Erschienen in: Deppermann, Arnulf/Linke, Angelika (Hrsg.): Sprache intermedial. Stimme und Schrift, Bild und Ton. – Berlin/New York: de Gruyter, 2010. S. 359-382. (Jahrbuch des Instituts für Deutsche Sprache 2009), https://doi.org/10.1515/9783110223613.359

Werner Holly

Besprochene Bilder – bebildertes Sprechen

Audiovisuelle Transkriptivität in Nachrichtenfilmen und Polit-Talkshows

Abstract

Als eine wichtige Form der intermedialen Einbindung von Sprache wird technisch kombinierte („sekundäre") Audiovisualität beschrieben, wie sie prototypisch im Fernsehen vorkommt. Nach allgemeinen Strukturen von sekundärer Audiovisualität wird der Begriff der Transkriptivität (nach Jäger) kurz dargestellt: das „Anders-Lesbar-Machen" von Zeichen im gleichen oder einem anderen Zeichensystem. Danach werden zwei Spielarten von Fernsehaudiovisualität behandelt: Nachrichtenfilme als Zusammenspiel von Sprechertext mit vorgefertigten Bildsequenzen, nach bestimmten Mustern von wechselseitiger Transkription, die Anforderungen der Darstellbarkeit und Glaubwürdigkeit genügen sollen. In Polit-Talkshows werden die Sprecherbeiträge von Kamerainszenierungen mit drei Funktionen transkribiert: (a) Abwechslung und Gliederung, (b) Sprecherprofilierung und (c) Profilierung von Beteiligungsrollen anderer Teilnehmer.

1. Sprache pur? Sprache medial und intermedial

Das Thema dieses Bandes: „Sprache intermedial: Stimme und Schrift, Bild und Ton" gibt zu denken. Dass sich Sprachwissenschaft auch für Stimme und Schrift interessiert, ist noch nicht weiter verwunderlich, obgleich in der Geschichte der Disziplin bekanntlich über weite Strecken ausgeblendet war, dass Sprache nicht nur geschriebene Sprache ist, womit außer der gesprochenen Sprache auch gleich die Spezifik der geschriebenen Sprache aus dem Blick geraten war. Sprache wurde viel zu lange wie ein amediales Etwas behandelt. Immerhin hat die Differenz von Schriftlichkeit und Mündlichkeit inzwischen die Dignität eines Handbuchgegenstands erreicht. Aber wie können Bild und Ton (also auch nicht-stimmlicher Ton) zum Gegenstand von Sprachwissenschaft werden? Bleibt man nur bei den Bildern: Dies mag nach modischer (vielleicht sogar schon wieder altmodischer) Anbiederung an einen „iconic" oder „visual" oder „pictorial turn" klingen, an die großen Schlagworte vom „Zeitalter der Bilder", von der „Visualisierung der Kommunikation" durch neue und inzwischen längst mittelalte Medien, an den Bild-Boom in allen möglichen Wissenschaften. Und dann: Sollten und können Linguisten sich überhaupt mit Bildern beschäftigen? Ist das nicht das Feld anderer?

Bevor ich mich in derlei Bedenken verliere, gebe ich kleine Beispiele, die andeuten sollen, dass Medialität und Intermedialität von Sprache jenseits aller Moden nicht nur linguistisch behandelt werden sollten, sondern auch behandelt werden müssen. Ich zitiere exemplarisch drei historische Aussprüche:

(1) „Platz an der Sonne"
(2) „Seit 5 Uhr 45 wird jetzt zurückgeschossen"
(3) „Ich bin ein Berliner"

Sie stehen verdichtend für historische Ereignisse, als Symbole für deutsche Geschichte des vorigen Jahrhunderts: für deutsches Weltmachtstreben, den deutschen Angriff auf Polen und die Situation im geteilten Berlin. Diese drei Phraseme und ihre Verankerung im kulturellen Gedächtnis zeigen zugleich, wie sich die mediale Situation im Laufe dieses Jahrhunderts verändert hat: (1) können wir bestenfalls zurückverfolgen bis zur Verschriftlichung der Rede des Staatssekretärs von Bülow, die er in der Reichstagssitzung vom 6.12.1897 gehalten hat; (2) haben wir als so genannten O-Ton im Ohr, in Wirklichkeit nicht – wie der Ausdruck *O-Ton* nahe legt – das Original, sondern eine Tonaufzeichnung von Hitlers Reichstagsrede vom 1.9.1939, die natürlich die Spur ihrer Technik enthält; (3) schließlich ist als Filmausschnitt aus Wochenschauen geläufig, aber die meisten Zeitgenossen haben die Rede, aus der das Zitat stammt, als Fernsehübertragung wahrgenommen. Auch die Nachgeborenen kennen nicht nur den Wortlaut, nicht nur die Stimme Kennedys mit seiner typischen Sprechweise, man sieht in den sattsam bekannten Filmdokumenten bis heute, wie auch sein sichtbares Auftreten zu seinem (seit Obama wieder vielbeschworenen) Charisma beiträgt: Er hatte nicht nur beste Redenschreiber, sondern war eben auch ein glänzender Rhetoriker, der wusste, wie gelungene *actio* bewerkstelligt wird. Die Bilder (siehe Abb. 1-4) dieses „medialen Geschichtsklischees"[1] zeigen Ernst in staatsmännischer Pose, dann Lächeln, Winken, verlegenen Buben-Charme mit Krawattenesteln; die Bildmontage gibt übrigens immer mal auch Naheinstellungen von einzelnen Zuschauern und ihrer Begeisterung.

Schon allein deshalb müssen wir Sprache intermedial betrachten. Wenn wir sprachliche Diskursereignisse angemessen beschreiben wollen, müssen wir ihre Medialität einbeziehen, sei es Schrift, sei es Stimme, aber eben nicht nur Stimme, sondern auch deren Überformung durch technische Medien, und vor allem ihre Einbettung in und Verknüpfung mit allen möglichen Arten von Sichtbarem und Hörbarem, den Körper im Raum – überhaupt Visuelles, und falls dabei technische Medien im Spiel sind: was und wie sie dies zeigen und hören lassen. Alles andere ist zwar Sprache pur, eine zu wissenschaftlichen Zwecken manchmal nützliche und deshalb erlaubte Reduktion, aber eben eine Reduktion, deren wir uns bewusst bleiben müssen und die wir nicht prinzipiell und „ohne Not" vornehmen dürfen.

[1] Siehe zu diesem Phänomen des kulturellen Gedächtnisses Holly (2003).

Abb. 1–4: Kennedy vor dem Rathaus Schöneberg

2. Audiovisualität

Im Folgenden werde ich eine Variante der intermedialen Einbettung und Verknüpfung von Sprache behandeln, die man als ‚sekundäre Audiovisualität' charakterisieren kann. Audiovisualität (als eine Form der Intermedialität oder Multimodalität) heißt zunächst nur, dass in den entsprechenden Kommunikationsformen beide *Modes*, Akustisches und Optisches, genutzt werden können. ‚Primär audiovisuell' sind alle *face-to-face*-Formen, auch die nicht-konversationellen Meso-Formen, zu denen man auch Vorträge zählen kann, in denen ja, wie gerade die antike Rhetorik wusste, die *actio*, also die gesamte körperliche, auch visuell wahrgenommene Performanz nicht wenig beiträgt. Wichtiger wird das Visuelle noch, wenn zusätzliche optische Stützen in Form von Tafelanschrieben, Folien usw. dazukommen. Visuell und auditiv operiert auch das räumliche kommunikative Arrangement, das dann in meinem zweiten Untersuchungsfeld, den Talkshows, eine erhebliche Rolle spielt.

Bei räumlicher Entkopplung der Beteiligten kann man dann von ‚sekundärer' Audiovisualität sprechen; hier besorgen Kameras Bilder und Mikrofone Töne; vor allem ist die Kombination beider Modes hier nicht mehr nur das spontane Produkt körpernaher Semiose (Sprachlaute, Mimik, Gestik, Kinesik), sondern es muss technisch bewerkstelligt werden, wie beide zusammenkommen. Die Kamera kann nun, wie meine Augen bei der Rezeption auch, Sprecher zeigen oder auch anderes. Während meine Augen aber nur über ein sehr begrenztes Potenzial in der Lieferung von „Ansichten" verfügen, kann eine Bildmontage alles Mögliche zu einem Sprachtext kombinieren, nicht zuletzt können auch andere Töne dazukommen.

Zeichenarten	<wahrnehmungsnah>	<arbiträr>
<körpernah>, <temporär>	*Gestik, Mimik*	*Lautsprache*
<körperunabhängig>, <fixiert>	*Bild, Film*	*Schriftsprache, abstrakte Symbole*

Abb. 5: Zeichenarten (nach Sachs-Hombach 2003, S. 96)

Das kleine Schema (Abb. 5) veranschaulicht auch die semiotische Leistungsfähigkeit der Tonfilmkombination, die Vorzüge arbiträr körpernaher mit denen körperunabhängiger und wahrnehmungsnaher Zeichen verbindet; zugleich sind beide, Lautsprache und Film, linear in einem Bewegungs- und Zeitablauf (der Film erzeugt zumindest die Illusion von Bewegtheit) und können so eine gemeinsame narrative oder berichtende Dynamik entwickeln, weil sie nicht nur additiv operieren. „Filme sind" – so Angela Keppler (2006, S. 73) – „ein virtueller Bewegungsraum, der uns nicht allein in Bewegungen des Sehens, sondern auch des Hörens und mit ihnen in eine komplexe Bewegung des Verstehens versetzt."

Ich werde hier beide Typen von sekundärer Audiovisualität behandeln, den, der überwiegend Sprecher mit ihrer Gestik und Mimik zeigt (dafür behandle ich das Genre ‚Polit-Talkshows'), und den, der meist anderes zeigt (in so genannten ‚Nachrichtenfilmen'). Letztere liefern Bilder, die besprochen werden, „besprochene Bilder"; erstere zeigen meist – oder besser ‚bebildern' – Sprecher, „bebildertes Sprechen". Wie Bilder und Sprache zusammenkommen, in meinem Fall, wie sie in den beiden Typen, die ich unterschieden habe, zusammengebracht werden, so dass sie zusammen ein Bedeutungspotenzial generieren, soll der Gegenstand dieses Beitrags sein. Ich werde also nichts über die Rezeption dieser Genres sagen, sondern beschränke mich auf das, was man in der Medienwissenschaft eine „Produktanalyse" nennt, die gewissermaßen präfigurierte Möglichkeiten der Rezeption herausarbeitet.

Zuvor will ich auf einige Literatur verweisen, die sich mit dem audiovisuellen Sprach-Bild-Zusammenhang beschäftigt hat. Erstaunlich ist, dass sich die Film- und Fernsehwissenschaft kaum für die Frage zu interessieren scheint, wie Wort und bewegte Bilder zusammengehen. Dies ist einer der Gründe, warum hier ein Feld für Linguisten ist, andere scheinen sich an Sprache nicht so recht heranzutrauen. Bei den wenigen Arbeiten, die wirklich auch bewegte Bilder und Sprache behandeln (die meisten analysieren statische Bilder und Schrift), geht es verständlicherweise meist um Spiel-

filme (z.B. Rauh 1987) oder um Nachrichtenformate im Fernsehen; die Nicht-Linguisten (wie etwa Brosius 1998) schreiben zwar über Bildfunktionen, das Sprachverstehen wird aber nur sehr oberflächlich thematisiert, noch pauschaler der Zusammenhang zwischen beiden; für Diskussions- oder Talksendungen beschränken sich die meisten (auch Holly/Kühn/Püschel 1986) auf pauschale Hinweise auf die Bildkomponente in angefügten Kapiteln (so auch die Politikwissenschaftler Meyer/Ontrup/Schicha 2000).

Als gewinnbringend hervorheben möchte ich hier zum einen die Tradition der britischen „social semiotics", die vor allem mit den Namen Gunther Kress und Theo van Leeuwen verbunden ist und auf die Hallidaysche Funktionalgrammatik zurückgreift, und zum andern die Arbeit von Harald Burger. Beide erfassen die Sprach-Bild-Beziehungen in Termini der formalen und semantischen Passung und der pragmatisch-funktionalen Bezugnahme. Dabei greift van Leeuwen auch auf die beiden Typen zurück, die schon Roland Barthes ([1964] 1990, S. 34) als „Funktionen" der „sprachlichen Botschaft in bezug auf die […] bildliche Botschaft" unterschieden hat: ‚Verankerung', durch die aus einem vagen Bildinhalt etwas ausgewählt und gewissermaßen festgenagelt wird, und ‚Relaisfunktion', die etwas Eigenständiges hinzufügt. Einen zusammenfassenden Überblick über „visual-verbal linking" gibt van Leeuwen (2005, S. 230) im folgenden Schema (Abb. 6):

Image–text relations		
Elaboration	Specification	The image makes the text more specific (illustration)
		The text makes the image more specific (anchorage)
	Explanation	The text paraphrases the image (or vice versa)
Extension	Similarity	The content of the text is similar to that of the image
	Contrast	The content of the text contrasts with that of the image
	Complement	The content of the image adds further information to that of the text, and vice versa ('relay')

Abb. 6: Überblick über „visual-verbal linking" (aus: van Leeuwen 2005, S. 230)

Man kann sehen, wie die Hallidayschen Begriffe ‚Elaboration', unterteilt in ‚Specification' und ‚Explanation', und ‚Extension' auf die Sprach-Bild-Beziehung übertragen werden. Zugleich tauchen altbekannte Begriffe für semantische Relationen wieder auf: Similarity, Contrast, Complement.

Burgers relativ reichhaltige Darstellung der Sprach-Bild-Relationen (2005, S. 400–424), die vor allem auf Fernsehen gemünzt sind, zieht drei Dimensionen heran: er gliedert sie 1. formal, 2. semantisch, 3. pragmatisch-funktional. Unter dem ersten Gesichtspunkt sind (nach Rauh 1987, S. 89 ff.) die Differenzen von synchron/asynchron, syntop/asyntop und intradiegetisch/

extradiegetisch, die zur Unterscheidung von On-, Off- und Oversituationen dient, relevant; außerdem die simultan/asimultan-Differenz. Für semantische Beziehungen unterscheidet er auf einer Skala von „konvergent – divergent" einmal ‚redundante' Beziehungen, dann ‚komplementäre' und schließlich ‚rhetorische' (Metonymie, Metapher). Funktional-pragmatisch unterscheidet er von „Text zu Bild" zwei: ‚Monosemieren' (mit und ohne explizite Deixis) und ‚Metakommunikative Kommentierung'. Für „Bild zu Text" gibt es fünf Funktionen: Referenzsicherung, Veranschaulichung, Authentizität, Aktualität, Weckung von Interesse. Die letzte Reihe knüpft an die zahlreichen Bildfunktionskataloge an, die vor allem seit Huth (1985) für informative Fernsehsendungen und auch in andern Zusammenhängen entwickelt worden sind.

Natürlich sind solche Kategorienraster, so nützlich und unumgänglich sie sein mögen, immer nur grobe Anhaltspunkte; sie können eine differenziertere Analyse anregen, aber nicht ersetzen. Vor allem wären weitere Erklärungen wünschenswert, die auf die spezifische Semantik der einzelnen Zeichenarten eingehen und durchsichtig machen, in welchen Fällen welche Relation hergestellt wird, kurz: die Frage geht nach typischen Mustern der Sprach-Bild-Kombination, die sich aus der autochthonen Semantik der beiden ableiten lassen. Im Weiteren will ich zunächst nach solchen Mustern für die Nachrichtenfilmanalyse suchen, dazu aber die Analyse auch theoretisch unterfüttern und einen zweiten, ganz anders ansetzenden Vorschlag für die Kamerainszenierung in Talkshows exemplarisch-ausschnitthaft präsentieren.

3. Transkriptivität

Zunächst und möglichst kurz zum theoretischen Unterbau. Fragt man nach der grundsätzlichen Struktur von Beziehungen zwischen Zeichen, kommt man in der Sprache zu syntaktischen Mustern, die in der Regel durch Formelemente kodiert sind. Solche sind zwischen Wörtern und Bildern in Tonfilmen selten zu finden, regelmäßig allenfalls in einem sehr rudimentären pragmatischen Modus, der durch Nebeneinander-Positionierung oder Gleichzeitigkeit der Wahrnehmbarkeit indiziert ist. Was gleichzeitig oder benachbart sprachlich und bildlich wahrnehmbar gemacht wird, wird wohl schon etwas miteinander zu tun haben, so darf nach der allgemeinen Griceschen Kooperationsmaxime vermutet werden. Spezifischere Bande werden dagegen semantische Bezüge liefern, die in unserem Wissen verankert sind, dadurch dass wir sprachliche und bildliche Informationen aufeinander beziehen können. In der Sprache sind natürlich auch explizite bilddeiktische Ausdrücke möglich. Forscher mit einer Neigung zu Eyetracking-Experimenten können die Bezugnahmen in Print- oder Online-

Textflächen aus dem Hin- und Herspringen der Augenbewegungen von Wort zu Bild erschließen, allerdings kaum beim Betrachten und Hören von Tonfilmen, wo das Auge zwar beobachtbar von Wort zu Bildinhalt geführt werden kann, aber sicher nicht umgekehrt.

Ludwig Jäger (2002, 2004, in diesem Band) hat die Idee entwickelt, dass die Generierung von Bedeutungen grundsätzlich durch die ‚transkriptive' Bezugnahme von Zeichen auf Zeichen zu erklären ist, nicht in erster Linie durch den Bezug von Zeichen auf Sachen, zu denen wir ja keinen unmittelbaren Zugang haben. Das impliziert zwar ein Problem des Anfangs, aber Jäger löst es durch eine „metaleptische" Figur: Auch aus einem noch bedeutungslosen Ausdrucksmaterial, einem ‚Präskript', wird durch ein Verfahren der Bezugnahme, das er gewollt metaphorisch „Transkription" nennt, nachträglich etwas Lesbares oder besser Verstehbares, ein ‚Skript', allerdings nicht in beliebiger Weise; es gibt Angemessenheitskriterien, die geltend gemacht werden können. Die Ontogenese der Sprach- oder der Malkompetenz liefert hierfür schöne Belege. Überhaupt kann man nun sagen, dass Transkription dazu dient, etwas „besser lesbar, anders lesbar, manchmal gerade weniger lesbar" zu machen und man kann einen intramedialen Bezug zwischen Zeichen derselben Art von einem intermedialen zwischen verschiedenen Zeichenarten, z.B. zwischen Sprach- und Filmbildzeichen, unterscheiden. Bilder machen zugehörige Sprachzeichen anders lesbar und umgekehrt. Hier noch ein Hinweis: Wenn man nun verkürzend davon spricht, dass Sprache Bilder und umgekehrt „transkribieren", dann ist das kein Ausweis von vorpragmatischem Denken, sondern die simple Gewohnheit, pingelige Redeweisen wie „mit Sprache transkribiert ein Sprecher oder Hörer ein Bild" etwas zu vereinfachen. Die Pointe der Transkriptionsidee ist ja gerade, dass man nicht an vorgängige Inhalte glaubt, die von Ausdrucksseiten bloß repräsentiert werden. Ohne Zeichenbenutzer bedeuten Ausdrücke nichts, wie denn auch?

Man kann aber fragen, was die Transkriptivitätsidee für die Analyse intermedialer Texte leistet. In meinen Augen wäre es schon genug, wenn sie hinreichend deutlich machte, wie die Generierung von Sprach-Bild-Bedeutungskomplexen im Rahmen eines allgemeinen Modells der Semiose anzusiedeln ist. Sie modelliert vor allem aber, dass solche Bedeutungskomplexe nicht in der bloßen Addition von Sprach- und Bildbedeutungen entstehen, sondern dass sie das Ergebnis höchst dynamischer Prozesse sind, in denen sie sich durch wechselseitige Bezugnahmen aufbauen, immer natürlich unter Einbeziehung aller Arten von Kontext- und Schemawissen, das auf beiden Seiten angestoßen werden kann.

Eine wirklich problematische Frage für eine konsequent handlungstheoretische Fundierung sehe ich eher darin, dass die durch eine nicht völlig kontrollierbare Kombination entstehenden Bedeutungspotenziale keineswegs alle intendiert, also gemeint sind, dass also so etwas wie ein semio-

tischer Überschuss entsteht, der von Rezipienten gedeutet werden kann, obwohl es sich nicht um Ergebnisse von kommunikativen Handlungen dreht, eher um Zufallsprodukte, die zum Ausdruck kommen, ohne zum Ausdruck gebracht zu werden, Phänomene der dritten Art vielleicht.[2]

Ich will es jetzt bei diesen kurzen Bemerkungen zur Transkriptivität belassen und zu meinen konkreten Analysefeldern weitergehen. Zuvor die These, die ich dabei verfolge:

Fernsehaudiovisualität kann analysiert werden als wechselseitige dynamische intermediale Transkriptivität, performiert nach Mustern einer Logik, die vor allem begründet ist in den spezifischen Potenzialen und Defiziten der jeweiligen Semantiken von gesprochener (teilweise auch geschriebener) Sprache und (überwiegend) bewegten Bildern, nach Bedürfnissen bestimmter Genres, wobei auch allgemeines und kulturelles Kontext- und Schemawissen herangezogen wird.

4. Sprach-Bild-Transkriptionen I: Besprochene Bilder – Nachrichtenfilme

Das Beispiel, das ich heranziehen werde, geht auf Material der Agentur APTN vom Juli 2005 zurück, das in drei verschiedenen Verarbeitungen durch verschiedene Sender dokumentiert ist. Ich beschränke mich hier aus Gründen der gebotenen Kürze (siehe zu weiteren Analysen des Materials auch Holly 2008 und 2009) auf die ZDF-Version, die mit 25 sec die kürzeste ist. Sie zeigt 6 Einstellungen, zu denen wir 4 Spracheinheiten hören. Das ZDF zeigt nur einen kleinen Ausschnitt des angebotenen Filmmaterials, wie man einer so genannten „Shotlist" der Agentur entnehmen kann, die sie zusammen mit einer „Storyline" möglichen Interessenten anbietet, beides übrigens schon Sprach-Bild-Transkriptionen. Dabei wird im ZDF die erste Einstellung des Ausgangsmaterials an die vorletzte Position montiert, ein entscheidendes Detail der letzten Einstellung ist übrigens weggelassen.

APTN-Shotlist: 1.7.2005, Turkey Shooting:

(1) Exterior of Justice Ministry building, ambulance outside
(2) Suspect runs out of building and through garden
(3) Police chasing and firing at suspect (AUDIO: gunfire)
(4) Tracking shot, police chasing suspect
(5) Suspect lying on ground surrounded by police
(6) Suspect sitting up

[2] Ausführlicher dazu Holly (in Vorbereitung).

(7) Police shooting suspect
(8) Bomb disposal expert in protective clothing, pan to suspect moving head
(9) Security asking camera to move away
(10) Police forming a cordon in front of suspect
(11) Bomb disposal expert moving in to suspect
(12) Bomb disposal expert cutting clothing off suspect …
(13) Second bomb disposal expert being sent away
(14) Cutaway to police
(15) Bomb disposal expert cutting clothing off suspect, examining object found on suspect, suspect moving head
(16) Cutaway of police and people gathered at scene
(17) Bomb disposal expert with tan cylinder found on suspect
(18) Police arguing with cameraman
(19) Officials at scene
(20) Bomb disposal expert
(21) Officials outside Ministry building
(22) Wide of officials outside building
(23) Bomb disposal expert
(24) Scene of crime officer putting on protective suit
(25) Pan from building to police standing in line

ZDF-Auswahl:

(2) Suspect runs out of building and through garden
(3) Police chasing and firing at suspect (AUDIO: gunfire)
(10) Police forming a cordon in front of suspect
(12) Bomb disposal expert cutting clothing off suspect …
(1) Exterior of Justice Ministry building, ambulance outside
(15) Bomb disposal expert cutting clothing off suspect, examining object found on suspect, (suspect moving head)

Nun werde ich dieses Filmchen, das in einem Nachrichtenblock gesendet wurde, anhand von Standbildern der einzelnen Einstellungen (E 2, 3, 10, 12, 1, 15) veranschaulichen, dann den Sprechertext, den man in vier Einheiten (I–IV) gliedern kann, wiedergeben und anschließend anhand einer Tabelle, die beide kombiniert, auf die transkriptiven Muster eingehen.

E (2)

E (3)

E (10)

E (12)

E (1)

E (15)

Abb. 7–12: Einstellungen des ZDF-Nachrichtenfilms

Sprechertext:

(I) *in der türkischen hauptstadt Ankara hat die polizei offenbar einen bombenanschlag auf das justizministerium verhindert* [2]

(II) *die beamten erschossen den mutmaßlichen attentäter nach einer kurzen verfolgungsjagd auf offener straße* [3, 10]

(III) *er soll zuvor versucht haben in das gebäude einzudringen und dort einen sprengsatz zu zünden* [12, 1]

(IV) *angeblich war er mitglied einer linksextremistischen terrorgruppe* [15]

Die Übersichtstabelle (Abb. 13) zeigt zum einen, dass nominierende/referierende Bezüge vom Sprachtext zum Bildtext gehen, die Referenzobjekte im Bild identifizieren, und zwar in einer vorausweisenden/kataphorischen Richtung (wie bei *polizei* oder *regierungsgebäude*) und in einer rückverweisenden/anaphorischen Richtung (wie bei *attentäter* und *verfolgungsjagd*). Das nominalisierende Kompositum *verfolgungsjagd* bezieht sich allerdings nicht nur auf ein einzelnes Referenzobjekt, sondern auf eine Proposition. Insofern liegt hier eine etwas andere Beziehung vor. In beiden Fällen gehen die Bezüge nach dem Muster ‚Mit Worten sehen', eigentlich merkwürdig, wenn man an die berühmte Formel vom Bild, das mehr als tausend Worte sagt, denkt.

Zugleich transkribieren die Bilder in jeweils umgekehrter Richtung die sprachlichen Ausdrücke zum Zwecke der Authentisierung, sie vermitteln uns den Eindruck von Augenzeugenschaft, und sie plausibilisieren das Gesagte durch Evidenz, indem sie einen simplen, aber anschaulichen Miniplot

unterlegen; hier so etwas wie: Mann wird von Polizei verfolgt und erschossen. (Im Übrigen können die Bildbeschreibungen der Shotlist als „ethnographische Daten" genommen werden, die es dem beschreibenden Wissenschaftler ersparen, eigene Bildbeschreibungen spekulativ auf den Sprachtext hin zu konstruieren. Ich habe hier lediglich zusätzliches Wissen getilgt, mit dem schon die Shotlist-Beschreibungen die Bilder anreichern.)

Sprachtext	**Bildtext**
(I) *in der türkischen hauptstadt Ankara hat die polizei offenbar einen bombenanschlag auf das justizministerium*	(2) Mann läuft aus Gebäude durch Parkstraße
verhindert (II) *die beamten erschossen den mutmaßlichen attentäter nach einer*	(3) Polizei verfolgt ihn und schießt
kurzen verfolgungsjagd auf offener	(10) Kordon von Polizisten
straße (III) *er soll zuvor versucht haben in das regierungsgebäude*	(12) Mann in Schutzanzug beugt sich über Liegenden
einzudringen und dort einen sprengsatz zu zünden	(1) Außenansicht von Gebäude
(IV) *angeblich war er mitglied einer linksextremistischen terrorgruppe*	(15) Mann in Schutzanzug beugt sich über Liegenden

Abb. 13: Transkriptionsbeziehungen I Sprachtext – Bildtext

Sprachtext	**Bildtext**
(I) *in der türkischen hauptstadt Ankara hat die polizei offenbar einen bombenanschlag auf das justizministerium*	(2) Mann läuft aus Gebäude durch Parkstraße
verhindert (II) *die beamten erschossen den mutmaßlichen attentäter nach einer*	(3) Polizei verfolgt ihn und schießt
kurzen verfolgungsjagd auf offener	(10) Kordon von Polizisten
straße (III) *er soll zuvor versucht haben in das regierungsgebäude*	(12) Mann in Schutzanzug beugt sich über Liegenden
einzudringen und dort einen sprengsatz zu zünden	(1) Außenansicht von Gebäude
(IV) *angeblich war er mitglied einer linksextremistischen terrorgruppe*	(15) Mann in Schutzanzug beugt sich über Liegenden

Abb. 14: Transkriptionsbeziehungen II Sprachtext – Bildtext

Darüber hinaus (siehe Abb. 14) verknüpft die überlappende Montage (wir sehen (3) bzw. (12) schon, während (I) und (II) noch zu hören sind), indem die Bilder an die gerade zu Ende gehenden Propositionen semantisch andocken und sie entsprechend transkribieren: (3) zeigt, wie (I) gemacht wurde (Detaillierung), (12) zeigt, wie das ‚Ergebnis' oder die ‚Folge' von (II) aussieht. Solche semantischen Relationierungen verdichten auf der Basis von allgemeinem Welt- und Schemawissen das Verstehen des Gesamttexts.

Mit diesen sehr elementaren Mustern der Bildsemantisierung und Bildaufschließung einerseits, der Sprachauthentisierung und Sprachveranschaulichung andererseits ist aber nur die Oberfläche der wechselseitigen Sprach-Bild-Transkriptionen erfasst. Natürlich gibt es aufgrund weiterer Leistungen von Sprache und Bild weitere, subtilere Beziehungen, die ich hier nur andeuten kann. Schaut man sich den Sprachtext genauer an, stellt man fest, dass er relativ autark formuliert ist. Er weist mit seiner Text-Struktur (Leadsatz, Ereignisverlauf, Vorgeschichte, Hintergrund) die typischen Züge einer Pressemeldung auf und könnte ebenso gut ohne Bildunterstützung verstanden werden.

Leadsatz	(I) *in der türkischen hauptstadt Ankara hat die polizei offenbar einen bombenanschlag auf das justizministerium verhindert*
Ereignisverlauf	(II) *die beamten erschossen den mutmaßlichen attentäter nach einer kurzen verfolgungsjagd auf offener straße*
Vorgeschichte	(III) *er soll zuvor versucht haben in das gebäude einzudringen und dort einen sprengsatz zu zünden*
Hintergrund	(IV) *angeblich war er mitglied einer linksextremistischen terrorgruppe*

Dies ist nicht bei allen solchen unterlegten Sprechertexten so. Manchmal gibt es mehr oder weniger explizite Verweise auf die Bildkomponente, so auch in einer CNN-Version desselben Materials, wo die Moderatorin aus dem Off die laufenden Bilder kommentiert und auf den chaotischen Charakter der bildlich vermittelten Szene verweist: *as you can see there is some of the chaos caught on tape.* In anderen Fällen sind die Hinweise weniger explizit, unterscheiden sich aber schon stark von dem Pressemeldungsstil, der ohne Bildtranskription möglich ist.

Zu diesem Stil gehören auch die nachrichtentypischen Heckenausdrücke, die das Gesagte relativierend immunisieren sollen. In jeder Aussage ist die epistemische Sprechereinstellung explizit abgeschwächt, durch zwei Modaladverbien (I *offenbar*, IV *angeblich*), ein vom Sprachhandlungsverb *mutmaßen* abgeleitetes Adjektiv (II) und inferentiellen Gebrauch des Modalverbs *sollen* (III). Zugleich werden handfeste sprachliche Bewertungen vorgenommen, durch deontische Wörter wie *bombenanschlag* (I), *attentäter* (II), *sprengsatz* (III), *linksextremistisch*, *terrorgruppe* (IV), die allesamt moralische und

sehr starke Negativ-Beurteilungen des Erschossenen zum Ausdruck bringen oder implizieren. Dagegen kann man der Formulierung *erschossen … auf offener straße* (II) eine gewisse Skandalisierung des polizeilichen Vorgehens entnehmen.

Die unterschiedlich deutlichen und unterschiedlich gradierten Bewertungen des Erschossenen und der Polizei können auch als ausgleichende Kommentare zu den Bildern gelesen werden, die umgekehrt zwar das gewaltsame Vorgehen der Polizei zeigen, nicht aber das des mutmaßlichen Attentäters, der sichtbar nur als Opfer in Erscheinung tritt, nicht als Täter. Hier liefert die Sprache in der ‚Relaisfunktion' einen eigenen Beitrag zum intermedialen Textganzen. Insofern transkribieren die wahrheitswertabschwächenden und abgestuft evaluativen Sprachelemente die Bilder in Richtung einer Relativierung und einer entdramatisierenden Tendenz, zu der auch die routinierte Prosodie der Sprecherinnenstimme beiträgt, so dass letztlich der gesamte Vorfall, vor allem das Vorgehen der Polizei zwar zunächst kritisiert, dann aber doch durch die dagegen stehenden Bewertungen des Erschossenen „normalisiert", d.h. als irgendwie balanciert eingeordnet werden.

Dies scheint auch erforderlich, denn die Bilder sind auf Dramatisierung und Sensationsverstärkung hin angelegt. Nicht ohne Grund versucht schon das Begleitmaterial der Agentur in einer Mischung aus Warnung und Kaufanreiz die Sensationsträchtigkeit der Bilder zu markieren, mit dem Hinweis auf der Einleitungsmaske: „+++CLIENTS PLEASE NOTE THAT PACKAGE CONTAINS FOOTAGE OF SUSPECT BEING SHOT+++". Auch die Qualität der Einstellungen und des O-Tons kann in diese Richtung wirken.

So geben die Einstellungen (2) und (3) (siehe oben) durch die Bewegung einen unmittelbaren Eindruck vom aktuellen Ablauf der Szene, die Kamera fängt mit wackelnden Bildern direkt ein, was als Ereignis berichtenswert ist, den Fluchtversuch des Protagonisten und die Verfolgung durch die Polizei, (3) enthält dazu noch Tonaufzeichnungen der Schüsse, was die Shotlist als „(AUDIO: gunfire)" und damit wiederum als dramatisches Element notiert. In der ZDF-Version ist zwar nicht zu sehen, wie der Mann niedergeschossen wird, aber die „Ergebnisbilder" (12) und (15) (siehe oben) zeugen durch zwei Elemente von der Dramatik der Szene: Sie sind erkennbar mit einem Teleobjektiv durch die Absperrungskette der Polizei gefilmt, zeigen also etwas, was nicht gezeigt werden soll. Das Bild von der Absperrkette (10) (siehe oben) begleitet die Aussage von der Erschießung, woraus man paradoxerweise schließen muss, dass sie eben nicht ganz *auf offener straße* stattfand (der Duden notiert dafür als Bedeutung: ‚vor den Augen aller, die sich auf einer Straße befinden'). Und man sieht den Bombenentschärfer in seinem gespenstisch wirkenden Schutzanzug, was als Anzeichen für die situative Gefahr gedeutet werden kann. Da man nicht erkennen

kann, dass der Beamte nach dem Sprengsatz sucht, und auch der Sprechtext dies nicht erläutert, bleibt es bei der optischen Dramatisierung ohne den ursprünglich informativen Wert der Szene.

Zusammenfassend: Für die Beschreibung der transkriptiven Muster sind Genre-Spezifika zu berücksichtigen. Die Filme, die ich analysiert habe, sind Teile von Nachrichtensendungen, die vorgeben, ein möglichst genaues und „objektives" Abbild der Wirklichkeit zu erzeugen, auch wenn wir wissen, dass dies nicht möglich ist. Es geht also um Ereignisse der Wirklichkeit, die deshalb für die (hier audiovisuellen) Texte zwei zentrale Aufgaben stellen; sie müssen: (a) darstellen, was passiert ist, und (b) beglaubigen, dass es so passiert ist; es geht also um Darstellbarkeit und Glaubwürdigkeit.

Fragt man danach, wozu die Bilder die Sprachtranskription verwenden, wenn schon nicht brauchen, kommt man auf einen Darstellbarkeitmalus bzw. -bonus von Bild und Sprache (siehe Abb. 15). Umgekehrt profitiert auch die sprachliche Darstellung von der Bildtranskription, insofern als fotografische Bilder die grundsätzliche Brüchigkeit von symbolischen Zeichen, ihren Glaubwürdigkeitsmalus, durch eine indexikalische Struktur zu kompensieren scheinen: Wider besseres Wissen, glauben wir Bildern, die angeblich nicht lügen, mehr; sie haben sicherlich einen Glaubwürdigkeitsbonus.

	Bilder	*Sprache*
darstellbar?	Darstellbarkeitsmalus: Ich kann nicht alles (genau) zeigen	Darstellbarkeitsbonus: Ich kann fast alles (genau) sagen
glaubwürdig?	Glaubwürdigkeitsbonus: Was ich zeige, hat Beweiskraft	Glaubwürdigkeitsmalus: Nicht alles, was ich sage, hat Beweiskraft

Abb. 15: Potenziale/Defizite (Boni/Mali) von Sprache und Bild

Dazu kommen aber weitere, subtilere Funktionen. Zusammenfassend kann man die wechselseitige Sprach-Bild-Transkription des kurzen ZDF-Berichts als ein wohlbalanciertes Zusammenspiel von bildlich vermitteltem Plot, der durchaus eine gewisse Dramatik vermittelt, und einem im Meldungsstil sachlich verankernden Sprachtext, der durch detaillierte Propositionen die Bilder aufschließt, dabei bewertend kommentiert und mit relativierenden bzw. legitimierenden Formulierungen alles Brisante wieder normalisiert. Verallgemeinernd kann man für das Verfahren der Nachrichtenfilm-Audiovisualität folgende Aspekte festhalten:

- „Oszillieren“ zwischen Sprache und Bild als wechselseitige ‚Transkription‘,
- typische Konstellationen/Muster und Abfolgen von Sprach- und Bildfunktionen,
- performative transkriptive Dynamik, wie in einem Reißverschluss,
- auf der Grundlage von eigenständigen semantischen Eigenschaften der verschiedenen Zeichentypen,
- auf der Grundlage von Kontext- und Schemawissen.

5. Sprach-Bild-Transkriptionen II: Bebildertes Sprechen – Polit-Talkshows

Es ist sofort klar, dass in Polit-Talkshows, die ich jetzt anhand eines Beispiels behandeln will, das Sprach-Bild-Verhältnis völlig anders als in den Nachrichtenfilmen gestaltet ist. Während dort reichhaltigstes Bildmaterial durch einen Voice-over-Sprecher gewissermaßen „besprochen“ wird, sind hier von Vornherein (fast nur) Sprecher im Fokus, die nun von Kameras gezeigt werden, deren Rede gleichsam „bebildert“ werden muss (siehe auch Holly in Vorbereitung).[3]

Diese fundamentale Umgewichtung soll nun aber nicht bedeuten, dass nicht in beiden Fällen wechselseitige Transkriptionen im Spiel sind. Oft sind schon beim Drehen der Nachrichtenfilmbilder *sprachliche* Konzepte und Schemata relevant; erst recht wirken bei der Auswahl der montierten Einstellungen auch Erfordernisse und Konventionen der Nachrichtensprache und – wie gezeigt wurde – deuten nicht nur die Sprechertexte die Bilder aus, sondern diese wirken auch auf das Sprachverstehen ein. Komplementär dazu ist zwar die Rede der Talkshowteilnehmer in einem gewissen Sinne primär, andererseits ist sie von Anfang an von der Tatsache beeinflusst, dass die Sprecher, ihre räumliche Präsenz, ihre körperliche Performanz und deren Inszenierung *gesehen* werden können, und zwar durch die Augen von Kameras. Wie jemand spricht und wie seine Rede verstanden wird, hängt auch davon ab, ob und wie ich ihn performierend sehen kann, wo er agiert und wie er mir gezeigt wird.

Das visuell Relevante in solchen performativen Gesprächssendungen kann entsprechend (siehe schon Holly/Kühn/Püschel 1986, S. 177–198) in drei Dimensionen erfasst werden:

a) Räumliches Arrangement: Setting, Design, Szenerie;
b) Körpersprachliches: Mimik, Gestik, Kinesik, Proxemik;
c) Kamerainszenierung: Einstellungen in Abfolgen (durch Umschnitte) mit Zooms, Schwenks, Fahrten; Inserts.

[3] Von den immer häufiger verwendeten Einspielfilmchen soll hier abgesehen werden; siehe dazu Klemm (in Vorbereitung).

Hier will ich mich ausschließlich mit dem dritten Aspekt befassen, der allerdings insofern die beiden ersten einschließt, als er darüber entscheidet, was von diesen und wie wir sie als Zuschauer überhaupt wahrnehmen können. Dabei verfolge ich – wiederum im Rahmen des Transkriptivitätsmodells – folgende These:

Die Kameraführung in Polit-Talkshows generiert durch die Selektion von Einstellungen und Umschnitten neue Bedeutungskomponenten, die sprachliche Äußerungen „transkribieren", d.h. überformen, implizit kommentieren und dadurch „anders lesbar" machen. Brisant ist vor allem, dass so dem Sprecher die alleinige Auktorialität entzogen wird und er die „performative Letztfassung" seiner Äußerungen nicht mehr selbst kontrolliert, sondern sich partiell an Instanzen technischer Medialität ausliefert, die dann auch nahezu unmerklich auf den Rezipienten wirken können.

Bei dem Versuch, diese Bild-Sprach-Transkriptionsprozesse deutend zu erfassen, lassen sich wiederum drei Aspekte oder Funktionen der Kamerainszenierung unterscheiden:

1) Profilierung des (thematischen und kommunikativen) Verlaufs durch Abwechslung, Gliederung, Strukturierung (auf der Makro- und Mikroebene);
2) Profilierung der Sprecher(selbst)darstellung durch dynamische Gestaltung der Sprecheridentität auf der Mikroebene;
3) Profilierung der Beteiligungsrollen anderer (Adressaten, Unterstützer – Gegner) und Verweise auf Kontexte (Personen als Frame-Repräsentanten) durch Gestaltung der Beziehungen zwischen den Protagonisten und zum Diskurs.

Die Kamerainszenierung betrifft also elementare Funktionsfelder des ubiquitären „Vorbereichs" jeder Kommunikation, nämlich ihre Organisation und die Bearbeitung von Beziehungen und Identitäten der Kommunizierenden.

Die Beispielsendung, auf die ich mich beziehe, ist die Ausgabe der Sendereihe „Maybrit Illner" vom 29. März 2007 mit dem Thema: „Lebenslänglich, trotzdem frei: Gnade für die RAF?"; sie stand im Kontext der Frage, wie sich der Bundespräsident zum Gnadengesuch des RAF-Häftlings Christian Klar verhalten solle. Als Protagonisten waren 5 Gäste geladen, die von der Moderatorin Maybrit Illner (auf der Abb. 16 unten 3. v.l.) zu Beginn folgendermaßen vorgestellt wurden:

– *Ina Beckurts – ihr mann wurde von RAF-terroristen ermordet* [2. v.r.];
– *Rupert von Plottnitz – er hat RAF-terroristen verteidigt* [ganz r.];
– *Claus Peymann – der Berliner intendant hat Christian Klar ein praktikum in seinem theater angeboten* [2. v.l.];
– *Roland Koch – der hessische ministerpräsident fordert: leute wie Peymann dürfen dieses schreckliche kapitel deutscher geschichte nicht noch verklären* [3. v.r.];
– *und Klaus Bölling – der ehemalige regierungssprecher meint: die geschichte der RAF ist nicht wirklich zu ende* [ganz l.].

Abb. 16: Setting und Teilnehmer „Maybrit Illner“, 29. März 2007

Bevor ich kurz auf die drei genannten Funktionen der Kamerainszenierung eingehe, gebe ich einen schematischen Überblick über die Einstellungen während der ersten Runde von Beiträgen der Gäste, die von der Moderatorin nacheinander durch Interviewfragen regelrecht abgerufen werden. Das Schema (Abb. 17) zeigt (von unten gelesen), wie die Beiträge immer länger werden (von anfänglich knapp 1 Minute bei Peymann auf über 4 Minuten bei Bölling), was natürlich auch damit zu tun hat, dass jeder folgende Sprecher auch das bisher von andern Gesagte aufgreifen und kommentieren möchte.

Zur ersten Funktion der Kamerainszenierung, die für Abwechslung, Gliederung und Strukturierung steht, soll hier nur kurz festgehalten werden (ausführlicher Holly in Vorbereitung): Das Schema zeigt auch, wie Bilder vom jeweiligen Sprecher (im Schema hell gefärbt) im Wechsel mit anderen Einstellungen zu sehen sind, die nicht nur den Sprecher, sondern auch anderes zeigen: andere Teilnehmer, die als Rezipienten der jeweiligen Sprecheräußerung ins Bild kommen, oder aber den Sprecher mit anderen zusammen, dazu Schwenks und Fahrten. Es ist auffällig, dass die Moderatorin öfters gegen Ende der Beiträge eingeblendet wird, so dass der Sprecherwechsel auch schon optisch vorbereitet wird.

Der häufige Wechsel der Bilder erzeugt also Abwechslung und Reizerneuerung, er entspricht damit aber auch unserem natürlichen Blickverhalten, das nicht über längere Zeit starr auf ein Objekt gerichtet bleibt. Es gibt also nur relativ kurze Zeitspannen der Kamerazuwendung; die Einstellungen vom Sprecher sind im Durchschnitt 9,5 sec lang, die von anderen 4,7 sec. Zur ständigen Reizerneuerung gehört auch, dass – anders als früher in solchen Sendungen – die Kamera fast niemals stillhält, sondern auch noch die Einstellungen einzelner Gäste durch minimale Zooms und Schwenks nahezu unmerklich variiert.

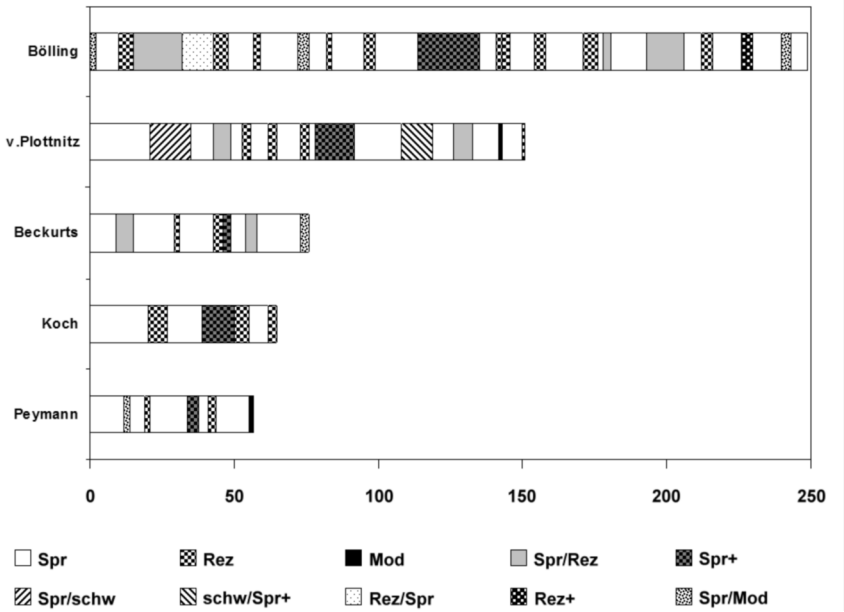

Abb. 17: Schema der Kameraeinstellungen in der 1. Runde von „Maybrit Illner"

Zu dieser mikrostrukturellen Gliederung kommt eine makrostrukturelle, die durch Umschnitte, größere Kamerabewegungen und gliedernde Einblendungen (sogenannte Bauchbinden) markiert werden. Immer häufiger werden auch Einspielfilmchen dazu genutzt, neue Subthemen oder andere Aspekte mit anderen Argumenten zum Thema einzubringen, wobei auch musikalische Elemente eine Rolle spielen.

Die zweite Funktion der Kamerainszenierung dient der zusätzlichen Profilierung des jeweils gezeigten Sprechers, der während seiner Rede zwischen 55% und 80% der Zeit (gemessen in dieser ersten Runde) allein im Bild ist; wie lange wir ihn allein sehen, hängt allerdings auch von der Gesamtlänge seines Beitrags ab und von der Häufigkeit, mit der er sich auf andere bezieht, was dazu führt, dass sie zur besseren Orientierung gezeigt werden. Die Konzentration der Kamera auf den Sprecher entspricht unserer natürlichen Erwartung, den Sprecher auch optisch zu identifizieren und ihm beim Reden zuzuschauen, damit wir auch die visuellen Botschaften, die er körpersprachlich mitliefert, verfolgen können. Natürlich wird durch die nahe Kameraeinstellung, die bei Sprecherbildern überwiegt, seine Mimik und Gestik überproportional fokussiert und damit verstärkt. Dazu kommen die aus der sozialsemiotischen visuellen Analyse von Bildern (Kress/van Leeuwen 1996; Jewitt/Oyama 2001) bekannten Funktionen von Kameraperspektiven, die ihren drei Dimensionen bestimmte, wenn auch prinzipiell ambivalente Bedeutungspotenziale zuweist:

– Einstellungsgröße: Distanz-Respekt/Intimität-Intensität
– Winkel horizontal: Involvement-Identifikation/Detachment-Skepsis
– Winkel vertikal: sozialer Status (Macht/Ohnmacht)

Wie schon angedeutet, handelt es sich nur um Potenziale, die erst in und von den jeweiligen Kontexten und Interpretationen aktiviert werden müssen. Den meisten Einfluss haben hier die Einstellungsgrößen und horizontalen Winkel, während die vertikalen Winkel in solchen Talkshows nur wenig variiert werden.

Zusammenfassend lässt sich festhalten, dass die Kameraführung permanent Ressourcen für eine sekundäre, sehr subtile „Kontextualisierung" des Sprechers generiert, die jeweils zwischen Polen sozialer Semiotik operiert. Mit den Augen der Kamera wird uns der Sprecher gewissermaßen aus einer bestimmten Haltung gezeigt: Spricht er (nach „Meinung" der Kamerainszenierung) so, dass man ihm näherkommen oder von ihm abrücken will, dass man ihm „auf den Zahn fühlen" muss oder respektvoll Abstand hält, dass man ihn frontal mit offenem Interesse von vorne betrachtet oder „skeptisch" von der Seite? Die Kameraprofilierung des jeweiligen Sprechers erlaubt also Rückschlüsse auf die Äußerung des Sprechers, sie kann gedeutet werden als visuelles *face work* im Sinne Goffmans, indem sie relevante Aspekte der Beziehung zum Sprecher optisch wahrnehmbar zum Ausdruck bringt.

Die dritte Funktion der Kamerainszenierung, die ich hier etwas ausführlicher behandeln will, betrifft die externe Kontextualisierung von Beteiligungsrollen und Frames, also die Frage, wie durch die Einblendung anderer Teilnehmer der jeweilige Redebeitrag im Hinblick auf Positionen anderer profiliert wird, um diskursive Spannungen und Verstärkungen zu erzeugen.

Wenn Hörer im Bild zu sehen sind, handelt es sich immer um eine gezielte Auswahl. Dabei gibt es grundsätzlich zwei Möglichkeiten. Die Entscheidung der Bildregie, einen Teilnehmer (oder das Publikum) zu zeigen, kann zurückgehen auf den Sprecher selbst, der durch direkte Anrede oder durch Erwähnung einen Diskussionspartner ins Spiel bringen kann. Oder aber es handelt sich direkt um eine Wahl der Bildregie, die damit die Kontextualisierung eines Anwesenden als betroffen, als zustimmend oder ablehnend vornimmt. Mit dem Zeigen eines anderen Teilnehmers werden also implizit Beteiligungsrollen zugeschrieben, es werden Diskussionsrollen und damit Beziehungen zwischen den Teilnehmern inszeniert und zugleich wird so der vorangegangene oder aktuelle Sprechertext in einer bestimmten Weise transkribiert, ihm werden andere Beteiligte zugeordnet. Dabei sind die Hörer fast immer „groß" oder „sehr groß" im Bild, denn man soll von ihren Mienen sehr differenziert ablesen können, wie sie zum Gesagten stehen, es geht um ihre mimischen Kommentare.

Darüber hinaus werden Personen als „Framevertreter“ gezeigt. Hat sich ein Teilnehmer explizit zu einem Thema oder Subthema geäußert, kann er von da an als Vertreter des entsprechenden semantischen Frames gelten, so dass seine Reaktion bei Wiedererwähnung des Frames von Interesse ist; so werden Kohärenzen visuell kontextualisiert, Frames auch optisch verstärkt. Zugleich werden Erwartungen geweckt, dass potenzielle Kontroversen sichtbar werden könnten, nach dem Motto: „Mal sehn, wie x dazu steht.“

Diese Verfahren sollen nun an einem Beispiel veranschaulicht werden. Im Verlauf der ziemlich langen Ausführungen von Klaus Bölling (siehe Abb. 18) werden gehäuft andere Beteiligte gezeigt, insgesamt in 18 Einstellungen (im Schema dunkler), gegenüber 14 ausschließlich vom Sprecher (im Schema heller). Ich greife nun die Einstellungssequenz 26–29 heraus, um zu sehen, wie der Sprechertext mit der Einstellungssequenz zusammenspielt.

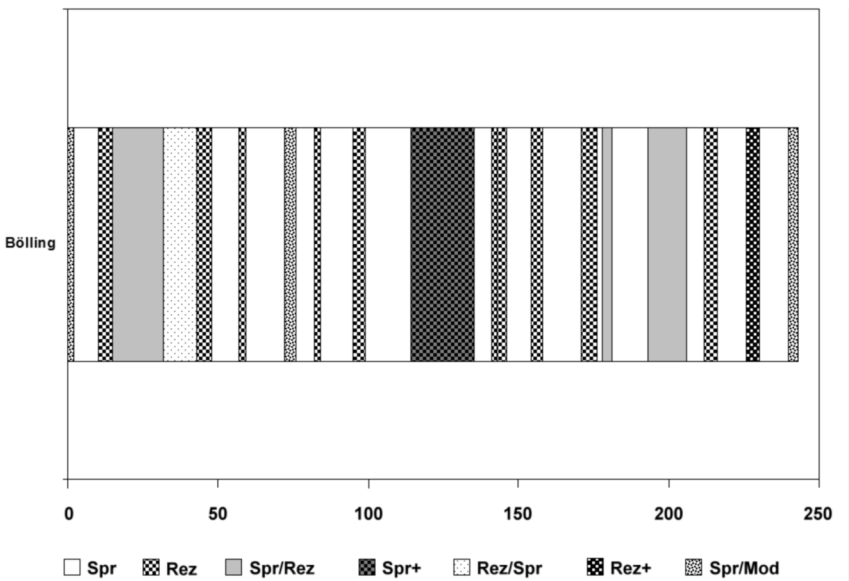

Abb. 18: Einstellungssequenz Bölling

Bölling, der während dieser Passage nur zwei Mal alleine im Bild ist, äußert dabei folgenden Sprachtext:

> [26 Bö] *bestimmte islamistische fundalisten* [sic] *sagen wir sind gesinnungstäter wir wollen die scharia und wir orientieren uns an den gesetzen des kóran und morgen* // [27 Schwenk Ohr Bö Koch-Illner-Peymann/Zoom] *frisch drauf los so geht das nicht wir müssen eh / alle rechtsstaatlichen gesetze regeln müssen wir / beachten und wenn herr Peymann äh diesem mann eine chance geben will da würde ich gar nicht protestieren* // [28 Bö] *ich sehe nur dass der herr Klar äh nicht nur gar keine reue zeigt und was reue kamman* // [29 Plottnitz] *spielen reue kamman heucheln sie kann auch mal* […]

Während Bölling im Kontext der Terrorismusdebatte für eine harte Haltung des Staates plädiert und dazu auch islamistische Fundamentalisten anführt (im Eifer als „Fundalisten" bezeichnet), wird als potenzieller Unterstützer Koch über den Hinterkopf von Bölling eingeblendet. Da Kochs Mimik aber zurückhaltend bis skeptisch bleibt, schwenkt die Kamera über Illner auf Peymann, als Bölling diesen gerade erwähnt. Daraufhin bleibt die Kamera auf Peymann und zoomt ihn ein bisschen näher heran, fast bis zur Großaufnahme, wie sie für solche Zwecke üblich ist (siehe Abb. 19–23).



Abb. 19–21: Bölling als Sprecher, Koch als Unterstützer, Schwenk über Illner

Als Bölling im weiteren Verlauf seiner Argumentation auf die Frage kommt, ob Christian Klar, um den es die ganze Zeit eigentlich geht, Reue zeigt, wird nun von Plottnitz eingeblendet (Abb. 24), der sich in seinem Statement explizit zur Frage geäußert hatte, ob die Terroristen Reue zeigen. Dort hatte er gesagt:

> *vielleicht noch ein satz zu der reue weil das auch überall ne rolle spielt* […] *ob es reuebezeugungen gegeben hat des weiß wissen nur die die die akten kenn // des sin ja nichtöffentliche verfahren*



Abb. 22–24: Schwenk auf Peymann, Peymann herangezoomt, von Plottnitz wegen „Reue"

Hier zeigen sich also die verschiedenen Verfahren in der Einblendung von Beteiligten: Man zeigt solche, die aufgrund von Vermutungen der Bildregie in einem bestimmten Verhältnis zum Gesagten stehen, dann erwähnte Beteiligte, schließlich solche, die im bisherigen Verlauf der Debatte für ein bestimmtes Thema relevant sind.

Es gilt also: Die (meist kurze) Einblendung von Mitdiskutanten (meist in Großaufnahme), die sie zu „markierten Hörern" macht, ist zugleich eine externe Weise, den Sprechertextabschnitt zu „kontextualisieren", als besonders relevant im Hinblick auf den Gezeigten, der zum Adressaten, Unterstützer oder Kontrahent bzw. zum Repräsentanten eines Themas stilisiert oder als solcher hervorgehoben wird; damit wird der Sprechertext (nachträglich) profilierend „transkribiert", zum Zweck der dramatisierenden Kommentierung. Dabei werden die Beteiligten typisiert: Ihnen werden von Anfang an (schon in der Vorstellung und immer wieder durch Inserts mit

Etikettierungen) stereotype Rollen im entsprechenden Diskurszusammenhang zugewiesen, z.B. als Provokateure, Scharfmacher, Kontrahenten, Betroffene, Unterstützer, Neutrale. Diese Typisierungen strukturieren die gesamte Dramaturgie, von der Einladung über die Sitzordnung, die Vorstellung bis hin zu den Inserttexten. Die Kameraführung inszeniert dann unterstützend entsprechende Interaktions- und Konfliktlinien.

Fragt man zusammenfassend danach, wie die Kamerainszenierung in Polit-Talkshows den Sprachtext transkribiert, lässt sich festhalten: Die Muster der Sequenzierung von Einstellungen durch Umschnitt kann man als „kulturelle Praktiken" der Bedeutungskonstitution auffassen. Im Zusammenhang mit Sprachperformanz handelt es sich um ein „sekundäres System", das Sprachliches überformt. Es ist ein implizites, unauffälliges, dennoch nicht beliebiges Vorgehen, das ohne explizites Regelwerk betrieben wird. Je nach Erfahrung und Ausbildungsstand der Macher ist es mehr oder weniger reflektiert. Meine bisherige Analyse stützt sich überwiegend auf die Produktanalyse und einige generelle Beobachtungen bei der Produktion solcher Sendungen. Dazu müssten aber Produktionsanalysen kommen, die im Stile der „work place studies" nach dem impliziten Wissen der Macher forschen.

Die hier angestellten Überlegungen erscheinen mir in dreierlei Hinsicht relevant. Medientheoretisch können sie als Beleg dafür gesehen werden, wie ein Mediendispositiv, eine technische Struktur, gleichsam „hinter dem Rücken" der Akteure wirkt. Kommunikationstheoretisch handelt es sich um mehrfach gefilterte Botschaften; im Sinne Luhmanns geht es um „Beobachtung von Beobachtung von Beobachtung" – der Zuschauer sieht, wie die Kamera sieht, wie die Akteure etwas sehen. Die Sichtweisen der Kamera sind dabei zumeist „transparent" und „naturalisiert", indem sie wirken, auch ohne dass wir sie bewusst wahrnehmen.

6. Kurzes Fazit

Technische („sekundäre") Audiovisualität generiert hochkomplexe Bedeutungen, und zwar durch die getrennte und kombinierte Inszenierung semantisch heterogenen Materials, das durch wechselseitige Transkription nach bestimmten Mustern aufeinander bezogen ist. Solche Muster müssen genre- und kommunikationsformspezifisch beschrieben und analysiert werden, was hier an zwei Beispielen von Fernsehsendungen ausschnitthaft versucht wurde.

In Nachrichtenfilmen geht es um die Darstellbarkeit und Glaubwürdigkeit von Ereignissen, wobei die kommunikativen Konstruktionen hergestellt werden durch typische Konstellationen von wechselseitig aufeinander

bezogenen Sprach- und Bildfunktionen, die sich auf der Basis von Kontext- und Schemawissen in einer performativen transkriptiven Dynamik wie einem Reißverschluss verbinden.

In Polit-Talkshows transkribiert die Kamerainszenierung Sprecheräußerungen zum Zwecke von a) Abwechslung und Gliederung, b) Sprecherprofilierung und c) Profilierung von Beteiligungsrollen anderer, so dass die Organisation des Gesprächs übersichtlicher und die Beziehungsgestaltung zu und zwischen den Beteiligten intensiver und für den Zuschauer attraktiver werden (Stichwort: Quote!).

Literatur

Barthes, Roland ([1964] 1990): Die Rhetorik des Bildes. In: Barthes, Roland: Der entgegenkommende und der stumpfe Sinn. Frankfurt a. M., S. 28–46. [Frz. Orig. 1964].

Brosius, Hans-Bernd (1998): Visualisierung von Fernsehnachrichten. Text-Bild-Beziehungen und ihre Bedeutung für die Informationsleistung. In: Kamps, Klaus/Meckel, Miriam (Hg.): Fernsehnachrichten. Prozesse, Strukturen, Funktionen. Opladen/Wiesbaden, S. 213–224.

Burger, Harald (2005): Mediensprache. Eine Einführung in Sprache und Kommunikationsformen der Massenmedien. 3., völlig neu bearb. Aufl. Berlin/New York.

Girnth, Heiko/Michel, Sascha (Hg.) (i.Vorb.): Multimodale Kommunikation in Polit-Talkshows. Stuttgart.

Holly, Werner (2003): „Ich bin ein Berliner" und andere mediale Geschichts-Klischees. Multimodale Stereotypisierungen historischer Objekte in einem Fernsehjahrhundertrückblick. In: Schmitz, Ulrich/Wenzel, Horst (Hg.): Wissen und neue Medien. Bilder und Zeichen von 800 bis 2000. (= Philologische Studien und Quellen 177). Berlin, S. 215–240.

Holly, Werner (2008): Audiovisuelle Sigetik. Über verborgene Bedeutungen im Bild-Sprach-Zusammenhang. In: Pappert, Steffen/Schröter, Melani/Fix, Ulla (Hg.): Verschlüsseln, Verbergen, Verdecken in öffentlicher und institutioneller Kommunikation. (= Philologische Studien und Quellen 211). Berlin, S. 147–169.

Holly, Werner (2009): Bildüberschreibungen. Wie Sprechtexte Nachrichtenfilme lesbar machen (und umgekehrt). In: Diekmannshenke, Hajo/Klemm, Michael/Stöckl, Hartmut (Hg.): Bildlinguistik. Berlin. [ersch. demn.].

Holly, Werner (i.Vorb.): Bildinszenierung in Talkshows. Medienlinguistische Anmerkungen zu einer Form von „Bild-Sprach-Transkription". In: Girnth/Michel (Hg.).

Holly, Werner/Kühn, Peter/Püschel, Ulrich (1986): Politische Fernsehdiskussionen. Zur medienspezifischen Inszenierung von Propaganda als Diskussion. (= Medien in Forschung + Unterricht, Serie A, 18). Tübingen.

Huth, Lutz (1985): Bilder als Elemente kommunikativen Handelns in Fernsehnachrichten. In: Zeitschrift für Semiotik 7, S. 203–234.

Jäger, Ludwig (2002): Transkriptivität. Zur medialen Logik der kulturellen Semantik. In: Jäger, Ludwig/Stanitzek, Georg (Hg.): Transkribieren. Medien/Lektüre. München, S. 19–41.

Jäger, Ludwig (2004): Die Verfahren der Medien: Transkribieren – Adressieren – Lokalisieren. In: Fohrmann, Jürgen/Schüttpelz, Erhard (Hg.): Die Kommunikation der Medien. (= Studien und Texte zur Sozialgeschichte der Literatur 97). Tübingen, S. 69–79.

Jäger, Ludwig (i.d.Bd.): Intermedialität – Intramedialität – Transkriptivität. Überlegungen zu einigen Prinzipien der kulturellen Semiosis.

Jewitt, Carey/Oyama, Rumiko (2001): Visual meaning. A social semiotic approach. In: van Leeuwen, Theo/Jewitt, Carey (Hg.): Handbook of visual analysis. London u.a., S. 134–156.

Keppler, Angela (2006): Mediale Gegenwart. Eine Theorie des Fernsehens am Beispiel der Darstellung von Gewalt. Frankfurt a.M.

Klemm, Michael (i.Vorb.): Zur Gestaltung und strukturellen Rolle der Einspielfilme in „Hart, aber fair". In: Girnth/Michel (Hg.).

Kress, Gunther/van Leeuwen, Theo (1996): Reading images. The grammar of visual design. London/New York.

Meyer, Thomas/Ontrup, Rüdiger/Schicha, Christian (2000): Die Inszenierung des Politischen. Zur Theatralität von Mediendiskursen. Wiesbaden.

Rauh, Reinhold (1987): Sprache im Film. Die Kombination von Wort und Bild im Film. Münster.

Sachs-Hombach, Klaus (2003): Das Bild als kommunikatives Medium. Elemente einer allgemeinen Bildwissenschaft. Köln.

van Leeuwen, Theo (2005): Introducing social semiotics. London/New York.