

A blurred, long-exposure photograph of a crowd of people walking through a modern, brightly lit interior space, possibly a shopping mall or office lobby. The image is dominated by blue and white tones, with streaks of light and motion blur creating a sense of dynamic movement. The people are mostly seen from behind or in profile, their forms softened by the blur.

Ralf Bohnsack  
Bettina Fritzsche  
Monika Wagner-Willi (Hrsg.)  
**Dokumentarische Video- und  
Filminterpretation**  
Methodologie und Forschungspraxis

Sozialwissenschaftliche Ikonologie: Qualitative Bild- und Videointerpretation  
Band 3, 2. Auflage | Verlag Barbara Budrich

Sozialwissenschaftliche Ikonologie:  
Qualitative Bild- und Videointerpretation

*Band 3*

Herausgegeben von:

Ralf Bohnsack

Aglaja Przyborski

Jürgen Raab

Thomas Slunecko

Ralf Bohnsack • Bettina Fritzsche •  
Monika Wagner-Willi (Hrsg.)

# Dokumentarische Video- und Filminterpretation

Methodologie und Forschungspraxis

2., durchgesehene Auflage

Verlag Barbara Budrich  
Opladen • Berlin • Toronto 2015

Bibliografische Information der Deutschen Nationalbibliothek  
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Gedruckt auf säurefreiem und alterungsbeständigem Papier.

Alle Rechte vorbehalten.

© 2015 Verlag Barbara Budrich, Opladen, Berlin & Toronto  
[www.budrich-verlag.de](http://www.budrich-verlag.de)

ISBN 978-3-8474-0683-9 (Paperback)  
**eISBN 978-3-8474-0826-0 (eBook)**

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Titelbildnachweis: [www.shutterstock.com](http://www.shutterstock.com)

Umschlaggestaltung: Bettina Lehfeldt, Kleinmachnow – [www.lehfeldtgraphic.de](http://www.lehfeldtgraphic.de)  
Lektorat und Satz: Ulrike Weingärtner, Gründau

# Inhalt

## Einleitung

*Ralf Bohmsack/Bettina Fritzsche/Monika Wagner-Willi*

Dokumentarische Video- und Filminterpretation.....11

## Bildungsforschung: Elementarbereich, Primarbereich und Familie

*Iris Nentwig-Gesemann/Katharina Nicolai*

Dokumentarische Videointerpretation typischer Modi der Interaktionsorganisation im Krippenalltag.....45

*Gerald Blaschke*

„Hier wird richtig gelernt!“ Eine dokumentarische Perspektive auf die Gestaltung des Übergangs in die Schule.....73

*Juliane Lamprecht*

ErzieherInnen und LehrerInnen im Austausch. Videoanalysen zur Sinnkomplexität interaktiver Körperlichkeit.....97

*Margret Xyländer*

Die Familie als Bildungsgemeinschaft. Zur dokumentarischen Videointerpretation einer abendlichen Ritualisierungspraxis.....111

## Bildungsforschung: Unterricht und Schule

*Bettina Fritzsche/Monika Wagner-Willi*

Dokumentarische Interpretation von Unterrichtsvideografien.....131

*Tanja Sturm*

Herstellung und Bearbeitung von Differenz im inklusiven Unterricht. Rekonstruktionen mithilfe der dokumentarischen Videointerpretation....153

*Matthias Martens/Dorthe Petersen/Barbara Asbrand*

Die Materialität von Lernkultur. Methodische Überlegungen zur dokumentarischen Analyse von Unterrichtsvideografien.....179

*Sabine G. Richter*

Klassenmanagement in Übergangssituationen des  
Hauptschulunterrichts. Dokumentarische Videointerpretation  
von Interaktionspraktiken im Umgang mit sozialer (Un-)Ordnung.....207

*Ralf Bohnsack/Juliane Lamprecht*

Das Unterleben in der Pause: die Aneignung von Territorien und  
Artefakten. Eine dokumentarische Videointerpretation.....235

*Astrid Baltruschat*

Unterricht als videografische Konstruktion.....267

## **Öffentliche Medien und Protest im öffentlichen Raum**

*Ralf Bohnsack/Alexander Geimer*

Filminterpretation als Produktanalyse in Relation zur  
Rezeptionsanalyse.....295

*Manel Hell*

Alkoholwerbung versus Alkoholprävention in komparativer  
Videoanalyse. Eine dokumentarische Videointerpretation.....319

*Stefan Hampl*

Zur Rekonstruktion von Montage und Farbkontrasten im Video  
am Beispiel der dokumentarischen Interpretation von Musikvideos.....349

*Julia Ringies*

Implizites Wissen im Dokumentarfilm. Die abbildenden  
BildproduzentInnen im Fokus der Analyse.....387

*Michael Neuber*

Choreografie des Protesthandelns. Körper als Medium  
politischen Protests.....411

## **Transkriptionssysteme**

*Stefan Hampl*

Videotranskription und das System MoViQ.....441

Texttranskription mit TiQ.....465

AutorInnen.....	467
Abbildungsverzeichnis.....	471
Danksagung.....	481
Abbildungen – Farbteil.....	485



# Einleitung



# Dokumentarische Video- und Filminterpretation

Ralf Bohnsack/Bettina Fritzsche/Monika Wagner-Willi

Im Bereich der qualitativen Methoden ist verbalen Phänomenen wie vor allem Gesprächen und Interviews auf dem Wege der Textinterpretation bereits seit den 1960er-Jahren in ihrer Eigenlogik Rechnung getragen worden. Das bedeutet, diese Texte so weit wie möglich unabhängig vom Vor-Wissen, insbesondere vom Kontextwissen der Interpretierenden, aber in genauer Analyse der von den Erforschten *selbst hergestellten* Kontexte zu interpretieren und sie somit als selbstreferentielle Systeme zu verstehen. Diese im Zusammenhang mit dem linguistic turn in den Geistes- und Sozialwissenschaften entstandene methodische Grundhaltung (vgl. Bohnsack 2009: Kap. 3.1) ist in den 1980er-Jahren in den anspruchsvollen rekonstruktiven Verfahren im Bereich der Textinterpretation zunehmend selbstverständlich geworden. Die wesentlichen Fortschritte im Bereich der qualitativen Methoden in den letzten dreißig Jahren waren eng mit den Verfahren der Textinterpretation auf der Basis eines Verständnisses von Texten als selbstreferentiellen Systemen verbunden.

Eine Vorreiterrolle kam hier der Konversationsanalyse von Harvey Sacks (1995) bzw. Sacks/Schegloff/Jefferson (1978) zu, die deshalb im Folgenden besondere Erwähnung finden soll. Die Konversationsanalyse hat die Entwicklung der Verfahren der Textinterpretation wesentlich beeinflusst. Seit den 1990er-Jahren hat sie auch im Bereich der Videoanalyse Bedeutung gewonnen. Umso erstaunlicher ist es, dass die sich in dieser Tradition verortenden Ansätze der Videoanalyse, wie sie grundlegend von Charles Goodwin (2001) und Christian Heath (1997) entwickelt worden sind, dem Visuellen oder genauer dem Bildhaften oder Ikonischen<sup>1</sup> einen derartigen Status selbstreferentieller Systeme dezidiert aberkennen. So betont Charles Goodwin (2001: 157), einer der bekannten Namen der qualitativen Videoanalyse, „dass visuelle Phänomene nur erforscht werden können, indem ein vielfältiges Set von semiotischen Ressourcen berücksichtigt wird [...]. Viele von diesen, wie z.B. die Struktur, die durch laufende Gespräche beigesteuert wird, sind in keinerlei Hinsicht visuell, aber die sichtbaren Phänomene [...] können ohne sie nicht zureichend analysiert werden“.<sup>2</sup>

- 
- 1 Genauer betrachtet, geht es – wenn wir der Differenzierung von Charles S. Peirce (1931-1935) folgen – hier nicht um *visuelle* (obschon diese immer wieder so genannt werden), sondern um *ikonische* Phänomene oder Zeichen (vgl. auch Eco 1994: 197ff.). Denn auch sprachliche Zeichen (welche im Sinne von Peirce zu den von den *ikonischen* zu unterscheidenden *symbolischen* Zeichen zählen) werden, wenn sie in schriftlicher Form vorliegen, *visuell* wahrgenommen.
  - 2 Zitate aus nicht deutschsprachiger Literatur sind dort, wo der nicht deutschsprachige Titel zitiert wird, von uns selbst ins Deutsche übertragen worden.

## *Die dokumentarische Videoanalyse im Unterschied zu einer Videografie in Ergänzung zur Gesprächsanalyse*

Während Goodwin es also nicht für möglich hält, visuelle Phänomene bzw. ikonische Phänomene ohne Berücksichtigung von Gesprächen, von verbalen Äußerungen, zu analysieren, findet sich demgegenüber in der Konversationsanalyse eine lange Tradition, Gespräche, also verbale Phänomene, ohne Berücksichtigung anderer semiotischer Ressourcen, insbesondere sichtbarer Phänomene, zu interpretieren. Die Ausklammerung dieser anderen Ressourcen wird mit der strikten empirischen Fundierung in Audiografien („audio recordings“) in gewisser Weise zum methodischen Prinzip erhoben (vgl. Sacks/Schegloff/Jefferson 1978: 8).<sup>3</sup> Eine methodologische Begründung für diese Ungleichgewichtigkeit des Sprachlichen und des Ikonischen finden wir in den Videoanalysen in der Tradition der Konversationsanalyse aber nicht.

Im Übrigen muss diese Auseinandersetzung auch nicht allein und nicht einmal primär methodologisch-grundlagentheoretisch entschieden werden. Entscheidend ist für uns die empirische Evidenz von exemplarischen Analysen, in denen gezeigt werden kann, dass grundlegende Orientierungen der Akteure in einer Videoanalyse auch ohne deren verbale Äußerungen herausgearbeitet werden können, wie sich dies auch in diesem Band findet (in besonders ausgeprägter Weise in den Beiträgen von Bohnsack/Lamprecht und von Hampl zur Montage) und wie Bohnsack (2009: Kap. 6 u. 7) es am Beispiel einer Fernsehshow exemplarisch demonstriert hat.

Die in der Konversationsanalyse fundierte Videoanalyse, die in der Sozialwissenschaft große Bedeutung gewonnen hat, ist somit primär in ihrer *ergänzenden Funktion zur Gesprächsanalyse* entwickelt worden.<sup>4</sup> Das Bild stellt hier also lediglich eine Ergänzung des Textes dar. Das gilt z.B. für die „Video-Interaktions-Analyse“ von Hubert Knoblauch (2005) und auch für die Video- resp. Filmanalyse von Angela Keppler (2006), die ebenfalls beide in der Tradition der Konversationsanalyse stehen. Ganz allgemein dominiert in der derzeitigen Praxis der sozialwissenschaftlichen Videoanalyse eine Vorgehensweise, welche die Dimension des Bildhaften lediglich ergänzend zur Ana-

---

3 Eine Ausnahme hierzu bilden insbesondere Arbeiten, wie sie von Kendon (1990) vorgelegt wurden. So werden dort in die Transkription von Gesprächen nonverbale Elemente wie Blickausrichtung oder Körperbewegungen integriert und in ihrer Bedeutung und Funktionalität für sprachliche Äußerungen untersucht. Dies trifft auch auf Videoanalysen von Heath (z.B. einer Arzt-Patient-Interaktion: 1997) zu. Gleichwohl bleiben diese Arbeiten insofern dem Prinzip der Selbstreferentialität des Sprachlichen resp. Textlichen verhaftet, als sie die Funktionalität nonverbaler Äußerungen für die Konversation fokussieren.

4 Erste Versuche, darüber hinauszugehen, finden sich auch bei Dinkelaker/Herrle (2009). Im Bereich der Dokumentarischen Methode finden sich ebenso Analysen, bei denen die Bilddimension in lediglich ergänzender Funktion zur Gesprächsanalyse einbezogen wird. Allerdings verstehen sich diese Arbeiten selbst als Gesprächsanalysen, so etwa das videogestützte Gruppendiskussionsverfahren bei Iris Nentwig-Gesemann (2006).

lyse des Verbalen hinzuzieht. Demgegenüber bildet in der Dokumentarischen Methode der Zugang zur Eigenlogik des Bildlichen bzw. Visuellen und der körperlichen Ausdrucksformen – einschließlich der Positionierung des Körpers im Raum und in Relation zu Objekten – einen der Schwerpunkte der Videoanalyse. In dieser Hinsicht handelt es sich um einen Zugang, der besonders geeignet ist, das Potenzial der Videoanalyse zur Erforschung der körperlich-materiellen und räumlichen Aspekte des Performativen auszuschöpfen (vgl. Wulf u.a. 2001; Fritzsche/Wagner-Willi 2013: 276). Wir müssen uns über die Konsequenzen im Klaren sein, welche eine lediglich ergänzende Funktion des Visuellen für die Analyse des Körperlichen bzw. die Rekonstruktion der Eigenlogik des körperlichen Ausdrucks, also der Bewegungen derjenigen, die Gegenstand einer Videografie oder die abgebildete AkteurInnen eines Filmes sind, hat.

### *Videografien und Filme als Eigenprodukte der Erforschten und als Erhebungsinstrumente der Forschenden*

In der empirischen Forschung lassen sich grundsätzlich zwei Arten von Filmen und Videos<sup>5</sup> unterscheiden: Auf der *einen Seite* haben wir es mit solchen Filmdokumenten zu tun, die in den jeweiligen kulturellen Lebenszusammenhängen oder Milieus, welche Gegenstand der sozialwissenschaftlichen Analyse sind, selbst produziert wurden. Dies gilt einmal für den *privaten* Bereich – beispielsweise für Familienvideos, die von Angehörigen der Familie aufgenommen worden sind. Diese geben uns dann Aufschluss über die Handlungspraxis und den Erfahrungsraum, das Milieu bzw. den Habitus einer Familie.

Ähnliches gilt aber auch für Filme und Bilder aus dem *öffentlichen* Bereich, also für den Bereich der *Massenmedien*, beispielsweise für Fernsehsendungen oder Werbefilme. Hier erhalten wir u.a. Einblicke in die durch die Medien transportierten oder propagierten Lifestyles sowie – je nach Vergleichshorizont – den Habitus derjenigen, die Urheber des Filmprodukts oder beispielsweise Moderator einer Fernsehshow sind (vgl. Bohnsack 2009: Kap. 6).

Von jenen visuellen Grunddaten, die insgesamt *Eigenprodukte der Erforschten* sind, lassen sich auf der *anderen Seite* solche unterscheiden, die zu *Forschungszwecken* produziert werden. Letzteres ist dort der Fall, wo – wie etwa bei Unterrichtsvideos – Videografien *als Erhebungsinstrument* der (sozial-)wissenschaftlichen Forschung Verwendung finden.

---

5 Bei der Verwendung der Begriffe „Video“ und „Film“ orientieren wir uns am Alltagssprachgebrauch, demzufolge Videos vor allem als Life-Aufzeichnungen verstanden werden – von Alltagssituationen wie aber auch medialen Ereignissen (bspw. Fernsehshows). Demgegenüber folgen Filme einem Drehbuch und sind das Produkt vielfacher Korrekturen und Überarbeitungen nicht nur seitens der Akteure hinter der Kamera, sondern auch im Bereich der Ereignisse vor der Kamera.

Im letzteren Fall liegt der Fokus der Analyse auf den Gestaltungsleistungen der *abgebildeten* Bild- resp. VideoproduzentInnen (vgl. ursprünglich: Bohnsack 2003). Das sind diejenigen, die *vor* der Kamera agieren, die also abgelichtet werden. Demgegenüber interessieren solche Videografien und Filmdokumente, die von den Erforschten selbst produziert worden sind, sowohl hinsichtlich der Gestaltungsleistungen der *abgebildeten* BildproduzentInnen wie auch derjenigen der *abbildenden* BildproduzentInnen. Das sind diejenigen, die *hinter* der Kamera agieren, und diejenigen, die auch noch nach der Aufzeichnung an der Bearbeitung von Film und Video beteiligt sind.

### *Abgebildete und abbildende Bild- resp. VideoproduzentInnen*

Die sich aus der komplexen Relation dieser beiden unterschiedlichen Arten von Bildproduzierenden<sup>6</sup> ergebenden methodischen Probleme sind dann leicht zu bewältigen, wenn beide zu *demselben* Erfahrungsraum, d.h. zum selben Milieu gehören. Dies ist beispielsweise dann der Fall, wenn ein Angehöriger der Familie ein Familienfoto oder -video produziert (Bohnsack 2009: Kap. 4.3) oder wenn Lehrer oder Schüler selbst ein Video über ihren eigenen Schulalltag produzieren (Baltruschat 2010). Im Falle von öffentlichen medialen Produktionen (Filmen und Fernsehshows) sowie von Milieufremden und von Forschenden aufgezeichneten Videos stellt uns die Relation der Gestaltungsleistungen von abbildenden und abgebildeten BildproduzentInnen vor erhebliche Probleme, da diese Relation weder in der wissenschaftlichen Videografie noch in der Filmwissenschaft umfassend bearbeitet wurde. Letztere befasst sich überwiegend mit den Gestaltungsleistungen der abbildenden BildproduzentInnen (Montage und Kameraeinstellung), erstere fast ausschließlich mit denjenigen der abgebildeten BildproduzentInnen (Bohnsack 2009: Kap. 5.6 u. 5.3).

Zwar gehören die Gestaltungsleistungen der *abbildenden* BildproduzentInnen im Falle der zu Forschungszwecken erstellten Aufzeichnungen nicht zum eigentlichen Forschungsgegenstand. Gleichwohl sollten diese zumindest ansatzweise in Rechnung gestellt, also methodisch kontrolliert werden (siehe dazu auch die Beiträge von Baltruschat sowie Bohnsack/Lamprecht und Fritzsche/Wagner-Willi in diesem Band). Dies betrifft u.a. die Kameraführung und -positionierung, die Wahl der Perspektivität und der Einstellungsgröße wie auch eventuelle Montageleistungen. Während im Falle der Fotos und Videos, die von den Erforschten selbst produziert worden sind, der Zugang zur Eigenlogik des Produkts im umfassenden Sinne – auf der Ebene der Gestaltungsleistungen sowohl der abgebildeten wie abbildenden BildproduzentInnen

---

6 Wir verwenden hier den Begriff der BildproduzentInnen auch dort, wo wir es mit Videos oder Filmen zu tun haben, da das Problem der Verschränkung und Überlagerung der Gestaltungsleistungen von Abbildenden und Abgebildeten nur die Bilddimension der Medien Video und Film betrifft und nicht deren verbale Dimension.

– interessiert, wird den Gestaltungsleistungen der letzteren im Falle von Fotos und Videos als Erhebungsinstrument nur zum Zwecke der methodischen Kontrolle Rechnung getragen. Hier steht der Zugang zum Ikonischen primär im Dienste der Eigenlogik der Gestaltungsleistungen im Bereich der Bewegungen und körpergebundenen Ausdrucksformen der Abgebildeten.

### *Produkt- im Unterschied zur Rezeptionsanalyse*

Neben der Differenzierung von abgebildeten und abbildenden BildproduzentInnen und derjenigen von Videos als Forschungsinstrument und als Eigenprodukt der Erforschten gilt es im Zuge einer methodischen oder methodologischen Präzisierung darüber hinaus, die *Analyse des Produkts*, also die Analyse des Films oder Videos selbst, von der Analyse der Rezeption des Produkts, also von der *Rezeptions- oder Mediennutzungsforschung*, zu unterscheiden. Im Falle der *Produktanalyse* zielt das Erkenntnisinteresse auf den Habitus und den Modus der Erfahrungskonstitution bzw. den Erfahrungsraum der *ProduzentInnen* und im Falle der *Rezeptionsanalyse* auf den Habitus und den Modus der Erfahrungskonstitution und den Erfahrungsraum der *Rezipierenden* sowie auf dessen mögliche Veränderungen (siehe zu dieser Differenzierung den Beitrag von Bohnsack/Geimer in diesem Band sowie Bohnsack 2009: Kap. 5.2). Auch im Bereich der *dokumentarischen* resp. *praxeologischen Rezeptionsanalyse* finden sich inzwischen umfangreiche Vorarbeiten (u.a. Geimer 2010a, 2010b; Michel 2005 u. 2006; Schäffer 2003).

### *Polysemie, Mehrdimensionalität und komparative Analyse*

Die wissenschaftliche Analyse eines Bild- oder Videoprodukts unterscheidet sich von der Common-Sense-Interpretation unter anderem auch durch die methodische Variation empirischer Vergleichshorizonte. Je nach *Vergleichshorizont* geraten unterschiedliche – einander aber nicht ausschließende – Dimensionen oder Erfahrungsräume des Bildes oder Videos in den Blick. Um es an einem Beispiel zu erläutern: Wenn wir ein von Lehrpersonen über ihren Schulalltag produziertes Video vor dem Vergleichshorizont eines Videos von SchülerInnen interpretieren, geraten uns milieutypische Differenzen innerhalb der Schule in den Blick. Wenn wir dieselben beiden Videos mit denen von Lehrpersonen und SchülerInnen einer anderen Schule vergleichen, gerät wiederum eine andere Dimension in den Blick: Es können nun durchaus Gemeinsamkeiten zwischen den zunächst lediglich different erscheinenden Milieus von Lehrpersonen und SchülerInnen derselben Schule sichtbar werden, die auf eine übergreifende Schulkulturtypik verweisen.

Auf diese Weise erscheint die Polysemie desselben Videos nicht – wie bei Roland Barthes 1990: 34) und auch in den Cultural Studies (vgl. u.a. Hepp

2004) – primär als methodologisches und methodisches Hindernis, sondern als entscheidende Herausforderung für die methodische Weiterentwicklung im Sinne einer Typenbildung. In der Dokumentarischen Methode wird dieser Herausforderung mit der komparativen Analyse und der *Mehrdimensionalität* der Typenbildung Rechnung getragen (vgl. Bohnsack 2010a). Die Produktanalyse der Dokumentarischen Methode erhöht ihre Validität bereits auf elementaren Ebenen der Interpretation durch die komparative Analyse, also die *systematische* Variation explizit eingeführter Vergleichshorizonte (siehe dazu auch Hampl 2010 u. 2015). Auf diese Weise wird dann auf einer höheren Ebene auch die mehrdimensionale Typenbildung möglich.

### *Ikonografie, Ikonologie und der Zugang zu Bild und Text als selbstreferentielle Systeme*

Wesentlich für die Dokumentarische Methode ist neben der methodischen Kontrolle der Vergleichshorizonte und des Kontextwissens aber vor allem der Wechsel von der Ikonografie zur Ikonologie, also von der Frage nach dem *Was* zur Frage nach dem *Wie* (genauer dazu weiter unten). All dies ermöglicht im Zusammenhang mit der „Einklammerung des Geltungscharakters“ (Mannheim 1980: 88), d.h. der Einklammerung von Ansprüchen der faktischen Wahrheit und normativen Richtigkeit (dazu u.a.: Bohnsack 2013: 177), und der systematischen Rekonstruktion der Formalstruktur den Zugang zum Bild und Video als selbstreferentiellem System und konstituiert den Wechsel der AnalyseEinstellung von der Common-Sense-Interpretation zur wissenschaftlichen Analyse.

Die Differenz zu den Theorien und Interpretationen des Common Sense, des Alltags, wird in der Dokumentarischen Methode also nicht mit dem Anspruch einer ‚höheren‘ Rationalität der sozialwissenschaftlichen Analyse begründet, sondern, wie bereits dargelegt, mit dem Wechsel der *AnalyseEinstellung*. Diese AnalyseEinstellung ermöglicht – über die Theorien und das *explizite* Wissen der Erforschten und deren Rekonstruktion hinausgehend – den Zugang zu deren *implizitem* Wissen, insbesondere deren handlungsleitenden Wissensbeständen.

Im Bereich des sogenannten interpretativen Paradigmas wie auch der Cultural Studies, welche in der sozialwissenschaftlichen Video- und Filmanalyse besonders einflussreich sind, findet sich ein derartiger Bruch mit dem Common Sense wie auch der Zugang zum impliziten Wissen in systematischer Weise nicht (genauer dazu Bohnsack 2009: Kap. 5.3). Im Bereich von Bild und Film erschließt sich der Zugang zum impliziten Wissen erst dann, wenn wir der *Eigenlogik* dieses visuellen Materials insbesondere in seiner Differenz zu Sprache und Text Rechnung tragen.

## Die Bewegungen der abgebildeten BildproduzentInnen

Den ikonischen Phänomenen in ihrer Eigenlogik im Sinne selbstreferentieller Systeme Rechnung zu tragen, ist auch Voraussetzung für den validen Zugang zu den Bewegungen, den körperlichen Ausdrucksformen, der abgebildeten BildproduzentInnen in ihrer Eigenlogik. Für die Klassiker der Filmwissenschaft Siegfried Kracauer (1964) und Béla Balázs (2001) vermochte das Medium Film einen bisher nicht gekannten Zugang zu elementaren Ebenen sozialer Realität zu eröffnen, nämlich zur Ebene der inkorporierten Gesten, der Gebärden und der Mimik. Für Siegfried Kracauer (1964: 387) geht es darum, „der Realität nahezukommen, wenn wir ihre untersten Schichten durchdringen“. Im Kontrast zu den klassischen Filmtheorien von Balázs und Kracauer, in denen dieser Ebene der Analyse, die in der Sprache der Kunstgeschichte als vor-ikonografische bezeichnet werden kann, noch eine zentrale Bedeutung zukam, stehen in der modernen Filmwissenschaft die Gestaltungsleistungen der abbildenden BildproduzentInnen im Zentrum (siehe dazu weiter unten).

Die elementaren Schichten oder Ebenen im Sinne von Balázs und Kracauer hat Erwin Panofsky (1975), der nicht nur als Klassiker der Kunstgeschichte, sondern auch der Filmwissenschaft gilt, auch als diejenigen „primärer oder natürlicher Bedeutungen“ oder als *vor-ikonografische* im Unterschied zur ikonografischen Ebene bezeichnet. Es ist diese vor-ikonografische oder – in der Sprache der Semiotik – die *denotative* Ebene, deren genaue Beobachtung und Beschreibung die wesentliche Grundlage der ikonologischen Interpretation und der dokumentarischen Bild- und Videointerpretation darstellt. Auf der ikonografischen Ebene fragen wir mit Bezug auf die Bewegungen der abgebildeten BildproduzentInnen danach, *was* das für eine Handlung ist (bspw. ein „Gruß“). Wir müssen somit Motive, genauer Um-zu-Motive, unterstellen (bspw.: „ich hebe die Hand, um zu grüßen“ oder „ich beuge den Rumpf, um mich zu setzen“). In dokumentarischer oder ikonologischer Interpretation stellt sich die Frage nach dem *Wie* der Herstellung dieser Handlung (bspw. ist das Heben der Hand „präventiv“, „unsicher“, „starr“). Einen Zugang zum *Wie* der Handlung vermittelt mir die genaue Beschreibung auf der vor-ikonografischen Ebene.

Ray L. Birdwhistell (1968: 380), der Klassiker der Bewegungsanalyse, erläutert das *Wie* am Beispiel des militärischen Grußes. Diese Handlung – ob schon hoch standardisiert – erhält eine enorme Variabilität weitergehender Bedeutungen durch das *Wie* ihrer Herstellung: „Durch den Wechsel in Haltung, Gesichtsausdruck, der Geschwindigkeit oder Dauer der Bewegung des Grüßens und sogar in der Wahl ungeeigneter Kontexte für die Handlung kann der Soldat den Empfänger des Grußes ehren, herabwürdigen, zu gewinnen versuchen, beleidigen oder befördern“.

Die Bewegungsabläufe auf der vor-ikonografischen Ebene differenzieren wir noch einmal in *Gebärden* oder *Gesten*<sup>7</sup> einerseits und *operative Handlungen* andererseits (siehe dazu die Grafik). Träger der *Gesten* können die Extremitäten sein (bspw. ‚Ausstrecken des Armes‘), der Rumpf (bspw. ‚Drehen oder Beugen des Rumpfes‘), der Kopf (bspw. ‚Senken des Kopfes‘), aber auch die Mimik (bspw. ‚Lächeln‘). Die Gesten lassen sich ihrerseits noch einmal in ihre *Elemente* differenzieren, die wir mit Birdwhistell (1952) als „Kineme“ bezeichnen. Auch die ersten videogestützten Analysen auf der Grundlage der Dokumentarischen Methode haben auf der vor-ikonografischen Ebene angesetzt, mit einer genauen Rekonstruktion von Gesten und operativen Handlungen und teilweise deren Elementen, den Kinemen. Dies gilt für die Analysen von Monika Wagner-Willi (u.a. 2005 u. 2013) über Schülerinteraktionen im Klassenraum oder auch für die Arbeit von Amelie Klambeck (2007) über PatientInnen mit „psychogenen Bewegungsstörungen“ und deren Verhalten während der Chefarztvisite.



Abbildung 1: Korporierte Praktiken der abgebildeten BildproduzentInnen

*Operative Handlungen* (bspw. ‚Sich-Hinsetzen‘, ‚Aufzeigen‘, ‚Hose-Hochziehen‘) umfassen in der Regel mehrere Gesten in ihrer Sequenzialität, die sich

7 Die Begriffe Gebärde und Geste werden hier synonym verwendet (dazu auch: Bohnsack/Przyborski 2014).

wiederum aus simultan oder sequenziell verbundenen Kinemen zusammensetzen. Wesentliches Merkmal der Unterscheidung der operativen Handlungen von den Gesten im elementaren Sinne ist, dass sie – und aus diesem Grund möchten wir sie bereits als Handlungen bezeichnen – mit zweckrationalen Motivkonstruktionen, d.h. mit Konstruktionen von Um-zu-Motiven, versehen werden können, wenn auch möglicherweise nur in rudimentärer Weise. Dies geschieht bspw. derart, dass die einzelne Geste (bspw.: ‚Beugen des Rumpfes‘, ‚Strecken des Armes‘) „selbst nur Mittel im Sinnzusammenhang eines Entwurfes“ ist (Schütz 1974: 119), beispielsweise des Entwurfes ‚Sich-Hinsetzen‘ oder des Entwurfes ‚Aufzeigen‘.

Durch diesen Entwurf, durch diese zweckrationale Konstruktion eines Um-zu-Motivs (‚B beugt den Rumpf, um sich hinzusetzen‘; ‚A streckt den Arm, um aufzuzeigen‘), werden das Sich-Hinsetzen und das Strecken des Armes zu einer *Handlung*. Hier lassen sich dann jeweils weitere Hierarchien von Um-zu-Motiven konstruieren, also bspw.: ‚A zeigt auf, um seinen Redewunsch zu signalisieren‘, d.h., um sich zu melden‘. Allerdings ist der Handlungsentwurf beim letzteren Beispiel nicht direkt am Bewegungsverlauf *beobachtbar*, er muss vielmehr als Entwurf, als Um-zu-Motiv, auf der Grundlage normativer Erwartungen und Rollenbeziehungen, d.h. durch die Einbindung in einen institutionellen Kontext, unterstellt oder attribuiert werden. In diesem Fall bewegen wir uns bereits auf der *ikonografischen Ebene*.

Nunmehr kann eine Eigenart der ikonologischen oder dokumentarischen Interpretation deutlich werden, die von elementarer Bedeutung ist (siehe dazu auch die Grafik). Dieselbe Bewegung (bspw. das ‚Heben der Hand bei mehr oder weniger gestrecktem Arm‘) kann immer auf zwei Sinnebenen zugleich interpretiert werden: zum einen zweckrational im Rahmen der Konstruktion eines Um-zu-Motivs (des ‚Sich-Meldens‘), mit der wir uns auf die Suche nach dem *subjektiv gemeinten Sinn* begeben. Zum anderen und zugleich kann die Geste auch (wenn wir nach dem *Wie* ihrer Herstellung fragen) als *Dokument* für das *Wesen* oder den *Habitus* des Akteurs (‚Fleiß‘, ‚Ehrgeiz‘, ‚Aufmerksamkeit‘, ‚Konkurrenz‘) interpretiert werden. Entscheidend ist dann, *wie* sich jemand hinsetzt oder *dass* er oder sie sich setzt: „Nicht das ‚Was‘ eines objektiven Sinns, sondern das ‚Daß‘ und das ‚Wie‘ wird von dominierender Wichtigkeit“ (Mannheim 1964: 134).

Das *Wie*, also der ikonologische oder dokumentarische Sinngehalt, erschließt sich vor allem auf der Grundlage einer Rekonstruktion der durch die Erforschten *selbst hergestellten Kontexte*, der *eigenen Kontextuierungen* der Akteure im Forschungsfeld. Die einzelnen Bewegungen und Äußerungen sind grundsätzlich in ihren ‚natürlichen‘ Kontexten zu erfassen. Dies gilt auch für die Textinterpretation und ist wesentliches Merkmal des Zugangs zu Text und Bild als selbstreferentiellen Systemen. So wird eine Äußerung bspw. erst im Kontext einer Erzählung adäquat interpretierbar. Diese wiederum ist im Kontext einer Diskussion unter Jugendlichen produziert worden und erschließt sich demzufolge in ihrer kollektiven, also gruppen- und milieuspezifischen Bedeu-

tung nur unter Berücksichtigung des Kontextes der *Reaktionen* (Redebeiträge) der anderen am Diskurs Beteiligten. Im Bereich der Analyse von Videos und Bildern bedeutet dies, die einzelne (nonverbale) Geste im *sequenziellen* Kontext jener Gesten zu interpretieren, auf welche die zu interpretierende eine Reaktion darstellt, wie aber auch derjenigen Gesten, die auf sie reagieren. Wesentlich ist aber nicht nur der *sequenziell*, sondern auch der *synchron* oder *simultan* hergestellte Kontext. Am Beispiel des ‚Aufzeigens‘ oder ‚Sich-Meldens‘ geht es hier unter anderem um die im jeweiligen Augenblick je gleichzeitige Haltung der Hand und der Finger in ihrer Beziehung zu derjenigen des Armes, des Oberkörpers und der Mimik.

Bei Christian Heath, einem der Wegbereiter der Videoanalyse, der in der Tradition der Konversationsanalyse steht, wird die Bedeutung des Kontextes für die Interpretation nonverbaler Kommunikation herausgestellt: „Unglücklicherweise findet sich die weit verbreitete Tendenz unter den Forschern nonverbaler Kommunikation, anzunehmen, dass die Bedeutung einer Geste unentwirrbar an ihre physische Form gebunden ist, statt an den Kontext, in dem sie auftritt“ (Heath 1997:187). Aber bei Heath wird die Kontextuierung dann – in Orientierung an der Kontextuierung sprachlicher Äußerungen – einseitig als eine sequenzielle verstanden.

Demgegenüber hat Ray Birdwhistell, der in der Palo-Alto-Schule zu verorten ist, zu der auch Bateson und Watzlawick gehören, bereits in den 1950er-Jahren bahnbrechende Arbeit im Hinblick auf die simultane Kontextuierung nonverbalen Handelns geleistet. Er hat in seinen empirischen Analysen gezeigt, dass sich der Sinngehalt von Gesten, den sogenannten Kinemorphemen, erst aus der genauen Rekonstruktion der sie konstituierenden Elemente, der Kineme bzw. Kine, erschließen lässt. Dies gilt auch für die Mimik, bspw. für die mimische Geste des Augenzwinkerns („wink“), die Birdwhistell (1952: 19) folgendermaßen beschreibt: „a) Das rechte Augen ist geschlossen, während das linke geöffnet bleibt. – b) Die Mundhaltung ist ‚normal‘. – c) Die Nasenspitze ist eingedrückt (Kaninchennase). – d) Der linke Augenhöhlenrand ist schräg („squinted“)“.

Diese korporierten Praktiken auf der Ebene von Kinemen müssen sich gleichzeitig, synchron oder simultan, vollziehen, um ihren Ausdruckscharakter, d.h. ihre ikonologische Bedeutung, zu erhalten (vgl. dazu auch den Beitrag von Bohnsack/Lamprecht in diesem Band). Die Geste des „Zwinkerns“ entfaltet ihre Signifikanz nur dann, wenn ihre Konstituentien, die elementaren Bewegungseinheiten der Kineme, *synchron* oder *simultan* zum Ausdruck kommen. Hiermit ist zunächst das synchrone Zusammenspiel innerhalb abgegrenzter Bereiche des Körpers gemeint (vgl. Birdwhistell 1952: 17). Idealerweise umfasst die Rekonstruktion des simultanen Zusammenspiels aber den gesamten Körper, orientiert sich an der „whole body‘ conception“, wie Ray Birdwhistell (ebd.: 8) betont, und bezieht schließlich auch die räumliche und interaktive *Positionierung der Körper zueinander*, also die szenische Choreografie, mit ein („A dreht B den Rücken zu“).

Hier wird eine für die Videointerpretation entscheidende methodische Herausforderung erkennbar: Die Interpretation der Semantik von korporierten Praktiken setzt nicht nur deren Rekonstruktion im *sequenziellen* Ablauf voraus, sondern auch in ihrer *Synchronizität* oder *Simultaneität* (zur genaueren Unterscheidung siehe weiter unten). Diese erschließt sich in valider Weise aber allein auf der Grundlage von ‚eingefrorenen‘ Bildern, also von Standbildern, die auch als stills oder Fotogramme bezeichnet werden. Es war unter anderem diese Bedeutung der Fotogramme für die Interpretation der Gesten und der Mimik, welche den Klassiker der Semiotik, Roland Barthes, veranlasst hat, diesen eine zentrale Bedeutung für die Semiotik des Filmes einzuräumen. Nach Barthes (1990: 64) „lässt sich in gewissem Maß [...] das Filmische paradoxerweise nicht im Film ‚am rechten Ort‘, ‚in der Bewegung‘, ‚in natura‘ erfassen, sondern bisher nur in einem wichtigen Artefakt, im Fotogramm.“

Methodologisches und methodisches Prinzip der dokumentarischen Interpretation ist es, diese Frage nach dem *Wie*, nach dem *modus operandi*, dem *Habitus*, zugleich auf unterschiedlichen Dimensionen desselben Falles zu stellen, sodass diese unterschiedlichen Interpretationsdimensionen einander wechselseitig zu validieren vermögen. Im Falle der Interpretation der Bewegungen der abgebildeten BildproduzentInnen sind dies vor allem die unterschiedlichen Ebenen einer Bewegung, diejenigen der *Kineme*, der Gesten, der operativen Handlung und der institutionalisierten Handlungen. Diese werden in der empirischen Analyse auf Homologien hin überprüft.

Das bedeutet aber auch, dass der Weg zum dokumentarischen oder ikonologischen Sinngehalt von korporierten Praktiken der abgebildeten BildproduzentInnen nicht *zwangsläufig* detaillierte Beschreibungen auf der *vor-ikonografischen* Ebene, derjenigen von *Gesten* oder *Kinemen*, voraussetzt. Denn beispielsweise kann der ‚Ehrgeiz‘ eines Schülers sich auch in der Art der interaktiven Bezugnahme auf der Ebene der *operativen Handlung* des ‚Sich-Meldens‘ zu dem Sich-Melden der anderen dokumentieren, indem hier nämlich etwa eine Konkurrenz zum Ausdruck kommt. Die *Validierung* und *Vertiefung* derartiger Interpretationen führt aber im Bereich der Bild- und Filminterpretation immer über die *vor-ikonografische* Ebene, welche somit die *primordiale*, vorrangige oder fundamentale Ebene darstellt. Die sozialwissenschaftlichen, aber auch die nicht-sozialwissenschaftlichen Disziplinen, welche auf Bild- und Filminterpretationen angewiesen sind, stehen dabei vor dem Problem, dass eine Beschreibungssprache auf der *vor-ikonografischen* Ebene weitgehend fehlt.

## *Die Gestaltungsleistungen der abbildenden BildproduzentInnen: Einstellung und Montage*

Die Gestaltung von Videos und Filmen ist stets in hohem Maße als Ergebnis der Gestaltungsleistungen der abbildenden BildproduzentInnen anzusehen. Dies gilt auch dann, wenn Videografien als Erhebungsinstrument eingesetzt werden – die Forschenden stehen hier kontinuierlich vor der Aufgabe, die eigene Perspektivität zu reflektieren und in bewusst gewählte Foki umzusetzen. Während im Bereich der *Audiografie* der Wechsel des Standorts des Beobachters, also des Mikrofons, allenfalls Veränderungen im Bereich der Tonqualität zur Folge hat, hat der Wechsel des Standortes und der Perspektive im Bereich der *Fotografie* und der *Videografie* – und dies ist ganz wesentlich – unausweichlich die Konstruktion neuer Realitäten und auch Simultaneitäten zur Folge. Eine offensive Auseinandersetzung mit der eigenen Perspektivität stellt sich aus Sicht der Praxeologischen Wissenssoziologie als methodische Kontrolle der Standortgebundenheit des Wissens der Forschenden dar. Als besonders geeigneten Analyseschritt, der der Standortgebundenheit der abbildenden BildproduzentInnen Rechnung trägt, erachten wir die Fotogrammanalyse (Fritzsche/Wagner-Willi 2013: 271). Denn das Fotogramm in seiner *Simultanstruktur* stellt nicht nur eine wesentliche Grundlage für die Rekonstruktion der Bewegungen und des Habitus der abgebildeten BildproduzentInnen dar, sondern auch der abbildenden.<sup>8</sup> Im Fotogramm dokumentiert sich *zum einen* die Art der Kameraeinstellung bzw. die *Einstellungsgröße*: die Detail-, Groß- und Naheinstellung, die halbnahe, amerikanische, halbtotale, totale und Weit- oder Panorama-Einstellung.

Neben der Einstellungsgröße erschließt sich auch der *Kamerastandort* auf dem Wege über das Fotogramm. In der filmwissenschaftlichen Literatur wird hier im Wesentlichen zwischen der Normal-, Auf- und Untersicht unterschieden (vgl. Hickethier 1996: 61f.). Mit dem Begriff des „Bildraums“ (a.a.O.: 71) wird in der filmwissenschaftlichen Literatur z.T. auf das Bezug genommen, was wir unter dem Begriff der „*Perspektivität*“ in der dokumentarischen Bildinterpretation herausarbeiten (vgl. Bohnsack 2009: Kap. 5.5.3 u. 7.3): die Rekonstruktion des perspektivischen Zentrums, des Fluchtpunkts und des per-

---

8 Die umfassende Bedeutung von Fotogrammen für die Filmproduktion und Filmregie wird in einer neueren Veröffentlichung besonders evident, welche die Funktion des Storyboards als eines Drehbuchs „in Skizzenform“ (Knigge 2012: 23) für die Arbeit bedeutender Regisseure – u.a. Alfred Hitchcock und Steven Spielberg – hervorhebt. Henkel (2012: 26) betont, dass „Hitchcock der Typus Regisseur war, der vor Drehbeginn die entsprechende Szene dank detaillierter Storyboards bereits exakt vor Augen hatte“. Der Abgleich von Storyboards mit einer entsprechenden Szene im Klassiker „Die Vögel“ beispielsweise macht deutlich: „Jede Mimik und Gestik der Hauptdarstellerin Tippi Hedren ist in den Zeichnungen ablesbar und wurde von ihr annähernd 1:1 ausgeführt“ (ebd.).

spektivischen Modus: Frontal- und Übereck- oder Schrägperspektive sowie Untersicht und Aufsicht (Luftperspektive).<sup>9</sup>

Aus dem Zusammenspiel der Einstellungsgröße, der Perspektivität und des Ausschnitts, der *Kadrierung*, ergibt sich dann auch die *planimetrische Komposition*, deren Rekonstruktion uns, da das Bild selbst eine Fläche ist, im Sinne von Max Imdahl (1994) unmittelbar zur Eigenlogik des Bildes führt. Die Rekonstruktion der planimetrischen Komposition besteht im Wesentlichen in der Rekonstruktion jener Linien, durch welche das Bild in seiner Flächigkeit zusammengehalten und dominiert wird. Diese Linienführung kann auch durch die Farbgebung hergestellt oder unterstützt werden. Ein erster innovativer Schritt zur Bedeutung der Farbgebung bzw. der Farbkontraste für die planimetrische Komposition und damit auch für die Montage in der dokumentarischen Videointerpretation findet sich im Beitrag von Stefan Hampl zur Analyse von Musikvideos in diesem Band.

Erst vor dem Hintergrund einer genauen Rekonstruktion der einzelnen *Fotogramme*, der stills, in deren jeweiliger Einstellung und Gesamtkomposition werden dann Veränderungen der Fotogramme überhaupt präzise identifizierbar. Letzteres ist aber Voraussetzung für eine präzise Rekonstruktion des *Einstellungswechsels* aufgrund der Kameraführung und ggf. der Montage oder – wie es in Bezug auf das Fernsehen heißt – der Bildmischung.

Dieser Bedeutung der Fotogramme trägt auch das von Stefan Hampl und Aglaja Przyborski für die dokumentarische Film- und Videointerpretation entwickelte *Transkriptionssystem MoViQ* („Movies and Videos in Qualitative Research“) Rechnung (siehe dazu den entsprechenden Beitrag von Stefan Hampl im vorliegenden Band).

## *Sequenzialität und Simultaneität erster und zweiter Ordnung*

Aus der methodologischen Reflexion der Differenz zwischen den Gestaltungsleistungen bzw. Bewegungen der *abgebildeten* BildproduzentInnen auf der einen und denen der *abbildenden* auf der anderen Seite ergeben sich zunächst zwei unterschiedliche Dimensionen der *Sequenzialität*, der Sequenzstruktur des Films: Abgesehen von der Sequenzabfolge der einzelnen Filmbilder während der Vorführung, welche die technische Grundlage des Mediums Film darstellt, haben wir zum einen die Sequenzialität der Bewegungsabfolge der abgebildeten BildproduzentInnen und aller abgebildeten Ereignisse. Diese können wir als *Sequenzialität erster Ordnung* bezeichnen. Zum anderen haben wir die durch *Einstellungswechsel* und *Montage* hergestellte Sequenzialität als

9 Die Selektivität im Bereich der Perspektive und des Ausschnitts (*Kadrierung*) kann auch aus Restriktionen des Raumes (bspw. eines Klassenraumes) und aus Auflagen der Autoritäten im Forschungsfeld resultieren, sodass daraus nicht so ohne Weiteres auf den Habitus der abbildenden BildproduzentInnen geschlossen werden kann (siehe dazu auch den Beitrag von Fritzsche/Wagner-Willi in diesem Band).

Gestaltungsleistung der abbildenden BildproduzentInnen, die wir als *Sequenzialität zweiter Ordnung* bezeichnen können, da sie die *Sequenzialität erster Ordnung* immer schon voraussetzt, indem sie auf der Grundlage ihrer Vorgaben operiert.

Analog können wir auch von einer *Simultaneität* sprechen, welche (wie oben bereits dargelegt) das Produkt der korporierten Praktiken und Gestaltungsleistungen der *abgebildeten* BildproduzentInnen ist, und sie von einer Simultaneität als Produkt der Gestaltungsleistungen der *abbildenden* BildproduzentInnen unterscheiden. Der Kunsthistoriker Max Imdahl (1996: 137), dessen Ausführungen wir folgen, wenn wir annehmen, dass ein Interpretationsverfahren, welches der Eigensinnigkeit der Bildhaftigkeit im Unterschied zum Text gerecht werden will, primär von dessen „*Simultanstruktur*“ auszugehen hat, versteht diese als ein „kompositionsbedingtes, selbst sinnstiftendes Zugleich“, bei der das „Ganze [...] von vornherein in Totalpräsenz gegeben ist“ (ebd.: 23).<sup>10</sup> Imdahl als Kunsthistoriker hat das Bild letztlich ausschließlich als Produkt der Gestaltungsleistungen der abbildenden BildproduzentInnen, der KünstlerInnen, betrachtet. Und es ist in seinem Sinne eine gesteigerte künstlerische Leistung, wenn es gelingt, einem in der Realität *sequenziell* sich vollziehenden Geschehen (bspw. der Gefangennahme Jesu im Meisterwerk von Giotto) im Rahmen der unhintergehbaren *Simultaneität* des Bildes Ausdruck zu verleihen und damit ganz neue Sinnqualitäten zu schaffen.

Aber auch jede ‚profane‘ Foto- und Videografie schafft durch die Wahl des Augenblicks und des Ausschnitts – ob sie nun will oder nicht – simultane Beziehungen zwischen AkteurInnen untereinander und zwischen den AkteurInnen und anderen Ereignissen und Gegenständlichkeiten, indem sie aus den durch die *abgebildeten* BildproduzentInnen hergestellten simultanen Beziehungen im Moment der Videografie *selegieren* muss. Beispielsweise werden in einer Unterrichtsvideografie durch eine Modifikation des Ausschnitts, also der Kadrierung, und des Moments der Aufnahme je spezifische simultane Beziehungen zwischen den Aktivitäten von SchülerInnen untereinander oder zwischen SchülerInnen und der Lehrperson in demselben Bild erfasst (und andere ausgeblendet).

Wenn es um die Simultaneität der Gestaltungsleistungen der *abgebildeten BildproduzentInnen* geht, so beziehen wir uns, wie oben dargelegt, grundlegend auf Ray Birdwhistell (1952: 15ff.). Um den unterschiedlichen Dimensionen der Simultaneität im Bereich der abgebildeten und der abbildenden

---

10 Vgl. dazu auch die Unterscheidung von diskursiven und visuellen Formen bei Susanne Langer (1984: 99): Visuelle Formen „bieten ihre Bestandteile nicht nacheinander, sondern gleichzeitig dar, weshalb die Beziehungen, die eine visuelle Struktur bestimmen, in einem Akt des Sehens erfasst werden“. Langer fasst unter „Sehen“ allerdings sowohl die *sinnliche Wahrnehmung* als auch die *Interpretation, die interpretative Sinnbildung*, also die *Semantik* des Visuellen, zusammen. Die Frage der Simultaneität im Bereich der sinnlichen Wahrnehmung ist als ein wahrnehmungspsychologisches Problem für unsere Betrachtung von sekundärer Bedeutung. Uns geht es um die Prozesse der interpretativen Sinnbildung, der Semantik.

BildproduzentInnen Rechnung zu tragen, können wir – analog zur Sequenzialität – auch hier eine *Simultaneität erster Ordnung* von der *abbildenden Simultaneität zweiter Ordnung* unterscheiden, da im Bereich von Foto- und Videografie letztere die erstere immer schon voraussetzt. Die Simultaneität erster Ordnung lässt sich auch als *Synchronizität* bezeichnen.

Während wir hinsichtlich der *Beschreibung* der Gestaltungsleistungen der abbildenden BildproduzentInnen auf die Beschreibungssprache der Filmwissenschaft zurückgreifen können, fehlt eine umfassende differenzierte Beschreibungssprache der Gestaltungsleistungen der abgebildeten BildproduzentInnen. Ebenso fehlt auch eine Beschreibungssprache der formalen Struktur der interaktiven Beziehungen der abgebildeten BildproduzentInnen, also der interaktiven Organisation ihrer körperlichen Ausdrucksformen. Im Kontext der dokumentarischen Gesprächsanalyse haben wir für den sprachlichen Austausch, die verbale Interaktionsorganisation der AkteurInnen, den Begriff der Diskursorganisation geprägt (ursprünglich: Bohnsack 1989; Przyborski 2004; Bohnsack/Przyborski 2006), in der elementare Modi der Sozialität ihren Ausdruck finden. Im Bereich der körperlichen Ausdrucksformen oder Bewegungen schlagen wir den Begriff der *Modi der Organisation korporierter Praktiken* vor, welcher dann mit dem Begriff der *Modi der Diskursorganisation* unter dem Oberbegriff der *Modi der Interaktionsorganisation* zusammengefasst werden kann.

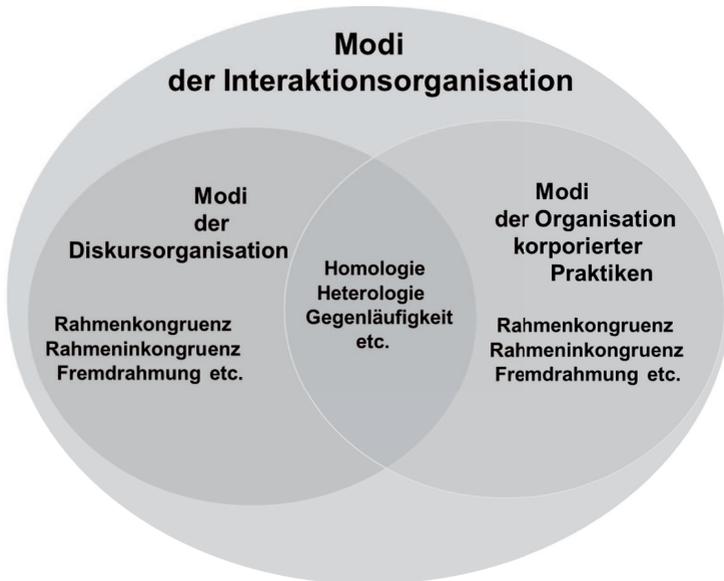


Abbildung 2: Modi der Interaktionsorganisation

## *Interaktionsorganisation: Diskursorganisation und Organisation korporierter Praktiken*

Insbesondere dort, wo sich die Äußerungen der AkteurInnen vor der Kamera wesentlich auf der nonverbalen Ebene und dort im interaktiven Bezug aufeinander entfalten, wie etwa im Bereich der Kleinkindpädagogik (siehe dazu den Beitrag von Iris Nentwig-Gesemann und Katharina Nicolai im vorliegenden Band), ist eine genauere Rekonstruktion und begriffliche Explikation der Formalstruktur der interaktiven Beziehungen, also der Modi der *Organisation korporierter Praktiken*, unerlässlich. Aber auch im Bereich der Unterrichtsforschung, in der die verbale Ebene im Zentrum steht, ist die Rekonstruktion der Modi korporierter Praktiken sinnvoll, wie die Beiträge von Martens/Petersen/Asbrand und Sabine Richter in diesem Band zeigen. Dieser Weg ist zwar schon früh von Monika Wagner-Willi ansatzweise beschritten worden (2005) – hinsichtlich der kategorialen Fassung, der begrifflichen Systematik, der Rekonstruktion von Modi der korporierten Praktiken und schließlich auch der übergreifenden Interaktionsorganisation stehen wir jedoch noch ganz am Anfang. In den genannten Beiträgen finden sich erste dahingehende Ansätze. So werden analog zur Rekonstruktion der Diskursorganisation in der Gesprächsanalyse auch in der Analyse korporierter Interaktionen sowohl unterschiedliche *Modi von Interaktionsbewegungen*, nämlich Propositionen, Elaborationen und Konklusionen, als auch unterschiedliche *Modi der Sozialität* unterschieden, und zwar inkludierende oder exkludierende Interaktionsmodi. Bei letzteren handelt es sich um oppositionelle oder divergente Interaktionsmodi. Zu den divergenten Interaktionsmodi zählen die Falsch- oder Fremdrahmungen. Letzteres bedeutet, dass A auf die Bewegungen von B systematisch in einer solchen Weise nonverbal reagiert, dass A dessen Bewegungen innerhalb seines eigenen Orientierungsrahmens vereinnahmt.

## *Die unterschiedlichen Systemebenen der dokumentarischen Videoanalyse*

Wenn wir die interaktiven Beziehungen der abgebildeten BildproduzentInnen auf der nonverbalen Ebene in den ihnen eigenen Ordnungs- und Organisationsprinzipien erfassen wollen, müssen wir sie ebenso als ein selbstreferentielles System, ein nonverbales Interaktionssystem, erschließen, wie wir dies auf der Ebene der sprachlichen Interaktion in Form von Texten gewohnt sind. Die Beiträge dieses Bandes unterscheiden sich dahingehend, auf welcher Ebene der Interpretation zwischen dem verbalen und dem nonverbalen oder inkorporierten Interaktionssystem differenziert wird – ob dies bereits auf der Ebene der formulierenden Interpretation oder erst auf derjenigen der reflektierenden Interpretation geschieht (siehe dazu auch die zusammenfassende Darstellung der

Beiträge). Dies ist sowohl von der Fragestellung als auch vom Datenmaterial abhängig. Eine – in diesem Sinne – radikale Variante geht dahin, für die beiden Interaktionssysteme eine je separate formulierende wie auch reflektierende Interpretation zu erstellen, also eine „Interpretation in der Bilddimension“ und eine „Interpretation in der Dimension von Text und Ton“, um diese dann am Ende in einer „reflektierenden Gesamtinterpretation“ zusammenzuführen (für letzteren Weg siehe Bohnsack 2009: Kap. 5.8 u. 6).

Die Bedeutung einer Differenzierung von nonverbaler und verbaler Dimension liegt u.a. darin, dass es dann möglich wird, methodisch kontrolliert zu überprüfen, ob beide Dimensionen einander wechselseitig zu validieren vermögen, und somit die Analyse auf ein höheres Niveau der Gültigkeit zu heben. So kann Margret Xyländer in ihrem Beitrag zur familialen Interaktion und Sozialisation in diesem Band herausarbeiten, dass die für einen bestimmten Familientypus zu identifizierende Doppelstruktur von großer Intimität oder Innerlichkeit auf der einen und einem Beobachtungs- und Überwachungscharakter auf der anderen Seite sowohl in der verbalen als auch in der nonverbalen Dimension, also homolog in Text und Bild, zu finden ist. In der Analyse des Verhaltens tobender Kinder auf dem Schulhof im Beitrag von Bohnsack/Lamprecht in diesem Band können die zunächst ausschließlich auf der Grundlage der Interpretation der korporierten Praktiken rekonstruierten Orientierungskomponenten durch die Interpretation der (nur sehr wenigen) sprachlichen Äußerungen validiert und metaphorisch verdichtet werden.

Weitgehend kann die Differenzierung der Analysen in die verbale und die nonverbale Dimension aber auch wesentliche Voraussetzung dafür sein, dass tiefer gehende und komplexere Strukturen herausgearbeitet werden können. Der Zugang zur Selbstreferentialität korporierter Praktiken ist insbesondere Voraussetzung dafür, Formen der Interaktionsorganisation zu erfassen, bei denen die korporierten Praktiken gegenläufig zur Ebene der Gesprächs- respektive Textanalyse operieren oder diese gar *konterkarieren* oder *unterlaufen*. Eine fundamentale Reziprozität innerhalb der interaktiven Beziehungen der abgebildeten AkteurInnen kann gerade darin ihren Ausdruck finden, dass die verbalen und die inkorporierten Praktiken durch ihre Gegenläufigkeit einander ergänzen. Letzteres wird im vorliegenden Band beispielsweise in dem Beitrag von Iris Nentwig-Gesemann und Katharina Nicolai in dem Bereich der Kleinkindpädagogik deutlich, und zwar in der Interaktion zwischen einem knapp dreijährigen Kind und der Erzieherin, welche dieses wickelt: Während sich das Kind auf der Ebene des nonverbalen, des korporierten Interaktionsstrangs in den Rahmen der Pädagogin fügt, sich diesem gegenüber responsiv oder reaktiv verhält, lässt sich letztere im verbalen Interaktionsstrang responsiv oder reaktiv auf den Rahmen des Kindes ein. Im Beitrag von Tanja Sturm aus der Unterrichtsforschung kann ferner herausgearbeitet werden, dass auf der verbalen Ebene die Beziehung zwischen der Lehrerin und den SchülerInnen asymmetrische Elemente aufweist, wohingegen sich auf der nonverbalen Ebene symmetrische Strukturen rekonstruieren lassen.

Die Relation der nonverbalen und der verbalen Systemebenen ist *eine* der zentralen Relationen der Videoanalyse. Sofern – wie im Falle des von Familienangehörigen produzierten Familienvideos – abgebildete und abbildende BildproduzentInnen identisch sind, ist die weitere interessierende Relation diejenige zwischen der Systemebene der *Simultaneität*, wie sie sich uns in den internen Relationen der Struktur der Fotogramme selbst dokumentiert, und jener anderen Systemebene, welche sich in den Relationen der Fotogramme zueinander, in deren *Sequenzialität*, dokumentiert. Bereits hier – in dieser Relationierung von Relationen – zeigt sich das, was der Filmtheoretiker Gilles Deleuze (1997: 24) meint, wenn er sagt: „Wäre das Ganze zu definieren, dann durch die Relation“.

Im Falle der Film- und Videoanalyse ist es also einerseits – im Bildbereich – die *interne Relationierung* der Bilder oder Fotogramme, also deren *simultane* Relationierung, welche wiederum in ihrer Relation zur Abfolge der Bilder in deren *sequenziellen Relationierung* zu rekonstruieren ist. Die Überprüfung der Relation von Simultanstruktur und Sequenzstruktur auf Homologien stellt ein wesentliches Element der Validierung im Bereich der Dokumentarischen Methode dar. Andererseits ist es die Relation dieses gesamten komplexen Bildbereiches zum Bereich des Verbalen, des Textlichen, welche das Ganze, also die Gesamtstruktur, von Filmen und Videos ausmacht. Die Relationierung ist zugleich Grundlage für weitere Differenzierungen wie auch für Validitätsprüfungen. Sobald wir in der Bilddimension zwischen der Dimension des Abgebildeten und derjenigen der Abbildung differenzieren, gilt es darüber hinaus, *zum einen* die Rekonstruktion der *Simultanstruktur* der abgebildeten Bewegungen mit der Rekonstruktion der *Sequenzstruktur* dieser Bewegungsabläufe zu relationieren, indem wir nach Homologien zwischen beiden Dimensionen suchen. Dabei ist – je mehr wir (im Falle der Videografie als Erhebungsinstrument) die eigene Erhebungsmethodik kontrollieren wollen – gleichsam ‚herauszurechnen‘, wie einerseits die Simultanstruktur erster Ordnung durch die Selektivität des Moments der Aufzeichnung und durch den gewählten Ausschnitt zur Simultanstruktur zweiter Ordnung wird. Andererseits gilt es zu reflektieren, ob und wie die Bewegungsabläufe in ihrer Sequenzialität erster Ordnung aufgrund der Kameraführung und eventuell der Montage (Bildmischung) durch das übergreifende System der Sequenzialität zweiter Ordnung überlagert und modifiziert werden.

## *Generelle Prinzipien der dokumentarischen Video- und Filmanalyse*

Die Dokumentarische Methode ist eine Methodologie und erkenntnistheoretische Position, die mit einer Grundlagentheorie (Metatheorie) verbunden ist, welche wir als Praxeologische Wissenssoziologie bezeichnen. Dieses theoretische wie auch methodologische Potential hat es ermöglicht, ein Spektrum

methodischer Zugänge im Bereich der Text- und der Bildinterpretation wie auch zu unterschiedlichen Gegenstandsbereichen zu eröffnen (im Überblick: [www.dokumentarischemethode.de](http://www.dokumentarischemethode.de)) und fortschreitend zu elaborieren. Indem die Dokumentarische Methode den übergreifenden Rahmen unterschiedlicher methodischer Zugänge bildet, ermöglicht sie auch sinnvolle Triangulationen, was insbesondere für die Video- und Filmanalyse von Bedeutung ist, stellt diese doch per se eine Triangulation von Gesprächs- und Bildanalyse dar.

Das methodologische Reflexionspotential der Dokumentarischen Methode bewahrt ebenso vor einer Fixierung auf Rezeptwissen. Als rekonstruktives Verfahren findet die Fortentwicklung der Methode in kontinuierlicher Rekonstruktion ihrer eigenen Forschungspraxis statt (dazu Bohnsack 2005 u. 2010b: Kap. 2). Gleichwohl liegen im Bereich der Gesprächsanalyse (resp. des Gruppendiskussionsverfahrens) und auch der Interview- und Bildanalyse aufgrund umfangreicher Forschungserfahrungen bewährte *Arbeitsschritte* in Erhebung und Auswertung bis hin zur Typenbildung vor, welche direkte Anleitungen für die Forschungspraxis darstellen. Aufgrund der Komplexität der Video- und Filmanalyse und der Vielfalt ihrer Gegenstandsbereiche haben sich hier unterschiedliche Verfahrensweisen ausdifferenziert, die in dem vorliegenden Band repräsentiert sind. Überwiegend sind ihnen folgende Prinzipien gemeinsam:

- der methodisch kontrollierte Übergang vom *expliziten* zum *impliziten Sinngehalt*, vom *Was* zum *Wie*, von der Unterstellung von *Motiven* zur Rekonstruktion des *modus operandi* der Handlungspraxis, des *Habitus*
- im Bereich der Bildinterpretation ist dieser Übergang derjenige von der Ikonografie zur Ikonologie
- der methodisch kontrollierte Übergang vollzieht sich auf dem Wege der Ausdifferenzierung der Arbeitsschritte der formulierenden und reflektierenden Interpretation des Textes wie des Bildes
- der Zugang zum Bild wie zum Text als *selbstreferentiellen Systemen* auf der Grundlage der Rekonstruktion der Formalstruktur von Texten und Bildern
- die Differenzierung zwischen den Bewegungen und Gestaltungsleistungen der *abgebildeten* und denjenigen der *abbildenden* BildproduzentInnen
- die Differenzierung der *Simultan-* und *Sequenzstruktur* in den Gestaltungsleistungen beider Kategorien von BildproduzentInnen
- der Zugang zur Formalstruktur der Interaktion der *abgebildeten* BildproduzentInnen: sowohl im Bereich der *verbalen* Interaktion, also der Gesprächsanalyse (Diskursorganisation), als auch im Bereich der *nonverbalen* oder *korporierten* Interaktion (Organisation korporierter Praktiken). In letzterer Hinsicht gilt es, neue Wege zu erschließen.
- der Zugang zur *Formalstruktur* der Gestaltungsleistungen der *abbildenden BildproduzentInnen* in ihrer Simultan- und Sequenzstruktur (Kameraeinstellung, Schnitt, Montage und Bildmischung). Hier kann an die filmwissenschaftlichen Kategorien angeknüpft werden.

- die systematische Auswahl der zu interpretierenden Passagen und Sequenzen nach dem Prinzip der *Fokussierung*
- das Grundprinzip der *komparativen Analyse*, welche Potentiale einer mehrdimensionalen Typenbildung und Generalisierung sowie der Kontrolle der eigenen Standortgebundenheit der InterpretInnen eröffnet
- die genannten Prinzipien führen zu einem *Wechsel der Analyseinstellung*. In einem solchen Wechsel ist die Unterscheidung einer *wissenschaftlichen* Video- und Filmanalyse von den Common-Sense-Interpretationen begründet.

### *Zusammenfassende Darstellung der Beiträge*

Die erste Rubrik des Bandes umfasst Interpretationen von Videografien, die zu Forschungszwecken erstellt wurden und die sich dem thematischen Komplex „Elementarbereich, Primarbereich und Familie“ zuordnen lassen. Eine Videointerpretation pädagogischer Interaktionen im Elementarbereich stellen *Iris Nentwig-Gesemann* und *Katharina Nicolai* mit ihrer Rekonstruktion typischer Muster der Interaktionsorganisationen im Krippenalltag vor. Zunächst verorten die beiden Autorinnen ihren Beitrag im Feld einer praxeologisch fundierten, rekonstruktiven frühpädagogischen Forschung. In Bezug auf die interessierenden Interaktionen zwischen Erzieherinnen und Kleinkindern geht es ihnen um die Bewertung von Beziehungs- und Interaktionsqualität, d.h., sie fragen danach, ob die Kinder als gleichwertige Interaktionspartner geachtet werden und ob es den Erwachsenen in der asymmetrischen pädagogischen Beziehung gelingt, auf ihre Handlungsmacht zu verzichten. Die Videografie einer Wickelsituation, die in diesem Beitrag analysiert wird, entstand im Zuge eines Praxisprojekts zu Alltagssituationen in der Krippe. Wie die Autorinnen darstellen, wählten sie zunächst auf der Grundlage eines inhaltlich thematischen Verlaufs mit der Gliederung in Ober- und Unterthemen und der Identifizierung von Fokussierungsakten Passagen für eine sequenzanalytische Feininterpretation aus. Diese wurden dann im Format eines detaillierten Beobachtungsprotokolls auf der vor-ikonografischen Ebene interpretiert. Anschließend wurden die Passagen einer reflektierenden Interpretation unterzogen, die eine ausführliche Analyse der Interaktionsorganisation, welche auch korporierte Handlungen umfasst, mit einschloss. Auf dieser Grundlage gelingt es den Autorinnen abschließend, inkludierende von exkludierenden Interaktionsmodi im frühpädagogischen Bereich zu unterscheiden und jeweils zu charakterisieren.

Mit dem Übergang vom Elementar- in den Primarbereich befasst sich der Beitrag von *Gerald Blaschke*, der eine Videointerpretation von sogenannten „Schnupperstunden“, d.h. den Unterrichtsbesuchen von Kindergartenkindern in ihrer (voraussichtlich) künftigen Schule, vornimmt. Er bezieht sich hierbei auf eine Studie, in deren Rahmen er eine Typologie der Praktiken der Gestal-