immer noch Kleineren entdecken lässt:

[D]er Interpret [kann] mit diesen methodischen Postulaten in einen Strudel der Detaillierung geraten [...], der ihn in den immer enger werdenden Trichter des inneren Sinnhorizonts einer Äußerung hinunterzieht. Die Forderung, sich in seiner Arbeit von der Maxime „Order at all points“ leiten zu lassen bzw. für „jedes Element des Textes“ eine Motivierungslinie zu explizieren, öffnet den Blick des Interpreten für die Details des registrierten Geschehens, ohne jedoch eine untere Detaillierungsgrenze festzuschreiben. So kann im Prinzip an jedem „Punkt“, an jedem Textelement der zergliedernde Blick erneut ansetzen, um auf einem bis dahin in der Interpretation übersprungenen, noch feineren Detaillierungsniveau die Suche nach einer bislang verborgenen Geordnetheit bzw. Motivierungslinie aufzunehmen. (Bergmann 1985: 315f.)

In Anlehnung an Simmel (1916) argumentiert Bergmann, dass diese „Atomisierungstendenz“ letztlich zu einer Gefahr für das Verstehen der zu analysierenden Abläufe werden kann; die Zergliederung einer Situation in zu kleine einzelne Elemente führt dazu, dass die „Erschei-

nung“, wie Simmel sie nennt, als solche nicht mehr wahrgenommen werden kann. Bergmann vermutet, dass diese Gefahr darin begründet liegt, dass die registrierende Konservierung zwar die Zeitlichkeit einer Situation zu bewahren vermag, diese aber genau hierdurch einer zentralen Eigenheit, nämlich ihrer Flüchtigkeit beraubt; Mitglieder sind aber in der Situation selbst mit genau dieser Flüchtigkeit, der Unwiederbringlichkeit der aufeinanderfolgenden Details der reziproken Herstellung der Situation konfrontiert, ihre lokale Analyse kann demnach die zeitmanipulative Perspektive der Analyse ex post niemals einnehmen. Simmel folgend sei demnach in der soziologischen Analyse darauf zu achten, dass die „Kontinuität des realen Geschehens“ nicht „auf der Strecke bleibt“ (Bergmann 1985: 316).

Der Sog des ‚Trichters des inneren Sinnhorizonts‘ gewinnt in der Auseinandersetzung mit Videodaten noch einmal an mehr Tiefe. Die potentiell zu beschreibenden Phänomene umfassen verschiedenste Kanäle: unterschiedlichste körperliche Ausdrucksformen wie Mimik, Gestik, Körperhaltung, die Positionierung im Raum, die Manipulation von Dingen, Berührungen und Blickkontakt können alle als prinzipiell situativ geordnet und aufeinander bezogen wahrgenommen (und müssen daher grundsätzlich analytisch ernstgenommen) werden. Die kleinteilige Verflochtenheit all dieser Elemente nachzuvollziehen, erfordert eine Präzision und Tiefe des Blicks, die zu einer zunehmenden Zersplitterung der Phänomene führt. In diesem Anstieg des analytischen Aufwandes und des Aufwandes bei der Transkription werden die beobachtbaren, beschreibbaren Phänomene zunehmend kleiner. Es besteht die Gefahr, dass Sacks' Maxime „take it that there is order at all points“ nicht mehr nur als ein Hinweis verstanden wird, Details nicht vorschnell als irrelevant auszuschließen, sondern eher als Aufforderung, diese Ordnung auf allen Ebenen der Detaillierung auch *nachzuweisen*. Die technischen Möglichkeiten dafür gibt es. Liegt aber der Fokus darauf, der Geordnetheit bis ins letzte nachzuspüren, wird es zunehmend schwierig, darüber hinaus im Blick zu behalten, welche Probleme hier durch die Interagierenden eigentlich gelöst werden.

5.2 Verlockungen der Allwissenheit

Vor dem Hintergrund, dass konversationsanalytische Arbeiten den Anspruch haben, als „primitive Wissenschaft“ (Lynch & Bogen 1994; Lynch 2000; Sacks 1992; Sacks 1985) die materiale Grundlage ihrer Erkenntnisprozesse offenzulegen und dadurch nachvollziehbar zu machen, galt und gilt das aufgezeichnete und transkribierte Datenmate-

rial als einzig zulässige Basis für die Analyse⁷. Ihm wird ein quasi-objektiver Status zugeschrieben, in dem Sinne, als diese epistemischen Objekte unter Wissenschaftler_innen herumgereicht und -gezeigt werden können. Mochten die Aufnahmen auch punktuell nicht vollständig sein, so war doch grundsätzlich klar, dass alles, was aufgezeichnet worden war, sich auf diese Weise ereignet hatte (Sacks 1992). Alle zusätzlichen Informationen über die zu analysierende Situation, etwa ethnographische Notizen oder Hintergrundinformationen, konnten nicht mit der gleichen Überzeugung miteinbezogen werden. Begründungen für Aussagen über Daten müssten stets „from the data themselves“ kommen, schrieben Schegloff und Sacks bereits in einem der ersten konversationsanalytischen Aufsätze (1973). Während bei Analysen von Telefongesprächen, die die ersten Daten der Konversationsanalyse darstellten, die Frage, was denn zu den „Daten“ gehöre, einigermaßen einfach zu beantworten war (nämlich alles, was von beiden Gesprächsbeteiligten wahrgenommen werden konnte, also die Wortbeiträge), stellte sich mit der Analyse kopräserter Interaktionen recht bald die Frage, ob die Beschränkung des Analysierbaren auf das Aufgezeichnete nicht möglicherweise eine deutliche Verzerrung der gemeinsam hergestellten Situation darstelle.

Die Hinzunahme von Videomaterial ermöglichte es, die analytische Rigidität aufrechtzuerhalten und dabei dennoch Ebenen der gemeinschaftlichen Herstellung von Sinn zu beachten, die sich nicht notwendig sprachlich manifestierten und so eine breitere Datenbasis zu erhalten. Gleichzeitig wurde jedoch offenbar, dass mit der Hinzunahme von Videoaufnahmen zwar die beobachtbaren Phänomene ausgeweitet wurden, gleichzeitig aber immer auch weitere Phänomene im Dunklen blieben – sei es aus zu schlechter Bildqualität (ist das wirklich ein Lächeln?), sei es aufgrund der gewählten Kameraposition, sei es aufgrund eines multifokalen Settings, das mit einer einzigen Kamera in seiner Komplexität nicht eingefangen werden konnte. Die konsequente Reaktion auf diese Erkenntnis bestand darin, durch die Hinzunahme weiterer und verschiedener Kameras die Daten weiter anzudichten. Zunehmend besteht der Standard darin, eine Interaktions-

situation vollständig und in Gänze einzufangen, wo nötig und möglich mit der Unterstützung weiterer technischer Geräte. Die Verwendung mehrerer Mikrofone sichert eine möglichst gute Tonqualität gesprochener Beiträge. Kameras fokussieren auf Bildschirme und Artefakte, wodurch Analysierenden schließlich besserer und präziserer Zugriff darauf eröffnet wird, als ihn (einige) Teilnehmende selbst haben.

In der KA mehrten sich inzwischen allerdings mahnende Stimmen, bei der Analyse nicht zu vergessen, dass Teilnehmende diese allwissende Perspektive nicht haben und auch nicht haben müssen. Mit Blick auf konversationsanalytische Studien im Kontext von Online-Begegnungen, die sich mit sehr ähnlichen methodologischen Fragen befassen müssen, wie sie hier aufgeworfen werden, warnen etwa Arminen et al. (2016) davor, der Zugänglichkeit von umfassenden und detailreichen Daten nicht die Teilnehmerperspektive zu opfern:

Caution should be preserved not to confuse the analysis of the situated interactional achievement with data on issues that have not been accessible to the participants studied and not to alienate the researcher from the members' social realities. That is, the analyst must take into account the limitations in the data available for participants in action. (Arminen et al. 2016: 304)

Die Verlockung, erhobene Details in die Analyse einfließen zu lassen, ohne zu überprüfen, inwieweit sie tatsächlich situiert und in der Begegnung für die Beteiligten eine Rolle spielen, ist groß. Damit wäre zwar grundsätzlich der Forderung entsprochen, aus „the data themselves“ heraus zu argumentieren – diese jedoch bilden möglicherweise weit mehr ab, als tatsächlich für die Interaktion von Relevanz ist und können vorschnell zur Begründung von Beobachtungen herangezogen werden („Die Lehrerin hat sich umgedreht, weil eine Schülerin in der letzten Reihe geniest hat“). Schegloffs Forderung, die „procedural consequentiality“ von Phänomenen für den weiteren Fortgang der Interaktion stets im Blick zu behalten, wird so auf neue Weise relevant. Und dann wäre eben die relevante Frage: Kann das Niesen einer Schülerin von einer Lehrerin, die etwas an die Tafel schreibt, als Anlass benutzt werden, um sich ermahnend zur Klasse umzudrehen? Ist es in diesem Sinne relevant für den weiteren Fortgang der Unterrichtsinteraktion? Oder: Kann man aus der Tatsache, dass sie sich nicht umdreht, schließen, dass sie es nicht gehört hat?

Wie man es auch dreht und wendet, es braucht ein praktikables Kriterium, um mit Video-Daten umzugehen, die tendenziell „alles“ einfangen, was in einer umgrenzten Situation (Goffman) vor sich geht und die den Analysierenden eine fast allwissende Perspektive ermöglichen (vgl. auch Bergmann 1985: 314).

⁷ Auch wenn die KA sich einer positivistischen Sprache bedient, so bleibt sie ihren ethnomethodologischen Wurzeln in gewisser Hinsicht treu, weil sie sich immer noch an konkreten, situativen Phänomenen abarbeitet, daran, wie Mitglieder etwas hier und jetzt (und nicht abstrakt in einem Modell) machen. Entscheidend sind die rohen Daten, die alles enthalten, um für „masters of natural language“ (Lynch 2000: 529) verständlich zu sein: „Scientific training should not be necessary for hearing what speakers say in an ordinary conversation, recognizing what sort of action is being accomplished and explicating how it is accomplished.“ (ebd).

An diesem Punkt mag es sinnvoll sein, auf eine Grundüberlegung zurück zu kommen, auf welcher die Konversationsanalyse aufruht. Eine zentrale These in Garfinkels Überlegungen betrifft die unheilbare Indexikalität praktischer Handlungen: die Tatsache, dass alles soziale Handeln kontextualisiert ist und nicht vollständig dekontextualisiert beschrieben oder verstanden werden kann (Garfinkel 1967: 4ff.). Die „Unheilbarkeit“ dieser Indexikalität – die Unmöglichkeit, zweifelsfrei und vollständig zu beschreiben, was eine beliebige Handlung oder Äußerung tatsächlich bedeutet – bearbeiten Mitglieder nach Garfinkel über eine Hinnahme spezifischer Vagheit. Bergmann fasst pointiert zusammen:

Genau dadurch, dass die Gesprächspartner sich vage und abwartend ausdrücken und sich mit tentativen Interpretationen begnügen, also Sinnhalte gerade *nicht* fixieren (und zwar weder als Sprecher noch als Hörer), wissen sie, worüber sie sprechen. (Bergmann 1988, SE 1: 39f.)

Interaktionsbeteiligte haben demnach in der Situation selbst weder vollständigen Zugriff auf alle relevanten Kontexte, noch benötigen sie diesen. Die Interpretationen der Situation der einzelnen Beteiligten sind keinesfalls deckungsgleich und müssen es auch nicht sein. Was in der gemeinsam hergestellten Situation relevant ist und was es bedeutet, wird in der reziproken, sequenziellen Interaktion ausgehandelt, unter großzügigen Annahmen, dass das, was das Gegenüber einbringt, schon seine Richtigkeit haben wird. Intersubjektivität muss verstanden werden als lokale Verstehensfiktion für alle praktischen Zwecke („for all practical purposes“, Garfinkel 1967: vii). Demnach findet sich das, worum es in dieser Situation geht, nicht unbedingt in den vielen Facetten des Sichtbaren und Aufzeigbaren. Es findet sich im wechselseitigen Bezug der Mitglieder aufeinander und in deren vorläufiger Einigung. So wenig, wie die Mitglieder zweifelsfrei wissen und verstehen können, worum es eigentlich geht, können Analysierende sich dies über noch präzisere Beschreibung der Situation erschließen. Insoweit die soziale Bedeutung der Situation eine praktische Aushandlungsleistung der an ihr Beteiligten ist, muss sich jegliche Analyse hieran ausrichten.

6 Neue Probleme im Umgang mit Videodaten in der Konversationsanalyse

6.1 Schwierigkeiten der Befremdung

Wie Bergmann in seinem zumindest für die deutschsprachige Konversationsanalyse sehr einflussreichen methodologischen Aufsatz „Flüchtigkeit und methodische Fixierung sozialer Wirklichkeit“ aufgezeigt hat, bieten Transkripte für die qualitative Forschung die Möglichkeit, soziale Situationen, die üblicherweise ausschließlich in der ihnen eigenen Zeitlichkeit und dem damit verknüpften Verstehensdruck zugänglich sind, dieser Flüchtigkeit zu entheben und sie damit aus einer gewissen Distanz heraus zu betrachten (Bergmann 1985). Sie dienen damit in erheblichem Maße der „Befremdung“, wie sie auch Amann & Hirschauer (1997) propagieren, wenn diese auch betonen, dass diese ebenso eine Leistung des_r geschulten ethnographischen Beobachtenden sein kann. Die Forscherin wird aus ihrem unmittelbaren, alltagsweltlichen Verständnis der Situation herausgerissen. Durch die Verlangsamung und Abstraktion wird das Geschehen in seiner aufwendigen, reziproken Konstruktion sichtbar, während die technische Registrierung gleichzeitig dafür sorgt, dass es (auf „naive“ Weise, Bergmann 1985: 301) durch dieses Verstehen unverändert, wiederholt zugänglich, anderen zeigbar bleibt. Bei der Transkription von Videodaten geschieht diese Befremdungsleistung, wie im Folgenden zu zeigen ist, zugleich zu stark und in zu geringem Umfang.

Während im Zuge des Visual/Video Turn nahelag, Videodaten auf dieselbe Weise wie Audiodaten in eine weiter fixierte Form zu überführen, stieß die Konversationsanalyse, wie oben beschrieben, bald auf große Herausforderungen. Die Kodifizierung von Transkriptionsnotationen für Videodaten gelangte schnell an Grenzen, wie sich etwa daran erkennen ließ, dass die verbesserte Neuauflage des Gesprächsanalytischen Transkriptionssystems, GAT2, ursprünglich einen Abschnitt zur Transkription von Videodaten enthalten sollte, ein Vorhaben, das jedoch schließlich aufgegeben wurde⁸. Zwar hat sich im Kontext der Multimodalen Gesprächsanalyse ein recht stabiles Notationssystem etabliert, dieses scheint jedoch von anderen Zugängen nicht übernommen zu werden. Andererseits lassen sich in der Gemeinschaft weiter intensive

⁸ Die angekündigte Revision der Notationen „visueller Komponenten“, wie sie im ursprünglichen Konzept vorgeschlagen wurde (Selting et al. 2009: 382), wurde nicht umgesetzt.

Suchbewegungen beobachten, etwa am Sonderheft *Representing data in video-based studies* der im Jahr 2018 neu gegründeten Online-Zeitschrift „Social Interaction. Video-Based Studies of Human Sociality“. Die Herausgeber skizzieren in ihrer Einleitung das positivistische Dilemma der Konversationsanalyse, ohne aber einen Ausweg zu präsentieren:

There is evidently no such thing as a perfect representation (...). But the aspiration has nevertheless always been to get closer to the "reality" by using new technologies, systems and methods for representation. Obviously, a slice of positivism is inherent in this aspiration. However, no matter what kind of sophisticated technological setup we can apply and no matter how accurate the annotation system gets, we are always dealing with interpretations. (Mortensen & Due 2019)

Das Sonderheft versammelt unterschiedliche Zugänge zur Repräsentation von visuellen Daten, die einen Einblick in die unterschiedlichen Lösungen geben sollen, welche Forschende mit Blick auf die Frage „what embodied actions to include in the transcription and how“ (Mortensen & Due 2019) entwickelt haben. Weitestgehend unreflektiert (aber s. Ayaß 2015) bleibt jedoch die Frage, *weshalb* Videotranskription eigentlich eine so hohe Diversität an Lösungen hervorbringt – anders als Audiotranskription, bei deren Konventionalisierung es im Wesentlichen lediglich darum ging, sich auf eine einheitliche Notationsform etwa für Pausen oder Betonung zu einigen, nicht aber darum, ob Pausen und Betonungen angemessene Phänomene der Transkription und Analyse waren. Dieser Frage wollen wir im Folgenden nachgehen.

6.2 Die Eigenheiten der Videotranskription

Audiotranskripte beruhen in ihrer Beschaffenheit wesentlich auf einem kulturell anerkannten, höchst differenzierten Zeichensystem: auf der Schrift. Schrift wiederum ist dadurch gekennzeichnet, dass sie eine enge (im Deutschen: phonetische) Korrespondenz mit der zugrundeliegenden Sprache aufweist (Chao 1968). Dass diese Korrespondenz nicht vollständig, sondern in vielen Bereichen konventionalisiert ist, fällt im Alltag im Allgemeinen nicht ins Gewicht: lesend nehmen wir geschriebene Wörter auf eine Weise als bedeutungstragend wahr, die ihren gesprochenen Äquivalenten entspricht. Unabhängig von der Darbietung nehmen wir **Hund**, HUND, *Hund* zunächst als Zeichen für das Konzept [Hund] wahr, ebenso wie wir das Konzept [Hund] unabhängig davon erkennen, ob es laut gesprochen, geflüstert oder besonders nasal hervorgebracht wird. Sprachtranskripte machen sich diese se-

miotische Korrespondenz zunutze, indem sie die Bedeutungsebene der Schrift beibehalten, dieser aber weitere Zeichen zuordnen, die der lautlichen, kontextualisierten Darbietung entsprechen sollen, vgl. etwa GAT2:

Die Transkription erfolgt in literarischer Umschrift und orientiert sich an der Orthographie, d. h. einer genormten Umsetzung der Lautsegmente in die Schrift. ... Sprechsprachliche Realisierungen, die der Bezugsnorm entsprechen, werden der Standardsprache entsprechend verschriftlicht, d. h. also nicht eigens notiert. (Selting et al. 2009: 360)

Die linguistische Unschärfe, die durch den weitgehenden Verzicht auf phonetische Umschrift entsteht, wird für die dadurch gewonnene unmittelbare semantische Zugänglichkeit in Kauf genommen. Audiotranskripte ordnen also einem bereits existierenden, stark konventionalisierten Zeichensystem einige wenige zusätzliche Zeichen zu, die wiederum ebenfalls recht stark konventionalisiert sind. Sie ermöglichen so eine klare Distanzierung und Befremdung durch eine *ungewöhnliche*, ausgebaute Darstellungsweise; gleichzeitig bewahren sie die Integrität des in der Originalsituation hergestellten koproduzierten semantischen Sinns, indem sie diesen weitgehend verlustlos übertragen (trans-skribieren). Die Transkripte bieten, wie von Bergmann (1985) beschrieben, gegenüber der Originalaufnahme eine Aufhebung der Flüchtigkeit und eine Distanzierung zum Geschehen, gleichzeitig bleibt der unmittelbar zugängliche Sinn jedoch weitgehend bewahrt. Die reduzierte Anzahl und stark konventionalisierte Form zusätzlicher Darstellungsregeln ermöglichen es geübten Analytiker_innen, Audiotranskripte ähnlich flüssig zu verarbeiten, wie etwa Musiker_innen in der Lage sind, Noten zu lesen. Audiotranskripte, die entlang der in der KA üblichen Systematiken erstellt wurden, haben einen mittleren Verfremdungsgrad, der Analytiker_innen aus ihrer „transformierenden Rekonstruktionspraxis“ des Alltags (Bergmann 1985: 310) zwar herausreißt, es ihnen aber immer noch ermöglicht nachzuvollziehen, um was es geht und wie dieses Faktum in der Interaktion prozedural entsteht.

Sollen nun neben sprachlichen (plus parasprachlichen) auch visuelle Komponenten im Transkript erfasst werden, erhöht sich die Komplexität des Produkts massiv. Ayaß stellt in ihrer Diskussion um den epistemischen Status von Transkripten für die Konversationsanalyse fest: „In transcribing audiovisual material, the time spent on the type of work does not simply increase; it explodes.“ (Ayaß 2015: 515) Dies liegt daran, dass ‚das Visuelle‘ nicht auf gleiche Weise eingrenzbar ist wie ‚das Sprachliche‘. Es umfasst grundsätzlich alles, was auf der Kameraeinstellung zu sehen ist – beispielsweise räumliche Organisation, Gesten, Körperhaltung, Berührungen, Mimik, Be-

wegungen, Manipulation von Gegenständen, vieles davon in simultanem Zusammenspiel. Darüber hinaus sind die einzelnen Kanäle nicht in klar abgegrenzte Phänomene unterteilbar; wo etwa eine Geste beginnt oder wo sie endet, ob es sich bei einer Bewegung um eine oder zwei ineinandergreifende Gesten handelt, ob Kopfrichtung und Blickrichtung als Einheit zu behandeln sind oder als einander ergänzende Phänomene zu gelten haben, sind aktive, analytisch geleitete Entscheidungen der transkribierenden Person⁹. Nicht nur in der Analyse, auch in der Transkription selbst verweigert sich das Originaldatum somit den semantischen Abkürzungen, die für Sprachtranskripte zur Verfügung stehen. Es wird eben nicht transkribiert, es müssen (sprachliche) Deskriptionen oder (visuelle) Repräsentationen vorgenommen werden, welche das, was im zeitlichen Fluss und in vielen simultan ablaufenden Ebenen geschieht, in möglichst präzise Zeichensysteme überführen sollen. Für die Abfassung sprachlicher Beschreibungen nicht-sprachlicher Handlungen ist wiederum eine „alltägliche Rekonstruktionspraxis“ unumgänglich (Bergmann 1985: 310).

Damit werden die aufbereiteten Daten um ein Vielfaches komplexer als die Originalaufnahme, deren inhärenter Sinnzusammenhang den Beteiligten und Beobachtenden im Allgemeinen unmittelbar – oder doch wenigstens grob – zugänglich ist. Denn die Deskription zergliedert die in der natürlichen Einstellung unmittelbar begreiflichen Sinnelemente in abstrakte Einheiten, die in der Rezeption wieder zu einem sinnhaften Phänomen rückinterpretiert werden müssen. Die durch das Transkript beförderte Befremdung zersplittert soziale Situationen bis zur Unkenntlichkeit. Sie werden nicht mehr verlangsamt, sondern desintegriert. Die Forschungspraxis aber, argumentiert Ronald Hitzler (1993: 230), finde „im Rahmen alltäglichen Verstehens“ statt – erst das wissenschaftliche Verstehen erfolge „im Rückzug in die theoretische Einstellung“. Um Phänomene isolieren und beschreiben zu können, muss man ihre Bedeutung im Gesamtzusammenhang zumindest als kompetentes Mitglied begriffen haben. Die Videotranskription verwehrt jedoch einen solchen kontrollierten, aber im Kern noch alltagsweltlichen Zugang zu den Zusammenhängen.

6.3 Analyse visueller Daten

Auch in der Analyse zeigt sich ein nicht zu unterschätzender Unterschied zwischen dem Umgang mit Audio- und Videotranskripten. Transkripte von Audiomaterial vollziehen einen Medien- und Kanalwechsel: das Ausgangsmaterial zielt auf den Hörsinn, wohingegen das Transkript auf den Sehsinn ausgerichtet ist. Es ist also Analysierenden ohne Sinneseinschränkungen ohne weiteres möglich (und vermutlich bei sehr vielen Analytiker_innen das übliche Vorgehen), gleichzeitig mit den Originaldaten wie dem Transkript zu arbeiten. Beide ergänzen einander und lassen so relevante Phänomene klarer hervortreten. Die Flächigkeit des Transkripts ermöglicht die Annotation und Hervorhebung von Phänomenen, die sich in der Flüchtigkeit der Aufnahme verlieren. Die Präzision und Zeitlichkeit der Aufnahme ermöglicht wiederum ein Korrektiv gegenüber der Gefahr der Über- oder Fehlinterpretation einzelner Segmente des Transkripts, deren Komplexität letztlich niemals adäquat abgebildet werden kann. Das Datenmaterial soll die Aufnahme bleiben – das Transkript ist ein analytisches Hilfsmittel, das dieses Datenmaterial aufzubrechen und zu abstrahieren erlaubt.

Videotranskripte hingegen sind auf denselben Wahrnehmungskanal ausgerichtet wie die Originaldaten. Die simultane Arbeit am Transkript und am Videomaterial ist demnach nicht möglich, denn der Aufmerksamkeitsfokus kann nur auf dem einen oder dem anderen liegen. Das Transkript unterstützt damit die Auseinandersetzung mit dem Videomaterial nicht, sondern bietet einen zweiten, weitgehend unabhängigen Analysefokus. Die Ergebnisse der Auseinandersetzung mit dem Transkript müssen so ans Videomaterial angepasst und an diesem validiert werden. Versuche, das Transkript ‚näher‘ an das Videomaterial zu bringen, etwa durch die Untertitelung der Videos, reduzieren zwar den Weg, den der Blick zurücklegen muss, heben aber das Problem der zwei inkompatiblen Aufmerksamkeitsfoki nicht auf. Darüber hinaus sind Untertitelungen, da sie an die zeitliche Entwicklung des Ausgangsmaterials gekoppelt sind, oftmals deutlich weniger komplex als Transkripte, und dienen häufig vornehmlich als Verstehenshilfe bei schlechter Audioausgabe.

Transkripte erweisen sich für die Analyse von Videoaufnahmen demnach oft als wenig hilfreich, gegebenenfalls sogar hinderlich. Die Rolle von Transkripten, die Verlangsamung, Befremdung, objektivierte Zugänglichkeit für andere, wird durch andere Strategien eingenommen: etwa reduzierte Abspielgeschwindigkeit bis hin zur Frame-by-frame-Analyse oder Screenshots. Während Ayaß vom epistemischen Charakter insbesondere der Videotranskripte spricht, scheinen diese als Erkenntnisobjekte zunehmend

⁹ Ayaß spricht deshalb, Goodwin 1994 folgend, auch vom Transkript als „object of knowledge“, in welches analytische Entscheidungen einfließen, und das deshalb auch niemals objektiv sein kann, sondern einen produktiven, analysebefördernden Zwischenstatus zwischen Fakt und Fiktion einnimmt (Ayaß 2015: 524).

in den Hintergrund zu treten. Gearbeitet wird vornehmlich am Ausgangsmaterial, unterstützt durch die punktgenaue Partiturdarstellung einzelner Sequenzen in Arbeitstranskripten, wie sie etwa vom Lehn (2018a) in seinem Lehrbuch in einer Art Blick auf die analytische Hinterbühne vorstellt. Diese Arbeitstranskripte werden aufgrund ihrer hohen Komplexität dort erstellt, wo ein Phänomen bereits identifiziert wurde und nun in seinen Eigenheiten beschreibbar gemacht werden soll. Sie nehmen damit im Erkenntnisprozess einen fundamental anderen Status ein als Audiotranskripte.

Andererseits erhöht sich die Bedeutung von Transkripten in der Darstellung. Für Harvey Sacks war das Transkript stets ein Merkmal von Wissenschaftlichkeit, indem es als objektiviertes Datenmaterial in der Darstellung der Ergebnisse nicht nur den Analysierenden, sondern auch der Leserschaft zur Verfügung stand (Sacks 1985). Obwohl dieser Objektivitätsanspruch schon lange angekratzt ist, beziehen sich konversationsanalytische Arbeiten doch immer noch sehr stark auf den Nachweischarakter des Transkripts, anhand dessen die eigenen Analysen plausibel und nachvollziehbar gemacht werden können. Die Selektivität des Videotranskripts erhöht dessen epistemischen Charakter immens: Es werden jene Phänomene hervorgehoben und der unmittelbaren Wahrnehmbarkeit zugänglich gemacht, die für den Nachvollzug des eigenen Arguments relevant sind. In der Arbeit mit Videodaten fallen demnach Arbeits- und Darstellungstranskript nicht ineinander. Das Arbeitstranskript ermöglicht die Präzisierung einer am Videomaterial gemachten Entdeckung; das Darstellungstranskript fokussiert auf das Phänomen und lässt die Komplexität der Situation demgegenüber in den Hintergrund treten.

Beides ist für die Analyse von Videodaten sicher nicht zu umgehen; vielmehr eröffnet dieser fokussierende Blick auf das Material häufig erst die Möglichkeit, Phänomene zu identifizieren und in ihren Abläufen zu begreifen, die sich ohne diese Form der Bearbeitung weder der Analyse noch dem rezipierenden Nachvollzug verfügbar machen würden. Die Stärke von Analysen auf der Grundlage von Videodaten liegt ja oftmals in der Lenkung des Blicks auf systematisch geordnete Phänomene, die aus der Vielschichtigkeit des Materials herausgelöst und in ihrer Bedeutung für die analysierte Situation erkennbar gemacht werden.

Der im Vergleich mit Audioanalysen frühzeitige Zugschnitt des Ausgangsmaterials auf Fokusphänomene kann jedoch dazu verleiten, nicht nur in der Darstellung eine Klarheit heraus zu präparieren, die möglicherweise im Ausgangsmaterial nicht auf gleiche Weise zu finden ist; der Umgang mit hochkomplexem Videomaterial kann

auch in der Analyse selbst dazu führen, die Komplexität frühzeitig dadurch zu reduzieren, dass aus einer noch stark in der alltäglichen Betrachtung verhafteten Auseinandersetzung mit dem Material eine frühe Festlegung auf Fokusphänomene erfolgt, die im Weiteren nicht mehr durch zirkuläre Rückgriffe auf das Ausgangsmaterial erhärtet wird. Das wissenschaftliche Verstehen geschieht dann auf einer Datenbasis, die bereits ohne Strategien der Befremdung weitgehend aus einer alltagsweltlichen Perspektive heraus reduziert wurde.

7 Schluss: Die Passung von Datum und Methode

Die Konversationsanalyse verdankt ihre Entstehung theoretischen Überlegungen, die Anfang der 60er Jahre in der Ethnomethodologie entwickelt worden waren, und der Verknüpfung dieser Überlegungen mit einem für damalige Verhältnisse in der Soziologie unüblichen Datentyp. Durch die Hinwendung zu und Auseinandersetzung mit Videodaten hat sie in ihrer weiteren Entwicklung einen starken Erkenntnis Schub erfahren. Sie musste sich mit dem vorher weitgehend unreflektiert angenommenen Primat der Sprache auseinandersetzen und konnte ihre zentrale Entdeckung der intrinsischen Geordnetheit sozialer Handlungen auf eine große Bandbreite weiterer Kontexte ausweiten. Dabei näherten sich punktuell auch Konversationsanalyse und Ethnomethodologie wieder an – die Ethnomethodologie griff auf Instrumentarien der registrierenden Konservierung der Konversationsanalyse zurück, um die Arbeit von Mitgliedern nachzuvollziehen, die Konversationsanalyse löste sich etwas von ihrer Sprach- und Strukturverliebtheit und weitete ihren Blick auf die Vielfalt lokal hergestellter sozialer Phänomene aus. Gleichzeitig ist aber eine methodologische Diskussion der Konsequenzen einer veränderten Datengrundlage weitestgehend ausgeblieben. Dies hat zur Konsequenz, dass die Affordanzen der Daten einen höheren Stellenwert einzunehmen drohen als das ursprünglich zugrunde gelegte Forschungsprogramm.

Videoaufnahmen stellen, trotz ihres Wesens der registrierenden Konservierung (Bergmann 1985), grundsätzlich einen anderen Datentypus dar als Audioaufnahmen. Sie verleiten Konversationsanalytiker_innen auf neue Weise zu Bemühungen, ‚objektive‘, ‚vollständige‘ Wiedergaben der zu untersuchenden Situationen zu erstellen und verschärfen mit einer Fülle von Einzelheiten den Drang, Ordnung auch noch im Kleinsten nachzuweisen. In vielen aktuellen Studien scheint der Analysefokus davon gelei-

tet, was das Material zugänglich macht. Die Zersplitterung von sozialen Situationen in ihre kleinsten Bestandteile macht es kaum noch möglich, die Orientierung auf die Relevantsetzungen der Beteiligten als Ankerpunkt in der Analyse beizubehalten. Analysen werden zunehmend zu kumulierten Detailbeschreibungen einzelner Beobachtungen. Ausgeblendet wird hierbei, dass es der ethnomethodologischen Konversationsanalyse immer primär darum ging, die reziproke und gemeinschaftliche Herstellung von Sinn unter wesentlicher Heranziehung der Mitgliedschaftskompetenz des_r Forschenden nachzuzeichnen. Diese reflektierte Kompetenz macht es erst möglich, den Abfolgen oder Gleichzeitigkeiten von Handlungen ihr sinnhaftes Aufeinanderbezogensein, ihre Sozialität zu entlocken. Dies aber ist das Ziel jeder konversationsanalytischen Studie: die Rekonstruktion sozialer Phänomene in ihrem (widerspenstigen) Eigensinn (Wolff 2005). Wo am Ende keine Antwort auf die Frage steht, welche lokalen Probleme Mitglieder auf systematische Weise in situ bearbeitet und gelöst haben, erschöpft sich die Analyse in reiner detailverliebter Deskription.

Dass sich diese Frage mit Blick auf Videodaten grundsätzlich beantworten lässt, dass es viele exzellente Studien gibt, die dies getan haben, steht für uns außer Frage. Die Ansicht, dass es zur Beantwortung dieser Frage unabdingbar nötig sei, auf Videodaten zurückzugreifen, wechselt allerdings das Phänomen mit seiner Dokumentation. Die klassischen Felder der KA, die sich mit „talk in interaction“ (Schegloff 1987; Heritage & Greatbatch 1991) in Institutionen beschäftigen, dürften von zusätzlichen Videodaten weniger profitieren, als die in jüngerer Zeit erschlossenen Felder, in denen soziales Handeln vornehmlich verkörpert oder mithilfe der Manipulation von Dingen geschieht. Die reziproke, beobachtbar-mittelbare (*accountable*) Herstellung von Sinn geschieht auf vielen Ebenen; deren Nachvollzug muss notwendig lückenhaft sein. Wo er die „Erscheinung“, wie Bergmann sie Simmel folgend nennt (Bergmann 1985) für andere erkennbar und nachvollziehbar, am Material gar selbst reproduzierbar werden lässt, ist er, für alle praktischen Zwecke der soziologischen Erkenntnis, gut genug.

Der Einzug von Videodaten in die Routinen von Konversationsanalytiker_innen hat – *seen-but-unnoticed* – sowohl den Umgang mit Datenmaterial als auch den Bezug zu den methodologischen Grundlagen verändert. Ersteres ist genuin ethnomethodologisch, denn, wie Bergmann schreibt:

Ethnomethodologie und Konversationsanalyse widersetzen sich prinzipiell einer Kanonisierung allgemeiner methodischer Regeln. Für sie sind die Entdeckungsschritte und analytischen

Verfahren ein untrennbarer Bestandteil des Phänomens, auf dessen Identifizierung und Erkundung sie gerichtet sind. Entsprechend dieser Maxime ist nicht nur das Phänomen selbst, sondern auch die Methode seiner Entdeckung und Analyse ein Ziel der ethnomethodologischen Untersuchung. (Bergmann 1988, SE2: 5)

Zweiteres jedoch ist problematisch, denn auf einen Methodenkanon kann dann verzichtet werden, wenn man sich auf eine bestimmte „analytische Mentalität“ (Bergmann 1988 unter Bezug auf Schenkein 1978) einigt, welche die Entdeckung von Phänomenen wahrscheinlich macht. Wo diese überschattet wird von technischen Fragen des Handwerks, droht die ethnomethodologische Aussagekraft zunehmend in den Hintergrund zu rücken.

Die Eigenheiten von Videodaten – ihre Komplexität, ihre Alltagsnähe, ihre sensorische Dichte – dürften nicht nur die Konversationsanalyse, sondern auch andere bislang weitgehend textbasierte Methoden vor Herausforderungen stellen. Die analytische Auseinandersetzung mit bewegten Bildern, ob dies nun über den Umweg der Transkription, der Beschreibung oder der Kodierung geleistet wird, erfordert andere Vorgehensweisen als jene mit (verschriftlichten) Tonaufnahmen oder Textdokumenten, und muss schließlich doch immer zu einer Überführung ins Textliche führen (s. auch Reichertz 2013). Diese Herausforderungen haben natürlich auch eine dezidiert technische Seite, deren Bewältigung notwendig und sinnvoll ist.

Das Potential, das Videodaten als Irritants für bewährte und möglicherweise inzwischen etwas behäbige Methoden darstellen können, wird jedoch verschenkt, wo die Auseinandersetzung sich in technischen Fragen erschöpft. Dabei sind unterschiedliche Entwicklungspfade für Forschungsmethoden denkbar, einen davon haben wir hier dargestellt: Eine unbedachte Verwendung neuer Datenformen kann dazu führen, dass Methoden gewissermaßen von innen ausgehöhlt und selbstvergessen weiterbetrieben werden. Gleichzeitig lässt sich am Beispiel der Konversationsanalyse sehen, dass ein falsch verstandener Innovationswille zur Setzung neuer Standards in den Methoden führen kann, etwa, wenn für eine wissenschaftliche Veröffentlichung Videodaten zur Bedingung gemacht werden. Die Erneuerung der Methoden zielt so aber auf handwerkliche Details und ruft diese letztlich unbegründet zu Orientierungsstandards aus. Ein weiterer Entwicklungspfad, den wir hier aber nicht näher beschreiben haben, lässt sich ebenfalls beobachten: Videodaten werden als dekorative Elemente in Studien eingesetzt, in denen darüber hinaus methodisch jedoch unverändert so verfahren wird wie zuvor. Schließlich lassen sich Tendenzen beobachten, den analytischen Einsatz von Videodaten selbst bereits als eigenständige Methode zu begreifen, den

Methodenzwang also weder von der Disziplin noch vom Gegenstand (vgl. Amann & Hirschauer 1997), sondern vornehmlich von der Datenform ausgehen zu lassen. Die Leichtigkeit, mit der im qualitativen Methodendiskurs neue Schulen und Methoden ausgerufen werden (Reichertz 2007), wirkt hier sicherlich unterstützend. Der reflexiven Auseinandersetzung mit den methodologischen und erkenntnistheoretischen Prinzipien von Methoden wird aber in all diesen Fällen aus dem Weg gegangen. Das Innovationspotential der neuen Datenform erschöpft sich vorzeitig in durch den technischen Fortschritt stetig neu aktualisierten Fragen der Datenorganisation, -aufbereitung und -darstellung, statt dort anzusetzen, wo Reflexion und Selbstvergewisserung immer sinnvoll sind: nämlich an ihren Grundlagen.

Für die Konversationsanalyse lässt sich sagen, dass sie in ihren Anfängen als einer der ersten sozialwissenschaftlichen Forschungsansätze andere Soziolog_innen ermächtigt hat, an der Arbeit von Kolleg_innen teilzuhaben und sie auf Augenhöhe zu diskutieren, nicht, weil sie besonders gute Soziolog_innen und Forscher_innen sind, sondern weil sie „Mitglieder“ sind und deshalb das empirische Datum (ein Audiotranskript, ein Tonbandschnipsel) und das darin sichtbar werdende Problem in der Interaktion hören/sehen können. Wir meinen, dass auch Videodaten dieses Potential haben. Sie „zeigen“ noch viel mehr. Es wird aber zentral darum gehen, auch in bewegten Bildern zu sehen und zu hören, welche Probleme Mitglieder aufgrund der Tatsache haben, dass sie sich begegnen und wie sie sie lösen. Das bedeutet eine starke Rückbesinnung auf die eigentliche Frage der Konversationsanalyse und würde weit über die vorherrschende gegenwärtige videobasierte Forschung hinausgehen, die lediglich das Problem der Koordination erkennen kann.

Literatur

- Amann, K. & S. Hirschauer, 1997: Die Befremdung der eigenen Kultur. Ein Programm. S. 7–52 in: S. Hirschauer & K. Amann (Hrsg.), *Die Befremdung der eigenen Kultur: zur ethnographischen Herausforderung soziologischer Empirie*. Frankfurt am Main: Suhrkamp.
- Arminen, I., C. Licoppe & A. Spagnoli, 2016: Respecifying Mediated Interaction. *Research on Language and Social Interaction* 49: 290–309.
- Ayaß, R., 2015: Doing Data: The Status of Transcripts in Conversation Analysis. *Discourse Studies* 17: 505–528.
- Bergmann, J., 1981: Ethnomethodologische Konversationsanalyse. S. 9–51 in: P. Schröder & H. Steger (Hrsg.), *Dialogforschung*. (Jahrbuch des Instituts für Deutsche Sprache, 1980). Düsseldorf: Schwann.
- Bergmann, J., 1985: Flüchtigkeit und methodische Fixierung sozialer Wirklichkeit: Aufzeichnungen als Daten der interpretativen Soziologie. S. 299–320 in: W. Bonß & H. Hartmann (Hrsg.), *Entzauberte Wissenschaft: Zur Relativität und Geltung soziologischer Forschung*. (Sonderband 3 der Zeitschrift „Soziale Welt“) Göttingen: Schwartz.
- Bergmann, J., 1988: *Ethnomethodologie und Konversationsanalyse* (Studienbrief mit 3 Kurseinheiten). Hagen: FernUniversität Hagen.
- Brand, S., 1976: For God's Sake, Margaret: Conversation with Gregory Bateson and Margaret Mead. *CoEvolutionary Quarterly* 10: 32–44.
- Burgoon, J., 1980: Nonverbal Communication Research in the 1970s: An Overview. *Annals of the International Communication Association* 4: 179–197.
- Chao, Y. R., 1968: *Language and Symbolic Systems*. Cambridge: Cambridge University Press.
- Corsten, M., 2018: Videoanalyse—Quo vadis? S. 799–817 in: C. Moritz & M. Corsten (Hrsg.), *Handbuch Qualitative Videoanalyse*. Wiesbaden: Springer VS.
- Dietrich, M. & G. Mey, 2018: Grounding Visuals. S. 135–152 in: C. Moritz & M. Corsten (Hrsg.), *Handbuch Qualitative Videoanalyse*. Wiesbaden: Springer VS.
- Garfinkel, H., 1967: *Studies in Ethnomethodology*. Englewood Cliffs: Prentice Hall.
- Goffman, E., 1981: *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Goodwin, C., 1994: Professional Vision. *American Anthropologist* 96: 606–633.
- Goodwin, C., 2018: *Co-Operative Action*. New York: Cambridge University Press.
- Greco, L., R. Galatolo, A. A. Horlacher, V. Piccoli, A. C. Ticca & B. Ursi, 2019: Some Theoretical and Methodological Challenges of Transcribing Touch in Talk-In-Interaction. *Social Interaction. Video-Based Studies of Human Sociality* 2: <https://doi.org/10.7146/si.v2i1.113957>
- Heath, C. & J. Hindmarsh, 2002: Analysing Interaction. S. 99–121 in: T. May (Hrsg.), *Qualitative Research in Action*. London: Sage.
- Heath, C. & P. Luff, 1992: Collaboration and Control. *Crisis Management and Multimedia Technology in London Underground Line Control Rooms*. *Computer Supported Cooperative Work (CSCW)* 1: 69–94.
- Heritage, J. & D. Greatbatch, 1991: On the Institutional Character of Institutional Talk: The Case of News Interviews. S. 93–137 in: D. Boden & D. H. Zimmerman (Hrsg.), *Talk and Social Structure: Studies in Ethnomethodology and Conversation Analysis*. Cambridge: Polity Press.
- Herrle, M., 2007: *Selektive Kontextvariation. Die Rekonstruktion von Interaktionen in Kursen der Erwachsenenbildung auf der Basis audiovisueller Daten*. Frankfurt am Main: Fachbereich Erziehungswissenschaft der Goethe-Universität.
- Hirschauer, S., 2001: Ethnografisches Schreiben und die Schweigsamkeit des Sozialen. Zu einer Methodologie der Beschreibung. *Zeitschrift für Soziologie* 30: 429–451.
- Hitzler, R., 1993: Verstehen: Alltagspraxis und wissenschaftliches Programm. S. 223–240 in: T. Jung & S. Müller-Doohm (Hrsg.), *„Wirklichkeit“ im Deutungsprozess. Verstehen und Methoden in den Kultur- und Sozialwissenschaften*. Frankfurt am Main: Suhrkamp.

- Hutchby, I. & R. Wooffitt, 2008: *Conversation Analysis*. Cambridge: Polity.
- Kautt, Y., 2017: Grounded Theory als Methodologie und Methode der Analyse visueller Kommunikation. *Forum Qualitative Sozialforschung/Forum Qualitative Social Research* 18, DOI: 10.17169/FQS-18.3.2859.
- Knoblauch, H., B. Schnettler, J. Raab & H.-G. Soeffner, 2006 (Hrsg.): *Video Analysis. Methodology and Methods: Qualitative Audiovisual Data Analysis in Sociology*. Frankfurt am Main u.a.: Lang.
- Konecki, K., 2011: Visual Grounded Theory: A Methodological Outline and Examples from Empirical Work. *Revija Za Sociologiju* 41: 131–160.
- Luff, P. & C. Heath, 2012: Some 'Technical Challenges' of Video Analysis: Social Actions, Objects, Material Realities and the Problems of Perspective. *Qualitative Research* 12: 255–279.
- Lynch, M. & D. Bogen, 1994: Harvey Sacks's Primitive Natural Science. *Theory, Culture & Society* 11: 65–104.
- Lynch, M., 2000: The Ethnomethodological Foundations of Conversation Analysis. *Text & Talk* 20: 517–532.
- Lynch, M., 2019: Garfinkel, Sacks and Formal Structures: Collaborative Origins, Divergences and the History of Ethnomethodology and Conversation Analysis. *Human Studies*: 183–198.
- Mair, M., P.G. Watson, C. Elsey & P.V. Smith, 2012: War-Making and Sense-Making: Some Technical Reflections on an Instance of 'Friendly Fire'. *British Journal of Sociology* 63: 75–96.
- Marstrand, A.K. & J. Svennevig, 2018: A Preference for Non-Invasive Touch in Caregiving Contexts. *Social Interaction: Video-Based Studies of Human Sociality* 1: DOI: 10.7146/si.v1i2.110019
- McIlvenny, P., 2019: Inhabiting Spatial Video and Audio Data: Towards a Scenographic Turn in the Analysis of Social Interaction. *Social Interaction. Video-Based Studies of Human Sociality* 2: DOI: 10.7146/si.v2i1.110409.
- Meier zu Verl, C., 2018: *Daten-Karrieren und epistemische Materialität. Eine wissenschaftssoziologische Studie zur methodologischen Praxis der Ethnografie*. Stuttgart: Metzler.
- Meyer, C., 2018: *Culture, Practice, and the Body. Conversational Organization and Embodied Culture in North-Western Senegal*. Stuttgart: Metzler.
- Meyer, C. & U. von Wedelstaedt, 2013: Skopische Sozialität: Sichtbarkeitsregime und visuelle Praktiken im Boxen. *Soziale Welt* 64: 69–95.
- Mohn, E., 2013: Differenzen zeigender Ethnographie: Blickschneisen und Schnittstellen der Kamera-Ethnographie. *Soziale Welt* 64: 171–189.
- Mondada, L., 2006: Video Recording as the Reflexive Preservation and Configuration of Phenomenal Features for Analysis. S. 51–68 in: H. Knoblauch, B. Schnettler, J. Raab & H.-G. Soeffner (Hrsg.), *Video Analysis. Methodology and Methods: Qualitative Audiovisual Data Analysis in Sociology*. Frankfurt am Main u.a.: Lang.
- Mondada, L., 2016: Challenges of Multimodality: Language and the Body in Social Interaction. *Journal of Sociolinguistics* 20: 336–366.
- Mondada, L. & R. Schmitt, 2010: Zur Multimodalität von Situationseröffnungen. S. 7–52 in: L. Mondada & R. Schmitt (Hrsg.), *Situationseröffnungen. Zur multimodalen Herstellung fokussierter Interaktion*. Tübingen: Narr.
- Moritz, C. & M. Corsten (Hrsg.), 2018: *Handbuch Qualitative Videoanalyse*. Wiesbaden: Springer VS
- Mortensen, K. & B.L. Due, 2019: Editorial. Special Issue: Representing Data in Video-Based Studies. *Social Interaction. Video-Based Studies of Human Sociality* 2: <https://doi.org/10.7146/si.v2i1.114682>
- Neville, M., 2015: The Embodied Turn in Research on Language and Social Interaction. *Research on Language and Social Interaction* 48: 121–151.
- Oevermann, U., T. Allert, H. Gripp, E. Konau, J. Krambeck, E. Schröder-Caesar & Y. Schütze, 1976: Beobachtungen zur Struktur der sozialisatorischen Interaktion. S. 274–295 in: M.R. Lepsius (Hrsg.), *Zwischenbilanz der Soziologie: Verhandlungen des 17. Deutschen Soziologentages vom 31. Okt. bis 2. Nov. 1974 in Kassel*. Stuttgart: Enke.
- Philips, S.U., 1976: Some Sources of Cultural Variability in the Regulation of Talk. *Language in Society* 5: 81–95.
- Potter, J., 2002: Two Kinds of Natural. *Discourse Studies* 4: 539–542.
- Rawls, A.W., 2008: Harold Garfinkel, Ethnomethodology and Workplace Studies. *Organization Studies* 29: 701–732.
- Reichert, J., 2007: Qualitative Sozialforschung – Ansprüche, Prämissen, Probleme. *Erwägen – Wissen – Ethik* 18: 195–2008.
- Reichert, J., 2013: Der bildliche Ausdruck als vertextbare Kommunikation? *Soziale Welt* 64(1–2): 37–50.
- Sacks, H., 1985: Notes on Methodology. S. 21–27 in: J. M. Atkinson & J. Heritage (Hrsg.): *Structures of Social Action. Studies in Conversation Analysis*. Cambridge: Cambridge University Press.
- Sacks, H., 1992: *Lectures in Conversation*. Vol. I: Fall 1964 – Spring 1968. Hg. v. Gail Jefferson. Oxford: Blackwell.
- Sacks, H. & E.A. Schegloff, 2002: Home Position. *Gesture* 2: 133–146.
- Sacks, H., E.A. Schegloff & G. Jefferson, 1974: A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50: 696–735.
- Schegloff, E. & H. Sacks, 1973: Opening up Closings. *Semiotica* 8: 289–327.
- Schegloff, E.A., 1987: Analyzing Single Episodes of Interaction: An Exercise in Conversation Analysis. *Social Psychology Quarterly* 50: 101 – 114.
- Schegloff, E.A., 1992: On Talk and its Institutional Occasions. S. 101–134 in: P. Drew & J. Heritage (Hrsg.), *Talk at Work. Interaction in Institutional Settings*. Cambridge: Cambridge University Press.
- Schenkein, J., 1978: Sketch of an Analytic Mentality for the Study of Conversational Interaction. S. 1–6 in: J. Schenkein (Hrsg.): *Studies in the Organization of Conversational Interaction*. New York: Academic Press.
- Schindler, L., 2012: Visuelle Kommunikation und die Ethnomethoden der Ethnographie. *Österreichische Zeitschrift für Soziologie* 37: 165–183.
- Schmitt, R., 2005: Zur multimodalen Struktur von turn-taking. *Gesprächsforschung Online* 6: 17–61.
- Schmitt, R., 2006: Videoaufzeichnungen als Grundlage für Interaktionsanalysen. *Deutsche Sprache* 1: 18–31.
- Schnettler, B. & A. Baer (Hrsg.), 2013: *Themenheft Visuelle Soziologie. Soziale Welt* 64 (1/2).
- Schütze, F., 1976: Zur Hervorlockung und Analyse von Erzählungen thematisch relevanter Geschichten im Rahmen soziologischer Feldforschung: dargestellt an einem Projekt zur Erforschung

von kommunalen Machtstrukturen. S. 159–260 in: Arbeitsgruppe Bielefelder Soziologen (Hrsg.), *Kommunikative Sozialforschung: Alltagswissen und Alltagshandeln*, Gemeindeforschung, Polizei, politische Erwachsenenbildung. München: Fink.

- Selting, M., P. Auer, B. Barden, J. Bergmann, E. Couper-Kuhlen & S. Günthner et al. (1998): Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte* 173: 91–122.
- Selting, M., P. Auer, D. Barth-Weingarten, J. Bergmann, P. Bergmann & K. Birkner et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung Online* 10: 353–402.
- Simmel, G., 1916: *Das Problem der historischen Zeit*. Berlin: Verlag von Reuther & Reichard.
- Smith, R., 2020: Seeing the Trouble: A Mountain Rescue Training Scenario in its Circumstantial and Situated Detail in Three Frames. *Ethnographic Studies* 17: 41–59.
- Tuma, R., B. Schnettler & H. Knoblauch, 2013: *Videographie: Einführung in die interpretative Videoanalyse sozialer Situationen*. Wiesbaden: Springer VS.
- vom Lehn, D., 2018a: *Ethnomethodologische Interaktionsanalyse. Videodaten analysieren und die Organisation von Handlungen darstellen*. Weinheim: Beltz Juventa.
- vom Lehn, D., 2018b: *Ethnomethodologische Interaktionsanalyse*. S. 183–196 in: C. Moritz & M. Corsten (Hrsg.), *Handbuch Qualitative Videoanalyse*. Wiesbaden: Springer VS.
- Wolff, S., 2005: Standards in der sozialpädagogischen Forschung. S. 115–134 in: C. Schweppe & W. Thole (Hrsg.), *Sozialpädagogik als forschende Disziplin: Theorie, Methode, Empirie*. Weinheim: Beltz Juventa.

Autorinnen

Sarah Hitzler

Universität Bielefeld
Fakultät für Soziologie
Postfach 10 01 31
33501 Bielefeld
sarah.hitzler@uni-bielefeld.de

Dr.in Sarah Hitzler, geboren 1979, wissenschaftliche Mitarbeiterin an der AG Qualitative Methoden der Fakultät für Soziologie an der Universität Bielefeld. Studium der Linguistik und Soziologie in Bielefeld und Amsterdam. Promotion an der Universität Bielefeld. Arbeitsschwerpunkte: Gruppeninteraktionen, institutionelle Interaktion, Herstellung von sozialer (Un-)Gewöhnlichkeit, Ethnomethodologie, Konversationsanalyse, qualitative Sozialforschung, Selbsthilfe. Veröffentlichungen: *Aushandlung ohne Dissens? Praktische Dilemmata der Gesprächsführung im Hilfeplangespräch*, Wiesbaden 2012; *Interaktion zwischen Personen mit und ohne kognitive Beeinträchtigung. Eine konversationsanalytische Untersuchung zur Rolle der Herstellung von ‚Gewöhnlichkeit‘*, in Dobslaw (Hrsg.), *Partizipation – Teilhabe – Mitgestaltung: Interdisziplinäre Zugänge*, Opladen 2018; *„DA“ wurde ich auch ganz blöde angeguckt‘. Das moralische Spannungsfeld zwischen gesellschaftlichem Anspruch und individueller Integrität in einer Selbsthilfegruppe zum Thema Übergewicht*, *Journal für Psychologie* 28(2), 2020.

Daniela Böhringer

Universität Duisburg-Essen
Institut Arbeit und Qualifikation
Forsthausweg 2
LE 528
47057 Duisburg
daniela.boehringer@uni-due.de

Dr.in Daniela Böhringer, geboren 1968, seit 2019 wissenschaftliche Mitarbeiterin am Institut Arbeit und Qualifikation der Universität Duisburg-Essen. Arbeitsschwerpunkte: Institutionelle Interaktion und Kommunikation, Konversationsanalyse, Ethnomethodologie, qualitative Sozialforschung, Wohlfahrtsstaatsforschung, Digitalisierung, Technik, Hochschulforschung. Studium der Soziologie, Ethnologie und Germanistik an den Universitäten Basel, Aachen und Graz. Promotion an der Karl-Franzens-Universität Graz, Österreich. 2008 – 2019 Wissenschaftliche Mitarbeiterin an den Universitäten Hildesheim und Osnabrück.

Publikationen: Böhringer, D. & Wolff, S. (2010): Der Computer als „Partner“ im institutionellen Gespräch. *Zeitschrift für Soziologie* 39: 233–251; Böhringer, D.; Karl, U.; Müller, H.; Schröer, W.; Wolff, S. (2012): Den Fall bearbeitbar halten. *Gespräche in Jobcentern mit jungen Menschen*. Opladen, Berlin, Toronto: Barbara Budrich; Böhringer, D. (2017): „Gut gemacht Mädchen!“ Geschlechterdifferenz und Geschlechterungleichheit an Hochschulen. S. 495–512, M. S. Baader, T. Freytag (Hrsg.) *Bildung und Ungleichheit in Deutschland*. Wiesbaden: Springer; Böhringer, D. (2020): Qualitative Forschungszugänge zu digitalisierter Sozialer Arbeit. S. 612–624 in: N. Kutscher, T. Ley, U. Seelmeyer et al. (Hrsg.), *Handbuch Soziale Arbeit und Digitalisierung*. Weinheim/Basel: Beltz Juventa.