

# Der bildliche Ausdruck als vertextbare Kommunikation?<sup>1</sup>

Von Jo Reichertz

**Zusammenfassung:** In dem Artikel wird die Frage erörtert, ob man die in (Bewegt-)Bildern jeder Art fixierten Inhalte und Bedeutungen im Rahmen *sozialwissenschaftlicher* Forschung verlustfrei in zeichenbasierte Inhalte und Bedeutungen transformieren und dann fixieren kann. Dies kann man – so die hier vertretene Position der hermeneutischen Wissenssoziologie – wenn man die zu interpretierenden Bilder als soziale *Handlungen* begreift und nicht als *ästhetische Strukturen und Muster*. Dabei muss zwischen der *gezeigten Handlung* (also der im Bild gezeigten Handlung) und der *Handlung des Zeigens* (also der, mit dem Bild zu zeigen) unterschieden werden. Beide Typen sozialen Handelns sind bereits sozial typisiert und somit auch in dieser Typik beschreib- und notierbar.

## 1. Was ist die Frage?

„Bilder kann man nicht wirklich in Texte überführen!“ Das sagen zurzeit einige Sozialwissenschaftler/innen (z.B. Engell 2010; Loer 2010; Knoblauch 2011), die sich mit der Analyse stehender wie laufender Bilder beschäftigen. Und wenn sie es sagen, dann tun sie es meist mit dem Verweis auf die Eigenständigkeit des Visuellen. Visuelles ließe sich prinzipiell nicht vertexten<sup>2</sup> – täte man es dennoch, ginge das Besondere, oder auch: das Wesentliche des Visuellen verloren. Der so gewonnene Text könne das Visuelle *nicht vollständig*, sondern *nur teilweise* erfassen – mithin seien Vertextungen von Bildern immer und notwendig unzureichend und für die Erfassung und Analyse von Bildern *nicht wirklich* geeignet.

Die etwas diffuse Formulierung ‚nicht wirklich‘ ist ein Versuch, die eigentümliche Praxis der Bildbeschreibung der Vertreter der Eigenständigkeit des Visuellen zu erfassen: Denn obwohl sich die Protagonisten dieser Position immer wieder und oft auch entschieden dafür aussprechen, dass Transkriptionen des Bildlichen (wenn überhaupt) nur partiell möglich sind, arbeiten alle, wenn sie Bilder interpretieren, mit (teils elaborierten) Formen der Datenformation (dichten Bildbeschreibungen, Transkriptionen, Diagrammen, Partituren). Niemand lässt nur das Bild sprechen, sondern vor der Interpretation wird das Bild in andere Symbole (meist Sprache) transformiert.

Mir geht es hier darum zu diskutieren, ob die *sozialwissenschaftlichen* Vertreter einer solchen Position Recht haben – also ob man<sup>3</sup> für diese Position gute Gründe vortragen kann. Denn es macht einen Unterschied, ob Sozialwissenschaftler/innen Bilder *im Rahmen sozialwissenschaftlicher Forschung* Bilder ausdeuten oder ob dies Kunst- oder Medienwissenschaftler/innen tun. Es geht mir also nicht darum, die Debatte um die Eigenständigkeit des Visuellen hier systematisch zu beschreiben, sondern darum, aus soziologischer Sicht eine Gegenposition vorzulegen, die sich wesentlich aus den Prämissen und Arbeitsergebnissen einer *hermeneuti-*

1 Ganz herzlich danken möchte ich Christine Moritz für ihre Anregungen und ihre Kritik. Beide waren sehr hilfreich.

2 In dieser Diskussion ist oft auch unklar, ob es das Ziel ist, Bilder zu versprachlichen, zu verschriftlichen oder zu vertexten. Ich entscheide mich hier für den Textbegriff, weil er der umfassendste ist. Denn Text wird hier verstanden als Verschriftung nicht nur gesprochener, mündlicher Kommunikation, sondern auch von musikalischer, mathematischer oder gestalteter Kommunikation (Musik, Diagramme, Storyboards, Tabellen etc.).

3 Es geht mir also nicht darum, mich mit den teils deutlich voneinander abweichenden Positionen verschiedener Vertreter/innen im Einzelnen auseinanderzusetzen, sondern um die grundsätzliche Position – weshalb nicht alle hier angesprochenen Kritikpunkte auf alle Protagonisten/innen in gleicher Weise zutreffen.

*schen Wissenssoziologie* speist (Reichertz / Englert 2010b) und die sich auf die Nutzung von Bildinterpretationen für die sozialwissenschaftliche Forschung bezieht.

Um das zu leisten, werde ich die in der aktuellen sozialwissenschaftlichen Literatur oft diskutierten Fragen erörtern (z.B. Bohnsack 2009; Breckner 2010; Keppler 2006; Knoblauch 2011; Knoblauch / Schnettler / Raab / Soeffner 2006; Loer 2010; Raab 2008; Reichertz 2012; Reichertz / Englert 2010 a, 2010 b; Schnettler / Knoblauch 2007, 2008; auch Lobinger 2012), ob und wie man Inhalte bestimmter Speichermedien<sup>4</sup> (Bewusst-sein, Bild, Sprechen, Text) verlustfrei (also identisch) in strukturell andere Speichermedien transformieren kann (Goodman 1997) – in diesem Fall, ob man die in (Bewegt-)Bildern jeder Art (Gemälde, Fotos, Grafiken, technische Bilder, Filme, Videos etc.) fixierten und zum Ausdruck gebrachten Inhalte und Bedeutungen verlustfrei in zeichenbasierte Inhalte und Bedeutungen transformieren und dann fixieren kann.

Einige der so formulierten Fragen sind meiner Ansicht nach trivial, andere falsch gestellt. Trivial sind sie, weil keine Abbildung mit dem Original identisch sein kann (wie Bateson das schlagend mit seinem ‚Karte-Territorium-Beispiel‘ gezeigt hat – Bateson 1987: 40 f), schon gar nicht, wenn die Abbildung mit Hilfe eines anderen Mediums erfolgt. Die Materialität des Mediums schreibt sich immer und unlösbar in die Inhalte ein (Flusser 1994). Identische Transformationen sind also nicht möglich – was auch zum einen niemand wirklich bestreitet, zum anderen aber auch nicht gefordert ist. Die entscheidenden Fragen sind vielmehr, ob es a) *hinreichend* genaue Transformationen für den Zweck der Erfassung von Informationen in Bildern geben kann und b) an welchen Stellen des Forschungsprozesses Transformationen visueller Inhalte in der Qualitativen Sozialforschung notwendig sind und an welchen nicht. Das sind die eigentlich interessanten und wichtigen Fragen, mit denen ich mich im Weiteren auseinandersetzen werde.<sup>5</sup> Die Frage nach der Korrektheit und Vollständigkeit der Transkription löst sich dann auf in die Frage nach der *Angemessenheit* von Transkriptionen. Angemessenheit bezieht sich dabei immer auf die Fragestellung der Forschung. Die hier vertretene These ist, dass sich bei *sozialwissenschaftlichen* Bildinterpretationen alles (im Hinblick auf die Fragestellung) Relevante auch transkribieren lässt – oft allerdings nur sukzessiv.

Falsch gestellt sind die o.a. Fragen, wenn damit auch gesagt werden soll, eine Bilddeskription sei nur dann gelungen, wenn der individuelle Eindruck des Bildausdrucks, den die Betrachter/innen (Simultanität, Vorsprachlichkeit) sich geschaffen haben, in der Notation richtig und vollständig wiedererzeugt wird. Noch viel schwieriger bzw. verworren würde die Lage, wenn gefordert würde (was aber bislang niemand ernsthaft tut), die Notation müsse so gestaltet werden, dass die individuelle Ausdrucksintention des ‚Autors‘ zweifelsfrei wiedererzeugt wird. Denn diese ist selbst dem Autor in der Regel nicht mehr zugänglich. Zugänglich sind nur die sozialen Typisierungen der Ausdrucksintention.

Solche Forderungen sind einerseits Ausdruck einer spezifischen, nämlich einer subjektivistisch und intentionalistisch angelegten Wirkungs- bzw. Aneignungstheorie, andererseits sind sie Forschungsfragen verpflichtet, die außerhalb der klassischen soziologischen Forschung liegen, nämlich den Fragen nach Bewusstseinsinhalten von konkreten Akteuren in bestimmten Momenten. Diese sind in der genannten Disziplin mit den verfügbaren Instrumenten weder verfügbar noch interpretierbar.

4 Speichermedien sind all jene Medien, die kommunikatives Handeln und Tun und deren Produkte (mehr oder weniger verlustfrei) fixieren und von der Situation ablösen und damit eine wiederholte, mehr oder weniger identische Reproduktion ermöglichen.

5 Der Beitrag stellt in diesem Sinne eine Fortsetzung der Überlegungen in Reichertz / Englert (2010 b) dar und wird in Reichertz (2013) hinsichtlich einer weiteren Differenzierung der Gütekriterien fortgesetzt.

Bei der Erörterung der oben angesprochenen Fragen kann man viel von der in den 1960er- und 1970er-Jahren geführten Debatte darüber lernen, ob und wann die Texttranskription von sprachlicher Interaktion angemessen ist und wann nicht.<sup>6</sup> Auch damals wurde, was mittlerweile nahezu vollständig in Vergessenheit geriet, hitzig darüber debattiert, ob die reale, sinnliche Interaktion und Kommunikation von Menschen sich mittels eines Verbal-Transkripts der Textspur dieser Interaktion und Kommunikation angemessen erfassen, also auch hier: ob die Welt sich weitgehend verlustfrei in Text transformieren lasse und ob die Welt mehr und etwas ganz anderes sei als Text (Garz 1994).

Virulent ist das alte Problem in den letzten Jahren deshalb geworden, weil die Sozialwissenschaften zusätzlich zum textbasierten Material verstärkt mit Bildaufzeichnung aller Art arbeiten, sei es, dass diese ihre Erhebungsinstrumente sind (videografische Aufzeichnungen, Fotografien der Forschungsfelder, auch Videografien der Interviews), sei es, dass sie ihre wichtigste Datensorte (Gegenstände der jeweiligen Untersuchung) sind. Beides hängt sicherlich ganz wesentlich damit zusammen, dass die technischen Medien der Bildaufzeichnung und der Bildbearbeitung so kostengünstig und für jeden handhabbar geworden sind (Reichert / Englert 2010 a, 2010 b). Veränderte in den 1960er-Jahren die Verbreitung kostengünstiger und transportabler Tonbandgeräte das Geschäft der Sozialwissenschaft, die bis dahin ihre Daten lediglich mit Papier und Bleistift und manchmal sogar mit Stenografie fixieren konnten, massiv, so hat die Verfügbarkeit preisgünstiger Videokameras (aller Art) und Auswertungssoftware die Praxis der Sozialwissenschaft erneut tiefgreifend verändert. Und weil heute so viele Menschen im Alltag stehende wie laufende Bilder (Videocams, Handys) produzieren und damit interagieren und weil auch viele Sozialwissenschaftler/innen deshalb mit stehenden wie laufenden Bildern arbeiten, gibt es die neue Debatte um die Angemessenheit und Zweckmäßigkeit von textuellen Repräsentationen.

## 2. Exkurs: Was ist ein Video und was eine Videoanalyse?

Doch an dieser Stelle taucht ein erstes Problem auf: Der Gebrauch der Worte ‚Video‘ oder ‚Videoanalyse‘ ist in der deutschsprachigen Debatte zur Zeit verwirrend und oft weiß man nicht, was eigentlich gemeint ist – was zu vielfältigen Missverständnissen führt. Da sicherlich noch einige Zeit vergehen, bis sich ein einheitlicher Sprachgebrauch eingestellt hat (siehe hierzu auch Knoblauch 2011), hier ein Versuch, etwas Ordnung und somit Orientierung aufzubauen.

Videos bestehen aus *laufenden Bildern* und unterscheiden sich damit von den *stehenden Bildern* kategorial. Bei den laufenden Bildern unterscheidet man wieder zwischen *Filmen* und *Videos*. Filme folgen einem Drehbuch, Videos in der Regel nicht. Aber auch innerhalb von Filmen und Videos gilt es, Verschiedenes auseinanderzuhalten. Denn auch hier gibt es Filme und Videos von Professionellen, die ihre Kunst gelernt haben und sich mit ihrer Arbeit auf ‚ihren‘ Diskurs beziehen, die Geschichte des Films reflektieren, bewusst mit Farben und Formen spielen, Regeln auflösen, verfremden, sich selbst thematisieren, mit den Ausdrucksregeln jonglieren, neue schaffen. Und natürlich gibt es auch hier die Halbprofessionellen und die Amateure und auch solche, die ohne Sinn und Verstand auf den Aufnahmeknopf drücken und hin- und herzoomen. Aber auch hier macht es weder Sinn, die Bewegung der Bilder in *stills* aufzulösen, noch macht es viel Sinn, diese stehenden Bilder ikonografisch oder ikonologisch zu interpretieren.

6 Beispielhaft dafür die Beiträge in der Zeitschrift *Studium Linguistik* in den Jahren 1975 bis 1978, abrufbar unter: <http://www.degruyter.com/view/j/zfgl.1977.5.issue-3/issue-files/zfgl.1977.5.issue-3.xml>.

Videos mögen manchmal auch (Bildende) Kunst sein (vgl. Bohnsack 2009). Dennoch führt es aus meiner Sicht nicht sehr weit, sie auch mit den Methoden der Deutung von Malkunst deuten zu wollen (z.B. mit Panofski 1978; Imdahl 1980; Gombrich 1978; Bättschmann 1984; allgemein Kaemmerling 1991). Der Versuch, Fotos mit den Mitteln der Malkunst zu deuten, wird den Besonderheiten von Fotos nur begrenzt gerecht. Gemalte Bilder unterscheiden sich nämlich kategorial von Fotos und es gilt auch: Fotointerpretationen unterscheiden sich kategorial von Videointerpretation, nicht lediglich graduell. Von der (Foto)Bild- zur Videointerpretation ist es kein kleiner Schritt, sondern ein sehr großer.<sup>7</sup> Videointerpretationen bestehen nicht aus der Summe von Einzelbildinterpretationen, sondern sie müssen das Eigene der Bewegbilddarstellung (z.B. Sequenzialität) erfassen können.

Unter ‚Video‘ soll hier erst einmal der materielle Träger (Magnetband oder digital codiert auf Datenträger) verstanden werden, in den zeitlich begrenzte Teile der audiovisuellen Teile einer bestimmten Interaktionssequenz mit Hilfe digitaler Kameratechnik eingeschrieben werden. Diese besondere und selektive Einschrift<sup>8</sup> ist mit geeigneten technischen ‚Lese- und Darstellungsgeräten‘ immer wieder identisch sichtbar zu machen, abzuspielen oder besser: *neu aufzuführen*. Die Neuaufführung der gespeicherten Ton- und Bildspuren auf (digitalen) Displays in bestimmten Rezeptionssettings wird auch oft als ‚Video‘ bezeichnet. *Videoanalysen* sind in der Regel Analysen von (meist kleinen Teilen) solcher Neuaufführungen.

Die Videobilder sind (mittlerweile) sofort (also schon während der Produktion) als Bilder und als digitale Datei verfügbar und kontrollierbar (also auch löschar) und in der Postproduktion leicht mittels Grafikprogrammen bearbeit- und gestaltbar. All dies ermöglicht eine schnelle Produktion ohne große Vorkenntnisse und eine billige Nachbearbeitung. Obwohl also auch Videoproduktionen ‚Filme‘ sind, muss man unterscheiden: Kinofilme und Fernsehfilme haben andere Produktionsbedingungen und deshalb auch eigene und eigenständige Bildsprachen (und andere Vorstellungen von den impliziten Zuschauer/innen) entwickelt. Videoaufnahmen, egal ob sie von Professionellen, Halbprofessionellen oder Amateuren erstellt wurden, unterscheiden sich deutlich von den Filmproduktionen, weil letzteren immer ein explizites und vorab entwickeltes Drehbuch zugrunde liegt (z.B. Hilt 2013). Videoproduktionen verfügen dagegen eher selten über ein Drehbuch (Ausnahme: scripted-Formate) und einen Einstellungsplan.

Die Aufnahmen werden in der Regel nicht konkret in ihrem Ablauf geplant, sondern die Ereignisse vor der Kamera übernehmen die Regie oder genauer: *scheinen* die Regie zu übernehmen – weshalb manche auch (im Übrigen zu Unrecht) annehmen, Videoaufzeichnungen

7 Ähnlich und ausführlicher Moritz: „Zumeist bleibt in dieser Vorgehensweise der Linearisierung von Einzelbildern jedoch ein Aspekt unberücksichtigt: Das Bild ist gerade kein Kontinuum, da keiner der piktoral einschlägigen Parameter (Flächen, Farben, Formen, Proportionen, Kontraste) auf dem Bild sich als ein Anfang oder als ein Ende bezeichnen ließe.[...] Begrenzung erfährt ein Bild nicht durch Anfang und Ende, sondern ausschließlich durch seinen Rahmen (Quadrage); aneinandergesetzte „Linien“ aus Einzelbildern bauen daher in dieser Vorgehensweise etwas auf, für das es in der Komplexität des Pikturalen keine Entsprechung gibt, und nahezu unbemerkt unterwandert auf diese Weise in der Forschungsarbeit das Diagrammatische die Spezifik des Filmischen.“ (Moritz 2011: 25).

8 Diese Einschrift ist besonders und selektiv, weil sie allein die Ton- und Bildspur der Ereignisse aufzeichnet. Zwar verzerrt diese Art der Aufzeichnung in der Regel nicht wesentlich die Töne und Bilder, aber ohne Zweifel besteht die reale Interaktion und Kommunikation aus mehr als nur der visuellen und akustischen Wahrnehmung. Insofern kann sehr wohl das Bild- und Tonmaterial den „Sinngehalt der Praxis“ (Loer 2011: 324) verzerren.

würden ein Ereignis vollständig dokumentieren<sup>9</sup> und deshalb seien Videoaufnahmen ein Mittel, den Nichtdabeigewesenen später und anderenorts das Vergangene so zu zeigen, wie es der Beobachter gesehen und erlebt hat – Videos seien also in der Lage, eine identische Erfahrung zu erzeugen.

Bei der Produktion und Analyse sind deshalb weitere Unterscheidungen von Bedeutung – so ist es z.B. von besonderer Bedeutung, wer aus welchen Motivationen heraus ein Video produziert hat oder genauer: als Kameramann oder -frau tätig wird. In der Regel wird zurzeit zwischen vier typischen Produktionsweisen und damit verbunden vier typischen Motivationen unterschieden:

1. Wissenschaftler/innen produzieren Videos zum Zwecke wissenschaftlicher *Forschung* (paradigmatisch hierfür: Heath / Hindmarsh / Luff 2010; Knoblauch 2011 und auch Tuma / Schnettler / Knoblauch 2013; Dinkelaker / Herrle 2009). Die Produkte werden in Interpretationsgruppen aufgeführt und zum Gegenstand der Analyse gemacht.
2. Amateure und Halbprofessionelle produzieren zur Dokumentation von *besonderen Festen* und *Anlässen* Videos (vor allem Raab 2008; auch Reichertz / Ivanyi 2002). Die Produkte werden in privatem Rahmen aufgeführt.
3. Da Videokameras die zeitgleiche Wiedergabe des Bildes auf einem Bildschirm ermöglichen, können Menschen ihr Handeln (z.B. beim Tanzen, Sport oder beim Sex; siehe hierzu Boll 2012) zeitgleich aufzeichnen und durch die Beobachtung des Bildschirms die Aufzeichnung und die Darstellung des Aufgezeichneten gestalten. Hier fallen Kameramann und Darsteller in eins. Diese Produkte können auch ins Netz gestellt oder privat aufgeführt werden
4. Halbprofessionelle oder Professionelle produzieren mit und/oder Skript ein Video, um es kommerziell zu verwerten, also um es an *Fernsehsender* zu *verkaufen* (Reichertz / Englert 2011 a; 2011 b, Bohnsack 2009, Jost et al. 2013). Die Videos werden an ein ‚verstreutes‘ Publikum ausgestrahlt.

Das Bestehen auf der In-Rechnung-Stellung der Produktions- und Herstellungsbedingungen von Videos soll hier nicht heimlich die Autorabsicht wieder ins Spiel bringen. Es geht gerade nicht um die Rettung der Autorintention, also der subjektiven Bewusstseinsinhalte im Prozess der Produktion. Es geht statt dessen darum zu zeigen, dass durch Produktions- und Herstellungsbedingungen eigene Möglichkeitsräume geschaffen werden, die auch inhaltliche Rahmen kreieren. Sie schaffen ein eigenes Genre und dieses Genre dient dann Macher/innen wie Nutzer/innen als Deutungsrahmen.

Diese Rahmen eröffnen und begrenzen, bedingen kommunikatives Handeln, und sie weisen dem Handeln meist auch ein gewisses Gewicht zu (Witz, Prüfung, Gebet, Entschuldigung). Die Rahmen legen nahe, was jeweils kommuniziert werden darf und was nicht und welche Folgen es hat (Reichertz 2009; Soeffner 1989). Diese Rahmen sind gesellschaftlich erarbeitet und verbürgt – sie sind Institutionen. Rahmen können in bestimmten Situationen interaktiv ausgefüllt, moduliert oder auch verändert werden. Fazit: Rahmen bilden das Fundament, auf dem Kommunikation, aber auch die Analyse erst möglich werden.

9 Eine reine Dokumentation von Leben kann nur als Idealtyp entworfen werden, von dem sich tatsächlich realisierte Produktionen systematisch unterscheiden müssen. Selbst die Kamera, die 24 Stunden lang aus einer festen Position das Geschehen auf einer Kreuzung aufzeichnet, ist keine Dokumentation, sondern dem so gerichteten Kamerablick ist immer schon eine Annahme darüber eingeschrieben, die besagt, was sehenswert und was nicht sehenswert ist. Filmen ist also eine andere Geste als Videografieren und das muss bei der Deutung Berücksichtigung finden. Es macht also einen auch deutungsrelevanten Unterschied, ob ein Video in und für die Forschung eingesetzt oder in und für andere Lebenswelten. Ausführlich dazu Moritz (2011: 34ff).

### 3. Ein Video zu sehen bedeutet, etwas gezeigt zu bekommen

Doch zurück zu der Ausgangsfrage, ob es so etwas wie eine bedeutungsgleiche Repräsentation von Bildern und Videos geben kann. Aus Sicht einer handlungstheoretisch und wissenssoziologisch orientierten *qualitativen Sozialforschung* (und nur um diese geht es hier) ist die Frage nach der bedeutungsgleichen Repräsentation schon falsch gestellt, weil sie impliziert, es ginge der Sozialforschung im Wesentlichen darum, die auf dem Bild befindlichen, geordneten und miteinander verbundenen Flächen, Punkte, Formen, Farben verlustfrei in einen grammatisch, semantisch und pragmatisch korrekten Text zu verwandeln. Natürlich kann das nicht gelingen, und selbst wenn es gelänge, hätte man doch nichts, was für die qualitative Sozialforschung von Bedeutung wäre. Gleiches würde passieren, wenn man versuchte, die Töne eines Musikstücks möglichst verlustfrei in Text zu transformieren. So lange man bei dem Transformationsversuch bei den Einheiten bleibt, die Ausdruck der Materialität des Mediums sind (also den Medieneinheiten), kommt man nicht sehr weit. Jede Transkription ist nicht nur weniger als die Originalszene, sondern immer etwas gänzlich *anderes*. So weit, so bekannt. Gleiches gilt auch für Bilder: Deren Transkriptionen sind nicht nur weniger als das Originalbild, sondern immer etwas gänzlich anderes.

Wer das Wort ‚rot‘ mit dem Medium ‚Bild‘ darstellen will, hat ein ernstes Problem, weil die Transformation des Wortes nicht darin bestehen kann, das Wort auf die Leinwand zu malen oder es abzufotografieren. Denn dann wäre es immer noch Schrift und nicht Bild. Wer die semantische Bedeutung des Wortes ‚rot‘ darstellen will, der hat das Problem nicht. Der malt einen roten Fleck und muss jetzt nur noch ikonografisch klarmachen, dass er die Farbe meint und nicht den Fleck.

Löst man sich also bei dem Problem der Erzeugung einer Repräsentation von den Medieninhalten und fragt nach der Möglichkeit, die Bedeutung der Medieninhalte mit Hilfe eines anderen Mediums zu repräsentieren, dann wandelt sich die eingangs gestellte Frage von der Formulierung, ob man etwas verlustfrei (also identisch) in ein anderes Medium transformieren kann, in die Frage, ob man die von einem menschlichen Akteur mit Hilfe eines Bildmediums zum Ausdruck gebrachte Bedeutung auch mit Hilfe eines anderen Mediums hinreichend genau und insbesondere erkennbar zum Ausdruck bringen kann. Mit ‚erkennbar‘ ist nun nicht gemeint, dass die im bildlichen, in ihrer materialen Konkretheit vollzogene Ausdruckshandlung in ihrer Einzigartigkeit erkennbar werden soll, sondern die in der Ausdruckshandlung *sozial typisierte Handlung*. Und Letzteres ist aus meiner Sicht durchaus machbar.

Ohne Zweifel sind ein Bild und ein Video etwas anderes als die jeweilige sprachliche Beschreibung, aber die entscheidende Frage ist, ob es gerechtfertigt ist zu sagen, dass ‚gezeigt wird, dass und wie ein Mann an seiner Pfeife zieht‘, wenn das Video Bilder eines Mannes zeigt, der gerade an seiner Pfeife zieht.<sup>10</sup> Und natürlich ist das gerechtfertigt. Denn alles, was Bedeutung hat, kann auch (mit Sprechen oder Schreiben, also in Sprache) ausgedrückt werden. Denn die Leitwährung von Bedeutung ist das Sprechen (und damit auch die Sprache). Auch wenn Bilder ohne Zweifel ‚eigen‘ und eigenständig sind, lassen sich die Eigenheiten (und das sehen wir anders als Mitchell 2008: 67) von Bildern in sprachliche Bedeutung ‚ummünzen‘.

Es geht also bei näherer Betrachtung des oben dargestellten Problems um die Frage, ob und wann es reicht, den ‚type‘ (zu den Begriffen ‚token‘ und ‚type‘ siehe Peirce 1976) einer Hand-

10 Nur der Vollständigkeit halber sei gesagt, dass der Betrachter natürlich nicht das Videoband sieht, also den Träger, auf dem die Bilder magnetisch fixiert sind. Zu sagen, man sähe ein Video, bedeutet im allgemeinen Sprachgebrauch, dass man nicht das Video sieht, sondern das, was dem Video technisch eingeschrieben und dann mit Hilfe weiterer Technik (Abspielgerät und Bildschirm) für Menschen sichtbar gemacht wurde.



lung textlich darzustellen oder ob man immer beim token – also der Konkretheit – bleiben muss.

Zentral für meinen Vorschlag ist es, Bilder als *Handlungen* zu begreifen: Mit Bildern zeigen die ‚Macher‘ des Bildes etwas. Die Bildproduktion ist in diesem Verständnis auch und wesentlich eine soziale Zeigehandlung, die etwas bewirken will. Weil es der sozialwissenschaftlichen Bildinterpretation um das Verstehen und Erklären von Handlungen geht, werden – und das ist eine zentrale Besonderheit dieses Ansatzes – ausschließlich Handlungen, also auch *Sprech- und Darstellungshandlungen erfasst und analysiert – nicht ästhetische Strukturen und Muster*. Dabei muss zwischen der *gezeigten Handlung (also der im Bild gezeigten Handlung)* und der *Handlung des Zeigens (also der, mit dem Bild zu zeigen)* unterschieden werden. Mit Ersterem, also der *Handlung vor der Kamera*, wird das Geschehen bezeichnet, das mit Hilfe des Bildes aufgezeichnet und somit im Bildfeld gezeigt wird, mit Letzterem, also der *Kamerahandlung*, der Akt der Aufzeichnung und Gestaltung, also der Akt des Zeigens durch die Gestaltung des Bildes (plus die Gestaltung des von dem Bild Aufgezeichneten).

Ein solch handlungstheoretisches Verständnis von Bildern ist keineswegs selbstverständlich. In der sozialwissenschaftlichen Bild- und Videoanalyse ist nämlich der Handlungscharakter des bildlichen Ausdrucks bislang recht wenig beachtet worden. Viele sozialwissenschaftliche Bild- oder Videoanalysen konzentrieren sich (aus unterschiedlichen Gründen) vor allem auf die *Bildinhalte*: Was und welche Personen sind zu sehen, was tun sie und was bedeutet das Gezeigte? Implizit wird dabei unterstellt, die Interpreten würden selbst unverstellt die Ereignisse sehen, die sie interpretieren, sie wären selbst Zeugen der Ereignisse, sie könnten die Ereignisse so beobachten, wie der/die Videograf/in sie gesehen hat. Oft gerät einer solchen Deutungskunst aus den Augen, dass sie nicht sieht, sondern dass ihr stattdessen etwas von anderen gezeigt wird, mit bestimmten Absichten, nach bestimmten Mustern, mit bestimmten Mitteln.

#### 4. Video als Kommunikation

Das wirft die Frage auf, wer eigentlich der handelnde und kommunizierende Akteur eines Bildes und eines Videos ist. Entgegen dem ersten Augenschein gibt es nicht nur die Handelnden, welche die Kamera zeigt, sondern (wie bereits oben gesagt) auch den Handelnden, der die Kamera in der Hand hält und die Kommandos gibt. Letzterer bestimmt durch seine Aufzeichnung, was vom Handeln vor der Kamera bleibt. Was letztlich also zählt, ist das Handeln des Kameramannes oder anders und allgemeiner: das *Handeln der Kamera*. Was man als Betrachter/in später sieht, ist nur das, was die Kamera ihn oder sie sehen lässt. Die Kamerahandlung ist also keine Zeigegeste in dem Sinne, dass sie sagen würde: ‚Schaut dort einmal hin!‘ und die Angesprochenen schauen in die angezeigte Richtung und machen sich alle ihr eigenes Bild. Nein, die Kamera zeigt nicht auf etwas hin, vielmehr schafft sie selbst ein Bild, das sie dem Betrachter vor die Augen hält. Die Geste der Kamera zeigt sich jedoch nicht selbst, sondern sie schafft mit bestimmten Zeichen eine eigene Weltansicht, sie drückt sich in besonderer Weise aus, sie will etwas vom Betrachter und sie setzt sich immer in Relation zum Betrachter und sie spricht zum Betrachter – selbst dann, wenn sie keine Worte benutzt. *Die Handlung der Kamera besteht also nicht im Zeigen, sondern im Kommunizieren*. Deshalb muss die Analyse von Kamerahandlungen immer auch Kommunikationsanalyse sein.

Das gilt auch für Wissenschaftler/innen, die im Zuge einer *Videografie* bewusst Ereignisse mit der Kamera dokumentieren und später dieses ‚Dokument‘ interpretieren wollen (Knoblauch 2011; Tuma / Schnettler / Knoblauch 2013). Denn auch wissenschaftlich ausgebildete Videografen sind nicht Gott, der mit gleichbleibender und interesseloser Aufmerksamkeit alles wahrnimmt, was sich den Sinnen darbietet, sondern immer wird vor dem Hintergrund der Forschungsrelevanzen Bestimmtes als *zeigenswert* ausgewählt. Alle richten ihre Kamera auf

bestimmte Personen und Ereignisse aus, bleiben dort so lange, wie es ihnen als wesentlich erscheint, fokussieren bestimmte Details, während sie anderes mit der amerikanischen Einstellung oder der Totalen festhalten – kurz: die Videografen zeigen dadurch, dass sie manches von den Ereignissen vor der Kamera in den Bildrahmen packen und anderes nicht, vor allem, was sie für zeigenswert halten; sie zeigen also erst einmal und vor allem *ihre eigenen* impliziten Vorstellungen, was aus ihrer Sicht wichtig und unwichtig ist. Deshalb muss auch diese Forschung grundsätzlich selbstreflexiv angelegt sein und das eigene Zeigehandeln mit interpretieren.

Immer handeln und kommunizieren also (mindestens) zwei Parteien: die *vor* der Kamera und die *hinter* der Kamera (diejenigen, die aufnehmen, die das Aufgenommene bearbeiten und damit abschließend entscheiden, was sehenswert und was unwichtig ist). Allerdings gibt es genre-, medien- und kulturabhängig unterschiedliche Verhaltensweisen der beiden Akteure. Manchmal versucht sich die Kamera unsichtbar zu machen (systematisch auch bei Videografien?), z.B. mittels Standkamera, die immer nur ein Objekt in den Blick nimmt und ihren Standort nicht verändert. Oder aber sie folgt fast unmerklich den Bewegungen oder den Blicken der Handelnden vor der Kamera. Die Kamera scheint dann nur zu beobachten, das ‚Objektive‘ aufzuzeichnen. Ihre Bewegungen sind motiviert durch das ‚Subjekt‘ vor der Kamera.

Sozialwissenschaftliche Ansätze der Bildinterpretation sollten also (insbesondere wenn die Bilder in praktischen Kontexten produziert wurden und praktischen Zwecken, also auch wissenschaftlichen Zwecken dienen) den Handlungscharakter von bildlichen Ausdruckshandlungen in den Blick nehmen. Wer sich bildlich ausdrückt, der kommuniziert, und weil er kommuniziert, handelt er. Dabei taucht bei bildlichen Ausdruckshandlungen die ‚Handlung‘ auf zwei Ebenen auf: Auf der Ebene der Darstellung und der Ebene des Darstellens. Eine bildliche Ausdruckshandlung besteht nämlich in der Regel (also nicht immer) aus einer auf dem Bild dargestellten sozial typisierten und damit erkenn- und beschreibbaren Handlung (z.B. Maria hält ihren toten Sohn im Arm) und einer durch das Bild sichtbar gewordenen Form einer sozial typisierten und damit erkenn- und beschreibbaren Handlung der Bildgestaltung (bestimmte Interpretation des Pieta-Motivs).

Da es immer um sozial typisierte Handlungen geht und deren sozial typisierte Bedeutung, also immer nur um *types* (das Typische) und nicht um *tokens* (das konkrete Einzigartige – siehe dazu erneut Peirce 1976), stellt sich die Frage nach der Transformierbarkeit von Bedeutung also neu und anders. Dann ist aber die Fassung der Bedeutung eines bildlichen Ausdrucks im Textausdruck kein grundsätzliches, sondern ein eher handlungspraktisches Problem.

Stellt man also bei der Frage der Erstellung von Notationen um von dem, was auf dem Bild *gezeigt* wird, hin zu dem, was durch den bildlichen Ausdruck an *Handlung* gezeigt und an Handlung gesetzt wird, dann ist eine Transformation vom Bild in den Text möglich, da Text wie Bild Ausdrucksmedien für sozial Typisiertes sind. Das daraus resultierende zentrale Problem besteht dann in der Bilddeutung und nicht in der angemessenen Vertextung. Allerdings kann ein Vertextungsversuch oder noch besser: eine systematische Vertextungspartitur, also eine Notation, dabei hilfreich sein, die Bedeutung zu ermitteln.

Insofern ist die Vertextung von bildlichem Ausdruckshandeln einerseits Mittel, andererseits Ergebnis der sozialwissenschaftlichen Bildanalyse. Zu einer solchen ‚Vertextung‘ (die natürlich auch mit ikonischen Symbolen arbeiten kann) gibt es bei der wissenschaftlichen Interpretation keine Alternative: Ein Verbleiben beim ‚Vorsymbolischen‘, welches vorgeblich der Eigensinnigkeit des Bildes und den visuellen Kompetenzen Rechnung tragen soll (Bohnsack 2009: 139; Keppler 2006: 104), ist nicht nur ohne Halt und Fundament, sondern auch im wörtlichen Sinne ‚bedeutungslos‘. Ohne Zweifel gibt es jenseits des Symbolischen etwas, aber was genau soll das sein, und: Wie soll es bei der Interpretation Berücksichtigung finden? Was



jenseits der Symbole ist, darüber müssen (um eine verwandte Formulierung Wittgensteins zu entleihen) die Sozialwissenschaftler/innen schweigen.<sup>11</sup>

## 5. Datenfixierung als Fundament der Bildinterpretation

Nach all diesen grundsätzlichen Erwägungen fragt sich, ob dies alles für wissenschaftliche Analysen im Alltag der Forschung von Belang ist oder nur der theoretischen Profilierung dient. Aus meiner Sicht ist die Klärung von erheblicher Bedeutung, entscheidet sich doch (auch) hier, ob die spätere Analyse *valide* ist. Denn von der Klärung der Frage nach der Transformierbarkeit der Bilder und Bewegtbilder hängt (wenn auch nicht allein) die Güte der Forschung ab, da die Güte von Interpretationsarbeiten ganz wesentlich von der Güte der zugrunde gelegten Daten abhängig ist.

Damit entpuppt sich die Klärung der oben diskutierte Frage als wichtiger Teil des komplexen Datenerhebungsprozesses, der selbst wieder aus *drei* Teilprozessen besteht: aus der *Datenerlangung* (Wie kommt man an Daten?), der *Datenzusammenstellung* (Welche qualitativen und quantitativen Eigenschaften hat das Sample?) und der *Datenfixierung* (Wie macht man die Daten für die Interpretationsgruppe und das spätere wissenschaftliche Publikum verfügbar?).

Die hier behandelte Frage ist (wie leicht ersichtlich ist) Teil des oft übersehenen oder als nicht wesentlich angesehenen Prozesses der Datenfixierung. Diesem Schritt im Forschungsprozess kommt jedoch eine besondere Bedeutung zu – wie Soeffners Formulierung aus den Kindertagen der qualitativen Sozialforschung auf den Punkt bringt: „Voraussetzung dafür, daß etwas zum ‚Datum‘ sozialwissenschaftlicher Analyse werden kann, ist, dass es als ‚Dokument‘ einer Handlung oder Lebensäußerung ‚diskursiv‘ vorliegt, d.h. daß es fixiert ist, immer wieder und in gleicher ‚Gestalt‘ von jedem beliebigen Interpreten angesehen, hin- und hergewendet und damit kontrolliert interpretiert werden kann, und daß schließlich aufgrund dieser dokumentarischen Diskursivität des Interpretationsgegenstandes die Interpretation sowie deren Methoden ihrerseits kontrolliert, verifiziert oder falsifiziert werden können.“ (Soeffner 1989: 58)

Soeffner betont, dass die Datenfixierung nicht nur dazu dient, das flüchtige Datum einzufrieren, also dem Datum eine *feste Gestalt* zu geben, sondern dass die feste Gestalt die Voraussetzung dafür ist, *diskursiv* das Datum zu deuten.<sup>12</sup> Zum Dritten ist die Datenfixierung dann unabdingbar, weil nur über diesen Weg für die *scientific community* die Interpretation zu überprüfen ist. Datenfixierung hat also *drei* Ziele, die sich in der konkreten Forschungsarbeit nicht immer zur Deckung bringen lassen: oft wird vieles, was für die Interpretation vor Ort für die Beteiligten, also in der Situation, wichtig war, nicht in die Datenfixierung aufgenommen – einfach deshalb, weil sich vieles (bislang) schlecht fixieren lässt oder weil für die Fixierung bestimmter Bedeutungseinheiten (noch) keine guten Mittel zur Verfügung stehen (z.B. die ‚Stimmung‘ von Musik oder Farben, für Gerüche). Sicher ist jedoch, dass ohne angemessene Datenfixierung die Güte der Forschung und deren Anerkennung wesentlich beeinträchtigt sind.

Dabei sichert die Datenfixierung nicht allein die *Validität* der Datenauslegung, sondern sie ist vielmehr eine wesentliche Voraussetzung dafür. Validität ergibt sich im Wesentlichen aus der Auswertungsmethode (siehe hierzu Loer 2010: 344ff) – jedoch nicht aus ihr allein, sondern Validität ergibt sich erst dann, wenn die Auswertungsmethode mit den richtigen Daten arbeitet

11 Das, was jenseits der Symbole ist, jedoch auf dem Bild zum Ausdruck kommt und zugleich als Ausdruck wahrgenommen wird, wird empirisch bei der konkreten Analyse der Bilder thematisch und kann dann wieder in die Transkription eingetragen werden.

12 Diese beiden Punkte hat Loer im Auge, wenn er im Anschluss an eine schöne Formulierung von Droysen feststellt: „Nur wenn das Material dauerhaft gegenwärtig und zugänglich ist, ist es für die Zwecke der Analyse geeignet.“ (Loer 2010: 327).

– weshalb jede Auswertungsmethode immer auch über eine Theorie der Daten verfügen muss, also die Frage klären muss, was die Daten eigentlich repräsentieren. Und damit schließt sich der Kreis: Datenfixierung ist also nicht nur Mittel der Interpretation, sondern immer auch Möglichkeit der Qualitätskontrolle qualitativer Forschung, und die ist immer mehr vonnöten.

## 6. Anforderungen an eine ‚gute‘ Datenfixierung von Videos

In den letzten Jahren wurde nun eine Reihe von Publikationen vorgelegt, in denen die Möglichkeiten der angemessenen *Videotranskription/-notation* und *Videoanalyse* erörtert und auch verschiedene (vor allem computergestützte) Systeme vorgestellt wurden (siehe den Überblick bei Moritz 2011 und Corsten / Krug / Moritz 2010; Moritz 2013; Tuma / Knoblauch / Schnettler 2013). Abschließend soll versucht werden, vor dem Hintergrund der weiter oben angestellten theoretischen Überlegungen Kriterien für die Einschätzung der Güte von Transkriptions- und Notationssystemen zu formulieren. Natürlich liegt jeder Einschätzung eine normative und theoretische Bestimmung zugrunde. Hier wird im Wesentlichen von der Position einer hermeneutisch arbeitenden Wissenssoziologie heraus argumentiert.

Wichtig ist erst einmal zu sehen, dass die *Videotranskription* sich zwar systematisch von der *Videoanalyse* unterscheidet, die beiden jedoch in der alltäglichen Forschungspraxis kaum voneinander zu trennen sind – weshalb man auch die Transkription *in der Regel* nicht an Externe vergeben, sondern immer als Teil der Forschungsarbeiten betrachten sollte. Eine Videoanalyse ist m.E. ohne Transkription / Notation nicht wirklich sinnvoll, während eine Transkription ohne gleichzeitige Analyse ‚geistlos‘ bleibt, also den ‚Spielsinn‘ nicht erfassen kann. Transkription bzw. Notation ist immer auch Interpretation. Und: Immer bringt eine sorgfältige Transkription oder Notation die Interpretation massiv voran.

Allerdings kann es in bestimmten Fällen nützlich sein, externe Fachkompetenz in die eigene Forschung zu integrieren: Wenn z.B. die Filmmusik eines Videos ‚transkribiert‘ werden soll, ist es sinnvoll, das Video von jemandem transkribieren zu lassen bzw. die Arbeiten extern an jemanden zu vergeben, der überhaupt in der Lage ist, Musik zu transkribieren und zu codieren. Bei der erschließenden Transkription von Musikdaten herrscht in der Sozialwissenschaft und der qualitativen Forschung noch ein ganz erheblicher Nachholbedarf.

Einige der wichtigsten funktionstüchtigen Transkriptions-/Notationsprogramme,<sup>13</sup> die zur Zeit verfügbar sind, heißen: *Moviscript* (Dokumentarische Methode Hampl 2010; Bohnsack 2009), *EXMARaLDA TASX* (Linguistik), *ANVIL* (Linguistik), *Feldpartitur* (Qualitative Sozialforschung, methodenneutral, Moritz 2010), *FOLKER* (Linguistik, Gesprächsforschung, Schmidt 2012), *f4/f5* (Qualitative Sozialforschung, Dresing / Pehl 2012) und seit Neuestem auch *trAVis*, das musikzentrierte Transkriptionsprogramm für audiovisuelle Medienprodukte von Jost et al. (2013).

Die Mehrzahl dieser Programme ist bereits sehr ausgefeilt. Dennoch ist die Lage noch recht unübersichtlich und einheitliche Standards sind noch in weiter Ferne. Dies ist gewiss auch so, weil die jeweiligen Notationssysteme sich aus einer bestimmten Fach- und Problemspektive und sich daraus ergebenden Relevanzen ergeben haben. Deshalb müssen sie auf die jeweilige Fragestellung der eigenen Forschung angepasst werden oder wie Mead zurecht formuliert: „Das Ausmaß, das die Analyse annimmt, hängt von den Erfordernissen des Problems ab“ (Mead 1983: 321). Jede Untersuchung benötigt ihre eigene Schärfentiefe und ihre eigenen

13 Von diesen Transkriptions- und Notationsprogrammen sind die elektronisch gestützten meist von einem spezifischen Forschungsinteresse geprägten Transkriptionsmuster zu unterscheiden. Sie können zwar in ein Programm eingepflegt werden, sind aber selbst keines – wie z.B. Notationsprotokoll von Gesche Joost (Joost 2008), CAFAS von Werner Faulstich und Holger Poggel (Korte / Faulstich 1991) und HaNoS von Jo Reichertz und Carina Englert (Reichertz / Englert 2010 a, 2010 b).

Begriffe. Oft ist es sicherlich hilfreich, wenn das Programm ‚bewährte‘ Begriffe und Zeichen vorhält und vorgibt, doch ohne die Möglichkeit, eigene und neue Begriffe in ein System einzuspeisen, verengt ein System immer den Blick auf die Bilder.

Natürlich muss nicht jede Transkription mit der gleichen Schärfentiefe operieren (siehe zu unterschiedlichen Ebenen und Schärfentiefen der Interpretation Moritz 2011: 59). Oft ist es aufgrund der eigenen Fragestellung sinnvoll, nur das zu notieren, was für die Forschungsfrage relevant ist (niemand sollte mit Kanonen auf Spatzen schießen), und natürlich ist es legitim und auch notwendig, sich nicht für alles zu interessieren. Interpretation ist immer auch Eingrenzung und Begrenzung (vor dem jeweiligen Forschungshorizont).

Dennoch gibt es ein paar Essentials, die bei allen Transkriptions- und Notationssystemen für Videos gegeben sein sollten, legen sie auf die Diskursivität bei der Interpretation und eine diskursive Überprüfung durch die *scientific community* Wert:

1. Die Notation von Videosequenzen sollte grundsätzlich in der *Partiturschreibweise* erfolgen. Allerdings sollte man bei der Anlage der jeweiligen Partitur beachten, dass es nicht zu einer „Wiederversprachlichung“ kommt, die den visuellen Aspekten der Analyse nicht gerecht wird und sich mehr von der Semantik der Sprache als von der Ikonographie und Semiotik der Bilder leiten lässt“ (Knoblauch 2011: 143).
2. Neben der Partitur kommt bei der Deutung dem *Video* eine zentrale Bedeutung zu. Dies bedeutet, dass das Video in der Partitur nicht nur als einzelnes *still*, sondern auch als *laufendes Bild* bei der Analyse *stets verfügbar* sein sollte – dies im Übrigen nicht deshalb, weil das Bild nicht gänzlich zu verschriften ist, sondern *weil* es zu verschriften ist. Weil das laufende Bild immer wieder verfügbar ist, kann die Transkription laufend angepasst werden. Angesichts der neuen Möglichkeiten der digitalen Bildaufzeichnung kann man bei der Auswertung sehr gut mit digitalisierten Bildsequenzen arbeiten.
3. Da Videos zeitlich sequenzierte Handlungen wiederaufführen, muss jede Partitur ebenfalls *sequenziell geordnet* sein. Nur so lassen sich dann Sequenzanalysen anstellen – unabhängig davon, ob man sich der hermeneutischen und deutschen (Loer 2010) oder der konversationsanalytischen, im angelsächsischen Raum verbreiteten Sequenzanalyse (Knoblauch 2011) bedient.
4. Die Erstellung der Partitur ist nicht allein Datenfixierung, sondern immer auch *Datenanalyse*. Deshalb sollte eine Partitur immer ein Arbeitsmittel sein, das nicht nur eine Notation erlaubt, sondern immer auch eine Deutungsarbeit ermöglicht und erleichtert. Die Partitur sollte also in und aufgrund der Arbeit in jeder Hinsicht erweiterbar sein.
5. Da bei unserer Interpretation die *moves* (= kleinste bedeutungstragende Bewegung – siehe dazu Reichertz / Englert 2010 b: 13ff) die kleinsten Sinneinheiten sind, sollte das Notationssystem in der Lage sein, *moves* auch als solche zu erheben und zu symbolisieren. Ein Notationsverfahren sollte also *move* für *move* protokollieren und nicht mehr Bild für Bild. Auf der Ebene der Kamerahandlung können diese *moves* auch größere Handlungszüge sein.
6. Partituren sollten immer *Bedeutungseinheiten* erfassen. Die Leitwährung dieser Bedeutungseinheiten ist die *bedeutungsvolle Sprache*. Eine Codierung von Bildern mit Hilfe von rein deskriptiven Protokollsätzen, die sich scheinbar jeder Sinnschließung enthalten (z.B. Akteur A macht eine vertikale Mikrobewegung mit dem linken oberen *musculus rectus superior*) zum einen grundsätzlich nicht möglich und zum anderen ist sie völlig unübersichtlich und hilft bei der Interpretation nicht weiter. Gleiches gilt für Notation mit Hilfe von einer Fülle vorab genau definierter Ikons. Eine solche Notation bringt nämlich nicht nur bei der Präsentation von Untersuchungsergebnissen Probleme mit sich, sondern auch bei der Analyse der Daten selbst, da die Bedeutung der Bilder

sich im Dickicht der Zeichen verliert.<sup>14</sup> Statt des Bildes hat man nun eine formelhafte Repräsentation des Bildes, wobei in der Interpretation die Formeln wieder in sprachliche Bedeutung rückübersetzt werden müssen, da die Formeln immer nur Kürzel für sprachliche Formulierungen sind.

M.E. wird in den nächsten Jahren ernsthaft zu diskutieren sein, wie weit man die ‚Transformation‘ von audio-visuellem Material in (sprachliche) Zeichen oder codes in der sozialwissenschaftlichen Forschung treiben kann oder zugespitzt: treiben darf. Gegen die Forderung zur vollständigen ‚Verschriftlichung‘<sup>15</sup> um jeden Preis spricht, dass die möglichst exakte schriftliche Fixierung bzw. Codierung audio-visueller Daten schon allein wegen der Vielzahl parallel ablaufender *moves* nur immer ausschnitthaft gelingen kann und zudem oft zu fast sinnleeren ‚Protokollsätzen‘ führt. Was wie intensiv von den Bildern in Sprache, Symbole oder *codes* gefasst werden soll, hängt jedoch ganz wesentlich von der Fragestellung der jeweiligen Forschung ab: Es versteht sich von selbst, dass man z.B. die Erfassung der Emotionen durch die Beobachtung von Gesichtsbewegungen ganz anders codieren muss (siehe hierzu: Ekman 2007 und sein *Facial Action Coding System*) als die Entwicklung der narrativen Strukturen in Videos oder die Aufmerksamkeitslenkung durch den Kameramann in Videografien. Entscheidend sind also Fragestellung und Forschungsinteresse. Deshalb gilt allgemein: So viel wie nötig transkribieren und so wenig wie möglich. Deshalb ist jede Transkription eines Bildes nicht nur eine Einzelanfertigung, sondern auch eine Einzelfallanalyse.

## Literatur

- Bätschmann, O. (1984): Einführung in die kunstgeschichtliche Hermeneutik, Darmstadt.
- Bateson, G. (1987): Geist und Natur, Frankfurt / Main.
- Bauer, M. / Ernst, C. (2010): Diagrammatik: Einführung in ein kultur- und medienwissenschaftliches Forschungsfeld, Bielefeld.
- Bohnsack, R. (2009): Qualitative Bild- und Videointerpretation, Opladen.
- Boll, T. (2012): Mediengebundene Lebenswelten und Beobachten mit anderen Augen, in: Schröer et al. (Hrsg.), *Lebenswelt und Ethnographie*, Essen, S. 411-424.
- Breckner, R. (2010): Sozialtheorie des Bildes, Bielefeld.
- Corsten, M. / M. Krug / C. Moritz (Hrsg.) (2010): Videographie praktizieren, Wiesbaden.
- Dresing, T. / T. Pehl (2012): Praxisbuch Interview & Transkription. Regelsysteme und Anleitungen für qualitative ForscherInnen, 4. Auflage, Marburg.
- Ekman, P. (2007): Gefühle lesen. Wie Sie Emotionen erkennen und richtig interpretieren, München.
- Engell, L. (2010): Playtime. Münchener Film-Vorlesungen, Konstanz.
- Flusser, V. (1994): Gesten. Versuch einer Phänomenologie, Frankfurt / Main.
- Garz, D. (Hrsg.) (1994): Die Welt als Text, Frankfurt / Main.
- Gombrich, E. (1978): Kunst und Fortschritt, Köln.

14 In bestimmten Kontexten, nämlich bei Erkenntnisprozessen, kann eine solche Transformation von Bildern oder Überlegungen in Ikons durchaus erheblichen Sinn machen. Schon Peirce hat auf die kreative Potenz von Diagrammen und diagrammatischem Denken hingewiesen (vgl. Peirce 1976, 1993; auch Reichertz 2007; Bauer / Ernst 2010; Schmidt-Burkhardt 2012).

15 ‚Verschriftlichung‘ wird hier als allgemeiner Begriff verstanden. Bewusst bin ich mir dabei durchaus, dass es (wie weiter oben bereits gesagt) zwischen einer ‚Verschriftlichung‘, einer ‚schriftlichen Fixierung‘ und einer ‚Codierung‘ deutliche Unterschiede gibt. Die Unterschiede zwischen einer Transkription (= Zuordnung eines stellvertretenden Zeichens) und einer Kodierung (= Zuordnung zu einem interpretierenden code) gestalten sich in der Arbeit zwar fließend, sind aber dennoch kategorial voneinander zu unterscheidende Erschließungsprozesse (vgl. auch Moritz 2011).

- Goodman, N. (1997): *Sprachen der Kunst*, Frankfurt / Main.
- Hampl, S. (2010): Videos interpretieren und darstellen, in: M. Corsten / M. Krug / C. Moritz (Hrsg.), *Videographie praktizieren*, Wiesbaden, S. 53-88.
- Hilt, R. (2013): Licht und Schatten in der Videoeigenproduktion von Kindern, in: C. Moritz (Hrsg.), *Videotranskription in der Qualitativen Sozialforschung, Annäherung an einen komplexen Datentypus*, Wiesbaden (im Druck).
- Imdahl, M. (1980): *Giotto. Arenafresken. Ikonographie, Ikonologie, Ikonik*, München.
- Joost, G. (2008): *Bild-Sprache. Die audio-visuelle Rhetorik des Filmes*, Bielefeld.
- Jost, C. / D. Klug / A. Schmidt / A. Reautschnig / K. Neumann-Braun (2013): *Computergestützte Analyse von audiovisuellen Medienprodukten*, Wiesbaden.
- Kaemmerling, E. (Hrsg.) (1991): *Ikonographie und Ikonologie*, Köln.
- Keppeler, A. (2006): *Mediale Gegenwart*, Frankfurt / Main.
- Knoblauch, H. / B. Schnettler / J. Raab / H.-G. Soeffner (Hrsg.) (2006): *Video Analysis. Methodology and Methods*, Frankfurt / Main.
- Knoblauch, H. (2011): Videoanalyse, Videointeraktionsanalyse und Videographie, in: *sozialer sinn* 12, S. 139-145.
- Korte, H. / W. Faulstich (Hrsg.) (1991): *Filmanalyse interdisziplinär*, Göttingen.
- Lobinger, K. (2012): *Visuelle Kommunikationsforschung*, Wiesbaden.
- Loer, T. (2010): Videoaufzeichnung in der interpretativen Sozialforschung, in: *sozialer sinn* 11, S. 319-352.
- Mead, G.H. (1983): *Gesammelte Aufsätze 2*, Frankfurt / Main.
- Mitchell, W. / J. Thomas (2008): *Das Leben der Bilder*, München.
- Moritz, C. (2011): *Die Feldpartitur*, Wiesbaden.
- Moritz, C. (Hrsg.) (2013): *Videotranskription in der Qualitativen Sozialforschung. Annäherung an einen komplexen Datentypus*, Wiesbaden.
- Peirce, C.S. (1976): *Schriften zum Pragmatismus und Pragmatizismus*, Frankfurt / Main.
- Peirce, C.S. (1993): *Semiotische Schriften. Band 3*, Frankfurt / Main.
- Panofsky, E. (1978): *Sinn und Deutung in der bildenden Kunst*, Köln.
- Raab, J. (2008): *Visuelle Wissenssoziologie*, Konstanz.
- Reichert, J. (2007): Der marodierende Blick, in: *sozialer sinn* 2, S. 267-286.
- Reichert, J. (2009): *Kommunikationsmacht*, Wiesbaden.
- Reichert, J. (2012): What you see is what you get. Ein Lob der Oberfläche, in: S. Düchting / K. Plüm (Hrsg.), *Surface*, Münster, S. 107-116.
- Reichert, J. (2013): Zur Gültigkeit der Transkription von Videos, in: C. Moritz (Hrsg.), *Videotranskription in der Qualitativen Sozialforschung*, (im Druck).
- Reichert, J. / N. Ivanyi (2002): *Liebe (wie) im Fernsehen. Eine wissenssoziologische Studie*, Opladen.
- Reichert, J. / C. Englert (2010 a): Kontrolleure in der Trambahn, in: M. Corsten / M. Krug / C. Moritz (Hrsg.), *Videographie praktizieren*, Wiesbaden, S. 25-51.
- Reichert, J. / C. Englert (2010 b): *Einführung in die qualitative Videoanalyse*, Wiesbaden.
- Soeffner, H.-G. (1989): *Auslegung des Alltags – Der Alltag der Auslegung*, Frankfurt / Main.
- Schmidt-Burkhardt, A. (2012): *Die Kunst der Diagrammatik*, Bielefeld.
- Schnettler, B. / H. Knoblauch (Hrsg.) (2007): *Powerpoint-Präsentationen. Neue Formen der gesellschaftlichen Kommunikation von Wissen*, Konstanz.

- Schnettler, B. / H. Knoblauch (2008): Videoanalyse, in: S. Kühl / P. Strodtholz / A. Taffertshofer (Hrsg.), Handbuch Methoden der Organisationsforschung, Hamburg.
- Tuma, R. / B. Schnettler / H. Knoblauch (2013): Videographie. Einführung in die interpretative Video-Analyse sozialer Situationen, Wiesbaden.

Prof. Dr. Jo Reichertz  
Universität Duisburg-Essen  
Institut für Kommunikationswissenschaft  
Universitätsstraße 12  
45117 Essen  
jo.reichertz@uni-due.de