

Aus: Diekmannshenke, Hajo / Klemm, Michael / Stöckl, Hartmut (Hg.) 2011: *Bildlinguistik. Theorie - Methoden - Fallbeispiele*. Berlin: 123-156.

Hans-Jürgen Bucher

Multimodales Verstehen oder Rezeption als Interaktion
Theoretische und empirische Grundlagen einer systematischen
Analyse der Multimodalität

1. Einführung: Zum Begriff der Multimodalität

1.1. Der multimodale Turn: neue Probleme – alte Theorien?

Wenn die moderne Kommunikationsgeschichte als „visuelle Wende“ als „Pictorial turn“ charakterisiert wird (Mirzoeff 1998; Mitchell/Thomas 1995; Sachs-Hombach 2003), so ist das nur die halbe Wahrheit. Der grundlegende Wandel der Kommunikationsverhältnisse besteht nicht darin, dass zunehmend Abbildungen die Textkommunikation ergänzen oder einschränken, wie das beispielsweise Nicolas Mirzoeff beschreibt: „the world-as-a-text has been challenged by the world-as-a-picture [...], the visual disrupts and challenges any attempt to define culture in purely linguistic terms“ (Mirzoeff 1998: 5). Der grundlegende Wandel besteht darin, dass neue und neuartige Mischformen der verschiedensten Kommunikationsmodi und Kanäle entstanden sind, die man als multimodale Kommunikationsformen bezeichnen kann. Bild und Text sind – aufgrund der philologischen Wissenschaftsgeschichte und ihres derzeitigen visuellen Paradigmenwechsels – nur die prominentesten Vertreter einer Vielfalt von Kommunikationsmodi wie Design, Typografie, Farben, Grafiken, Piktogramme oder operationale Zeichen. Jedes Formular macht deutlich, dass wir nicht nur den Text verstehen müssen, sondern auch die spezifische Anordnung der einzelnen Textelemente – das Layout – oder den Aufforderungscharakter bestimmter Zeichen, etwas anzukreuzen oder auszufüllen. Die Tageszeitung hat sich im Laufe ihrer 400jährigen Geschichte von einem buchähnlichen Textmedium in ein Multi-Medium (Blum/Bucher 1998) verwandelt, in dem Texte, Fotos, Grafiken und Design zusammen eine komplexe Kommunikationsform bilden. Der Fernsehbildschirm einer Nachrichtensendung zeigt inzwischen nicht mehr nur den Text vortragenden Sprecher, sondern hat Hintergrundabbildungen, Inserts, Senderlogos und inzwischen auch in machen Fällen mehrere durchs Bild laufende Tickertexte. Aus der analogen Wählscheibe des Telefons mit Zahlen von 1 bis 10 ist ein interaktives Display mit vielfältigen Symbolen und Icons geworden, die operative Funktionen für die unterschiedlichsten Gebrauchsweisen erfüllen. Wissenschaftliche Vorträge, früher von einem Redner vorgetragen, sind inzwischen komplexe Kommunikations-

ereignisse mit Performance-Charakter, bei denen neben dem Redner nicht nur die Wandtafel, sondern auch mediale Präsentationen zum Einsatz kommen.

Die von Mitchell formulierte Einsicht, dass „all media are mixed media, and all representations are heterogenous; there are no „purely“ visual or verbal arts“ (Mitchell/Thomas 1995: 5) ist vor diesen Hintergrund nachvollziehbar, auch wenn vielfach in der Forschung noch eine Tendenz besteht, Kommunikation zu purifizieren – in der Linguistik zu reinen Textangeboten, in der Bildwissenschaft zu reinen Bildern.

Mit der Auffassung „all texts are multimodal“ gehen Kress und van Leeuwen noch einen Schritt weiter (Kress / van Leeuwen 1998; vgl. auch Baldry / Thibault 2005: insb. 19). Einerseits wird die Liste der Modi über Text und Bild hinaus um Musik, Layout, Typographie, Farbe etc. erweitert. Andererseits werden nicht nur medialen Kommunikationsformen verschiedene Kommunikationsmodi zugestanden, sondern allen Kommunikationsformen. Das Sprechen wird von Gestik, Mimik und Intonation begleitet, die Schriftkommunikation weist immer auch ein visuelles Arrangement der Schriftzeichen auf einer Seite als Gestaltungseinheit auf. Insofern ist daraus die Schlussfolgerung zu ziehen dass alle Kommunikationsanalyse multimodal ausgerichtet sein muss und zeigen sollte, wie sich Sinn und Bedeutung eines Kommunikationsbeitrags aus den unterschiedlichen Modi zusammensetzt. Stellvertretend für eine ganze Reihe von Theorien zur Multimodalität sei das Modell von Baldry/Thibault genannt, demzufolge Bedeutung in multimodalen Texten multiplikatorisch und nicht additiv entsteht und das davon ausgeht, dass dieser Prozess der Bedeutungszuweisung von einem Ressourcen-Integrationsprinzip („resource integration principle“) gesteuert ist (Baldry / Thibault 2005: 19).

Unter dieser Perspektivenerweiterung stellen sich die klassischen Probleme der Sprachwissenschaft und der Kommunikationsanalyse, wie das Problem der Kohärenz, der Kompositionalität, des Verstehens oder der kommunikativen Kompetenz in neuem Licht dar. Kommunikative Zusammenhänge können nicht nur inter- und intratextuell sein, sondern auch inter-modal. Das Problem der Kompositionalität stellt sich nicht nur auf der textlichen Ebene als Frage nach dem Aufbau des Textsinns aus den lexikalischen, semantischen und pragmatischen Einheiten. Angesichts der Ko-Okkurrenz verschiedener semiotischer Ressourcen und intermodaler Relationen (Iedema 2003) ist auch zu fragen, welchen Beitrag Elemente aus verschiedenen Modi zum Gesamtsinn eines Kommunikationsbeitrags leisten und wie diese Leistungen integriert sind.

Damit ist vom Gegenstandsbereich her die Sphäre der Textanalyse überschritten, so dass offen ist, in wieweit linguistische Theorien zur Klärung multimodaler Fragestellungen ausreichen. Mit der Erweiterung linguistischer Analysen auf Text-Bild-Zusammenhänge (Muckenhaupt 1986; Stöckl 2004; Holly 2005; Martinec/Salway 2005) sind bereits Ansätze erkennbar, die diese Ausweitung des Gegenstandsbereichs der Kommunikationsanalyse in Angriff genommen haben. Dabei lassen sich grob zwei Richtungen unterscheiden: Ansätze, die Text-Bild-Relationen ausgehend von linguistischen Theorien analysieren (Muckenhaupt 1986; Stöckl 2004; Holly 2005) und Ansätze, die einen bildtheo-

retischen Ausgangspunkt haben (Martinec/Salway 2005; Barthes 1990 (Orig. 1982)).

Einen Schritt weiter gehen die Forschungsarbeiten zu Formen multimodaler Kommunikation: sie berücksichtigen auch andere Kommunikationsmodi wie das Design (Kress / van Leeuwen 2001; Bucher 2007; Bateman 2008), Musik und Geräusch (Stöckl 2007), Farben (Kress / van Leeuwen 2002) oder die Typographie (Stöckl 2004, Stöckl 2005, Stöckl 2008, Antos / Spitzmüller 2007). Eine Theorie der multimodalen Kommunikation und des multimodalen Verstehens müsste so allgemein formuliert sein, dass sie für alle Modi gleichermaßen gültig ist. Text-Bild-Zusammenhänge sind dementsprechend ein Spezialfall intermodaler Relationen und unterliegen denselben analytischen Anforderungen wie beispielsweise Text-Musik- oder Text-Design-Zusammenhänge. Die folgenden Überlegungen beschränken sich auf multimodale Formen der Medienkommunikation im Fernsehen, Internet und in Printmedien.

1.2. Multimodalität und Non-Linearität

Ein Blick in die Geschichte der Medien zeigt, dass sich multimodale Erscheinungsform in ihrem Verlauf erst allmählich herausgebildet und ausdifferenziert haben. So waren die Zeitungen bis in die zweite Hälfte des 19. Jahrhunderts reine Textmedien in der Tradition des Buches (Bucher 1998; Barnhurst/Nerone 2001) und auch die Informationsangebote des Internet haben sich von textorientierten Nachrichtensammlungen erst allmählich zu den komplexen Angeboten aus Text, Bild, Video, statischen und dynamischen Grafiken gewandelt (Bucher 2006). Auch das Interface des Fernsehens hat nicht von Beginn an die heutige multimodale Komplexität besessen. Erst schrittweise wurden beispielsweise dem Nachrichtensprecher oder der Nachrichtensprecherin weitere Elemente, wie Hintergrundabbildungen, Themenüberschriften, Senderlogos oder zusätzliche Tickerbänder, hinzugefügt.

Für alle drei Medien lassen sich dementsprechend zwei Entwicklungstendenzen unterscheiden: die *Multimodalisierung* und die *Delinearisierung*. Die Medienkommunikation wird nicht nur über verschiedene Modi abgewickelt, sondern sie eröffnet dem Rezipienten auch die Freiheit, die Abfolge, in der die Angebote rezipiert werden, selbst zu bestimmen. Dieser Freiraum ist im Falle der Zeitung und des Internet bedeutend größer als im Falle des Fernsehens, das trotz allem ein zeitgebundenes Medium geblieben ist. Allerdings eröffnet der Fernseh Bildschirm durch die simultan präsentierten Elemente auch die Möglichkeit, über Rezeptionsreihenfolge und Ausblendung zu entscheiden. Das Phänomen der Multimodalität lässt sich dementsprechend ohne Berücksichtigung der Non-Linearität nicht sinnvoll analysieren. Auf der Grundlage dieser konstitutiven Verbindung schlägt Lemke (2002: 301) den Begriff „Hypermodality“ vor: „Hypermodality ist the conflation of multimodality and hypertextuality“. Eine Perspektive, die Non-Linearität und Multimodalität integriert, ist insofern produktiv, als sie einen umfassenderen Rahmen für die Analyse der entsprechenden Phäno-

mene bereitstellt. Während in der Multimodalitätsforschung die Frage nach dem Zusammenhang einzelner Elemente innerhalb eines multimodalen Clusters im Zentrum steht, eröffnet die hypermodale Sichtweise auch Fragen, die auf die Kohärenz und die Erschließbarkeit des gesamten Kommunikationsangebotes zielen. Die partikulare oder grammatische Sichtweise wird dadurch um eine holistische oder pragmatisch-kontextuelle Sichtweise ergänzt. Im Folgenden wird der Ausdruck „Multimodalität“ aufgrund seiner Etabliertheit beibehalten, wobei die non-lineare Seite stets mitgemeint ist.

Die bisherigen rezeptionsbasierten Systematisierungsvorschläge für multimodale Kommunikationsformen haben sich ausschließlich auf Kriterien wie körpernahe und temporäre Zeichenart versus körperunabhängige und fixierte (mediale) Zeichenart, oder wahrnehmungsnah versus arbiträr gestützt (Sachs-Hombach 2003: 96f.; Holly 2009: 3). Das Kriterienpaar linear versus non-linear blieb bislang unberücksichtigt. Dabei ist es unverzichtbar, um die spezifische Multimodalität medialer Kommunikationsformen aus einer Rezeptionsperspektive erschöpfend zu systematisieren. Mit dem erweiterten Kategoriensystem lassen sich folgende mediengebundenen Kommunikationsformen unterscheiden:

- *Lineare Kommunikationsformen*, wie Buchtexte, Hörfunkbeiträge; oder Fernsehfilmbeiträge,
- *Non-Lineare Kommunikationsformen* wie Online-Angebote, Zeitungen, Plakate,
- *Hybride Kommunikationsformen* mit linearen und non-linearen Elementen, wie Fernsehbeiträge mit Zusatz-Visualisierungen, Online-Angebote mit Videobeiträgen, Theaterinszenierungen mit simultanen Schauplätzen, oder Vorträge mit Medieneinsatz.

Die Grenzen zwischen diesen drei Typen sind nicht immer leicht zu ziehen. Auszeichnungen in linearen Texten oder deren Anreicherung mit Abbildungen können auch eine non-lineare Rezeption hervorrufen. Ein wesentliches Resultat der neueren Kommunikationsgeschichte ist gerade das Entstehen hybrider Formen.

Auf der Grundlage der vorausgegangenen Klärungen ist eine Bearbeitung der folgenden Problemfelder für eine Theorie der multimodalen Kommunikation und des multimodalen Verstehens konstitutiv:

Das Problem der Kompositionalität: Welchen Beitrag leisten einzelne Elemente eines multimodalen Clusters zu dessen Gesamtsinn? Voraussetzung für eine Klärung dieser Frage ist nicht nur die Beschreibung des Zusammenwirkens von Elementen in verschiedenen Modi sondern auch eine Klärung, was als einzelnes Element zählt. Insofern sind zwei Perspektiven erforderlich: eine holistische Perspektive, die vom Sinn der ganzen multimodalen Kommunikationseinheit – einer Zeitungsseite, eines Beitragsclusters, einer Webpage oder einer Website, einer Bildschirmdarstellung, einer Sendung oder eines Sendungsbeitrags – ausgeht; und eine partikulare Perspektive, die die kleinste bedeutungstragende Einheit eines Kommunikationsbeitrags ins Zentrum stellt – das Layout, die Typografie, die Farbe, das Foto, die Bildzeile, die Grafik, die Überschrift, den

Text, die Hintergrundmusik, das Ikon oder das Video. Zu diesem Problemfeld zählen auch die Frage nach der spezifischen Leistung einzelner Modi sowie die daraus ableitbare Frage nach modusspezifischen Deutungsmustern (Holly 2005: 342-344.) Zu klären ist in diesem Kontext auch die so genannte „semantische Multiplikation“ (Fei 2004: 239), also der Sachverhalt, dass der Gesamtsinn eines Kommunikates mehr ist als die Summe der Bedeutung einzelner Elemente. (Vgl. dazu Abschnitt 2)

Das Problem der Rezeption: In welcher Weise bauen Rezipienten ein Verständnis eines multimodalen Clusters auf? Voraussetzung für die Analyse dieses Problems ist eine dynamische Betrachtungsweise der Probleme des Verstehens. Zwar ist das Verstehen selbst kein Prozess, aber der Sinn eines multimodalen Angebotes lässt sich in der Regel nicht auf einen Schlag erfassen, sondern setzt einen Aneignungs- oder Erschließungsprozess – die Interpretation – voraus. Für eine Klärung dieses Prozesses lassen sich zwei unterschiedliche Modelle unterscheiden: die so genannte „*Saliency-Theorie*“ die davon ausgeht, dass auffallende Elemente des Rezeptionsgegenstandes den Aneignungsprozess steuern (Itti/Koch 2000). Die Rezeption wäre demzufolge ein bottom-up gelenkter Vorgang. Das Gegenstück dazu sind die *Schema-Theorien*, die davon ausgehen, dass Kompetenzen, kognitive Muster oder Intentionen für den Verlauf des Aneignungsprozesses entscheidend sind. Die Rezeption wird unter dieser Perspektive als Top-Down-Prozess verstanden (Schnotz/Bannert 2003). In diesem Beitrag soll gezeigt werden, dass für die Bearbeitung des Problems der Rezeption der Begriff der Interaktivität als Zentralbegriff genutzt werden kann (dazu Abschnitt 3).

Die beiden Problemfelder hängen eng miteinander zusammen: die Frage der Kompositionalität gehört in den Bereich der Semantik und befasst sich mit den koordinierten Verwendungsmöglichkeiten unterschiedlicher Zeichentypen. Das Problem der Rezeption geht aus von realen Verwendungsweisen verschiedener Zeichentypen in multimodalen Kommunikationsangeboten und fällt somit in den Bereich der Pragmatik. Die Unterscheidung der beiden Problemfelder lässt sich auch nutzen, um zwei verschiedene Verwendungsweisen der Begriffe „interaktiv“ „interagieren“ und „Interaktivität“ zu differenzieren: Während es im Falle der Kompositionalität um die Interaktion zwischen verschiedenen Elementen geht, ist im Rahmen des Rezeptionsprozesses die Interaktion zwischen Rezipient und Angebot angesprochen. Es wird auch zu zeigen sein, wie diese beiden unterschiedlichen Formen der Interaktivität miteinander zusammenhängen.

2. Sinn und Bedeutung multimodaler Angebote – die semantische Frage

Für multimodale Kommunikationsformen ist die Frage der Kompositionalität eng verbundenen mit dem Problem der semantischen Multiplikation: die Rekonstruktion des Sinns eines multimodalen Angebotes bedeutet auch zu erklären, in welcher Hinsicht die Bedeutung eines multimodalen Angebotes mehr ist als die Summe des semantischen Potentials seiner Teile. Dafür lassen sich drei typische Strategien unterscheiden: der *relations-grammatische Ansatz* führt die Gesamt-

bedeutung eines multimodalen Kommunikationsangebotes auf ein begrenztes Repertoire von intermodularen Relationen zurück, der *sozial-semiotische Ansatz* sieht im Design ein spezifisches Symbolsystem, das als kompositionelle Ressource multimodaler Angebote fungiert. Eine dritte Strategie, die man *pragmatisch-dynamisch* nennen kann, erklärt die Kompositionalität aus wechselseitigen Deutungen der kombinierten multimodalen Elemente.

2.1. *Der relations-grammatische Ansatz: Typologien multimodaler Relationen*

Eine der Strategien, multimodales Verstehen zu erklären besteht darin, eine Typologie von Zusammenhängen zwischen verschiedenen Elementen eines multimodalen Clusters aufzustellen. Dahinter steht die Annahme, dass Zusammenhänge zwischen Text-, Bild-, und Layout-Elementen nicht zufällig, sondern regelhaft und damit auch rekonstruierbar sind. Ein multimodales Cluster verstehen, heißt dann, die – irgendwie existierenden – Relationen zwischen modalen Elementen einem bestimmten Typus zuzuordnen. Als Vergleichsobjekt wird dabei häufig auf Text-Bild-Beziehungen zurückgegriffen. Kriterium für die Unterscheidung von Relationstypen ist die jeweilige *Direktionalität*: sind Text und Bild gleichwertig, bestimmt das Bild den Text oder der Text das Bild. Bei der Übertragung der Rhetorical Structure Theory auf multimodale Formen der Kommunikation, wird diese Anzahl der Relation auf über zwanzig ausgeweitet (Bateman 2008: Kap. 4, insb. 149). Überlegungen dieser Art gehen zumeist entweder auf Roland Barthes zurück, der im Rahmen einer „Rhetorik des Bildes“ u.a. auch Text-Bild-Relationen in ein System geordnet hat. Oder sie stellen Ableitungen aus dem Relationssystem dar, das Halliday für Textzusammenhänge entwickelt hat (Halliday 2004). Für die Zusammenhänge zwischen Sätzen unterscheidet Halliday zwei Grundrelationen, die Projektionen und die Expansionen. Ein Satz projiziert einen anderen entweder als Aussage oder als Idee. Oder der zweite Satz erweitert (expands) den ersten durch Konkretisierung (Elaboration), durch Ergänzung (Extension), oder durch Situierung (Enhancement) (vgl. dazu Halliday 2004: Kap. 7.2. mit Beispielen auf 380). Es ist leicht erkennbar, dass diese Relationstypen sich aufgrund ihrer Allgemeinheit problemlos auf Text-Bild-Zusammenhänge übertragen lassen: Das Porträt-Foto konkretisiert den Bildtext, indem es die genannte Person zeigt; das Unfallfoto ergänzt den Text um szenische Einzelheiten, die im Text nicht genannt sind; oder der Text situiert ein Foto, das ohne den Text unverständlich bliebe. Gute Beispiele für letzteres sind die in Wochenmagazinen turnusmäßig abgedruckten „Fotos der Woche“, deren Besonderheit gerade erst durch den entsprechenden Erklärtext gesehen werden kann. Den Typologien von Relationen liegt die Annahme zugrunde, dass, so wie einzelne Textsegmente die Kohärenz eines Textes konstituieren, auch die Kohärenz eines multimodalen Dokuments aus den Fotos, Grafiken, Texten und Layoutelemente abgeleitet werden kann.

Roland Barthes unterscheidet zunächst zwei grundlegende Funktionen, „die die sprachliche Botschaft in Bezug auf die bildliche Botschaft besitzt“ (Barthes

1990: 34): erstens die Verankerung des Bildes durch den Text, die Barthes als Kontrolle für das Bildverstehen auffasst: „Durch den Text beaufsichtigt der Schaffende [...] das Betrachten des Bildes“ (ebd. 35). Die zweite Funktion nennt Barthes Relaisfunktion, bei der Text und Bild in komplementärem Verhältnis stehen, wie das beispielsweise in einem Filmdialog der Fall ist: „die Worte sind dann, genauso wie die Bilder, Fragmente eines umfassenderen Syntagmas, und die Einheit der Botschaft entsteht auf einer höheren Ebene“ (ebd. 36).

Für die Analyse des Dokumentendesigns und der Entwicklung von Qualitätsstandards greift Karen Schriver diese Relationstypologien auf und unterscheidet ihrerseits fünf Schlüsselrelationen zwischen Texten und Abbildungen (Schriver 1997). Text und Bild können redundant sein, wenn beide dasselbe beinhalten, sie können komplementär sein, wenn ihre verschiedenen Inhalte gemeinsam und wechselseitig zum Gesamtverständnis beitragen; sie können ergänzend („supplementary“) sein, wenn eines der Elemente das dominierende andere mit zusätzlichen Angabe unterstützt; Text und Bild können integrativ oder synthetisch verwendet sein („juxtapositional“), wenn bei unterschiedlichem Inhalt „the key ideas are created by a clash or a semantic tension between the ideas in each mode“ (Schriver 1997: 413). Die fünfte und letzte Relation bezeichnet sie als „stage-setting“, „in which one mode forecasts the content, underlying theme, or ideas presented in the other mode“ (ebd.). So einleuchtend auf den ersten Blick diese Grammatiken der Relationen erscheinen, im Detail betrachtet werfen sie eine ganze Reihe von Problemen auf.

Eine erste Schwierigkeit ergibt sich aus der Voraussetzung, auf der alle Unterscheidungen von Relationen aufbauen: der Inhalt der jeweils zusammenhängenden Elemente muss unabhängig von der Art ihres Zusammenhangs bestimmbar sein. Nur wenn beispielsweise feststeht, dass ein Bild etwas zeigt, was durch den Text konkretisiert wird, kann eine supplementäre Relation angenommen werden. Was ein Bild zeigt, ist aber nicht eine Funktion des Bildes, sondern eine Funktion seines Verständnisses, das beispielsweise entscheidend von Wissen des Betrachters abhängt. Dass Text und Bild in einem Gegensatz oder einer „semantischen Spannung“ stehen (vgl. Schriver), müsste aus einem Abgleich des jeweils separaten Bildverständnisses und des Textverständnisses ableitbar sein. Die Fälle, die Schriver vor Augen hat, sind aber Fälle, in denen wir zunächst nicht verstehen, wie Text und Bild zusammenhängen und erst eine wechselseitige Analyse uns zu einem Verständnis verhilft. Dazu ein Beispiel: Die taz hat auf ihrer Titelseite vom 5. September 2008 ein Foto von Cem Özdemir mit folgendem Überschriften-Text betitelt:

Das Öbamale

Der „anatolische Schwabe“ Cem Özdemir wird Grünen-Chef SEITE 3, 14

Wer den sprachlichen Ausdruck „Das Öbamale“ nicht versteht, sieht weder einen Gegensatz noch eine andere Relation zwischen Text und Bild, sondern zunächst gar keinen Zusammenhang. Nur vor dem Hintergrund des Wissens über die Rolle von Barack Obama für die Zukunft Amerikas kann geschlossen werden, dass der

„anatolische Schwabe“ Özdemir damit gemeint ist. Offen bleibt dabei ob das „Ö“ auf „Özdemir“ oder auf „Öko“ anspielt. Als Fazit lässt sich festhalten: die sehr allgemeinen Relationen von Redundanz, Komplementarität, Supplementarität, oder Kontrast lassen sich erst auf der Basis grundlegenderer Zusammenhänge feststellen. Sie sind bereits Generalisierungen eines Verständnisses und nicht dessen Erklärung. Insofern sind sie auch nicht dafür geeignet, ein Verständnis des Zusammenhangs zweier Kommunikationselemente zu begründen. Relationsgrammatische Theorien setzen in ihren Beschreibungen von Zusammenhängen zwischen Elementen eines multimodalen Clusters das Problem des Verstehens bereits als geklärt voraus. Die für eine theoretische Klärung erforderlichen Aspekte wie der jeweilige Wissenstand, oder die interne Handlungsstruktur eines komplexen Kommunikationsangebotes werden für eine Klärung des Problems nicht systematisch herangezogen. Auch die Theorien multimodalen Verstehens, die auf der Rhetorical Structure Theory aufbauen, weisen diese Defizite auf und führen dementsprechend zu keiner befriedigenden Erklärung des Verstehensproblems. Diese Ansätze erben gewissermaßen die Schwäche der RST, die Organisation von Texten befriedigend im Rahmen einer Verstehenstheorie zu klären (vgl. dazu Fritz 2008).

Das grammatikalische Modell der multimodalen Relationen liefert auch keine Kriterien dafür, welche Zusammenhänge überhaupt relevant sind. Nicht alle Elemente einer Zeitungsseite, einer Website, eines Fernsehbildschirms müssen in jeder Hinsicht untereinander zusammenhängen. Zu Lösung dieses Problem sieht Bateman „the need for structure“ (2008: 145), also die Notwendigkeit eines übergeordneten Rahmens, innerhalb dessen geklärt werden kann, welche Elemente wofür relevant und untereinander kohärent sind. Erst auf der Grundlage eines solchen „Frames“, den Bateman auf der Basis einer Genre-Theorie klären will, kann bestimmt werden, welchen funktionalen Beitrag ein Element zum intendierten kommunikativen Zweck des gesamten Beitrags leistet (vgl. Bateman 2008: Kap. 5; Bateman et al. 2007).

Ein zweiter Einwand betrifft den bereits angesprochenen Status, den die Relationen für das Verständnis eines multimodalen Angebotes spielen. Dass sie jeweils eine tiefere Schicht der Interpretation voraussetzen, wird deutlich, wenn man ihre Erklärungen betrachtet. So klärt beispielsweise Halliday die Relation der *Extension* folgendermaßen: „adding some new element, giving an exception to it, or offering an alternative“ (Halliday 2004: 378). Die *Ausarbeitung (Elaboration)* wird folgendermaßen erklärt: „restating in other words, specifying in greater detail, commenting or exemplifying“ (ebd.).

Wie die Erklärungen zeigen, ist die tiefere Schicht der Interpretation eine Zuordnung der jeweiligen Äußerung, oder des jeweiligen Elements eines multimodalen Clusters zu einem Handlungsmuster, hier der Handlungsmuster ZUSATZINFORMATIONEN GEBEN, EINE AUSNAHME ODER EINE ALTERNATIVE ANFÜHREN, UMFORMULIEREN, SPEZIFIZIEREN oder KOMMENTIEREN. Die Basis des Verständnisses von Zusammenhängen müsste demzufolge also im Rahmen einer Theorie des sprachlichen Handelns und nicht im Rahmen einer Semiotik, einer Zeichentheorie geklärt werden.

Die Einwände 1 und 2 treffen auch die von Stöckl vorgeschlagene Typologie, die er empirisch aus dem Vorkommen verschiedener Text-Bild-Konstellationen herleitet (Stöckl 2004: Kap. 5.6.). Sieht man einmal davon ab, dass die von ihm vorgelegte Relationsliste – z.B. „Parallelisierung von Sprache und Bild“, „metonymische Konzeptassoziation“, „Symbolisierung“ oder „Metaphorisierung“ – nicht analytisch, sondern synthetisch entstanden ist, so zeigen seine Beschreibungen dieser Relationen ebenfalls, dass sie nicht in einer Theorie fundiert, sondern abgeleitet sind. So wird beispielsweise die Text-Bild-Relation „Symbolisierung“ folgendermaßen erklärt: „Mit der bildlichen Darstellung wird die Aussage des sprachlichen Textes generalisiert – das Bild steht also quasi als seine symbolische Zusammenfassung“ (Stöckl 2004: 258). Diese Beschreibung setzt voraus, dass „die Aussage des sprachlichen Textes“ ebenso unabhängig vom Bild ermittelbar ist, wie das, was das Bild unabhängig vom Text symbolisiert. Das von Stöckl dafür herangezogene Beispiel zeigt, dass das gerade nicht möglich ist: erst der Text klärt, als was das Bild zu sehen ist und erst vor diesem Hintergrund kann die Relation zwischen Text und Bild als „Symbolisierungszusammenhang“ rekonstruiert werden. Bei genauerer Betrachtung von Stöckls Relationsbeschreibung wird erkennbar, dass die Text-Bild-Relationen innerhalb der Erklärungen dieser Zusammenhänge als Handlungszusammenhänge beschrieben werden: mit dem Bild wird der Text generalisiert oder zusammengefasst. Auch hier zeigt sich: die Grundlage von Text-Bild-Zusammenhängen oder allgemeiner, von intermodalen Relation ist handlungstheoretischer Art. Eine Systematisierung und Typologisierung dieser Relationen muss an den Handlungsmustern ansetzen, nach denen sie verstanden werden können.

Ein dritter Einwand betrifft die Anwendbarkeit der Relationstypologie auf Elemente der Gestaltung und des Layouts. Deren Besonderheit liegt darin, dass sie sich gegenüber den Modi von Text und Bild parasitär verhalten: Typografie kann ohne Text nicht vorkommen, Farbe nicht ohne Schrift oder Abbildung, und das Layout setzt immer anzuordnende Elemente voraus. Gestaltungsmerkmale haben – so könnte man sagen – keinen propositionalen Gehalt. Dementsprechend scheitert eine relationsgrammatische Analyse solcher Einheiten daran, dass ihnen unabhängig vom jeweiligen Kommunikationsbeitrag keine Bedeutung zukommt, die sie aber zweifellos haben. Eine Erklärung des kommunikativen Sinns des Designs muss deshalb grundlegend anders ansetzen als es der relationstypologische Ansatz vermag. Es wird später zu zeigen sein, dass eine handlungstheoretische Fundierung des Designs von größerer Reichweite ist.

Als Fazit lässt sich für die relationsgrammatikalischen Ansätze festhalten: Relationen lassen sich nicht aus den einzelnen Elementen ableiten, zwischen denen sie bestehen sollen. Wie die Beispiele gezeigt haben, kontextualisieren sich die Elemente gegenseitig. Die Zusammenhänge zwischen ihnen lassen sich verstehen als abgeleitet von Sequenzen sprachlicher Handlungen. Dies kommt auch darin zum Ausdruck, dass sich die Erläuterungen der Relationstypen lesen wie Listen für solche Handlungsmuster. Voraussetzung für eine theoretisch befriedigende Lösung des Problems intermodaler Zusammenhänge ist es, diese nicht als Funktionen der jeweiligen Elemente zu sehen, sondern als Funktion des Verste-

hens, wie Elemente unterschiedlicher Modi zusammenhängen. Dass Platzierung und Aufmachung aus einem Text einen Aufmacher machen, ergibt sich nicht aus einer Art Addition der Bedeutung des Designs und des Beitragstextes unter Annahme einer spezifischen Relation zwischen den beiden Elementen. „Meaning is multiplicative rather than additive“ heißt es bei Baldry / Thibault (2005: 21). Auf der Grundlage einer handlungstheoretischen Semantik ist das Design eine Form des sozialen Handelns, bei dem Auszeichnungselemente wie Platzierung, Typographie, Farbe, Linien etc. verwendet werden, um Beiträge auszuzeichnen und anzuordnen. „Design is making sense of things“, wie es bei Krippendorf (2006: Introduction) heißt.

2.2. *Soziale Semiotik: die kompositionelle Funktion des Kommunikations-Designs*

Eine der derzeit einflussreichsten Theorie zur multimodalen Kommunikation ist die soziale Semiotik („Social Semiotics“). Ausgangspunkt dieses Ansatzes ist die bereits im ersten Kapitel eingeführte Diagnose „that language is no longer the central semiotic mode“ (Kress 1998: 73). Multimodalität wird definiert als der Gebrauch und die Kombination verschiedener semiotischer Modi – Sprache, Design, Fotos, Film, Farbe etc. –, wobei die verschiedenen Modi sich gegenseitig verstärken oder ergänzen können oder aber hierarchisch geordnet sind. Kress/van Leeuwen (2001) gehen dementsprechend davon aus, dass Kommunikation vielschichtig („multi-layered“) ist und dass demzufolge „meaning is made in many different ways, always, in the many different modes and media which are co-present in a communicational ensemble“ (Kress/van Leeuwen 2001: 111). Erklärtes Ziel der sozialen Semiotik ist eine übergreifende Grammatik für die verschiedenen Modi, die sowohl Bedeutung und Sinn dieser vielschichtigen kommunikativen Praktiken als auch das jeweiligen Zusammenwirken der einzelnen Modi rekonstruierbar macht. „We need a term that can encompass oil painting as well as magazine-layout, the comic strip as well as scientific diagram“ (Kress/van Leeuwen 1996: 3). Der Begriff des Design („visual design“) stellt diese übergeordnete Klammer dar, die alle multimodalen Erscheinungsformen verbindet.

Die Klärung des Design-Begriffs im Rahmen der sozialen Semiotik ist deshalb so zentral, weil das Design für die Umsetzung einer der drei Metafunktionen multimodaler Kommunikationsangebote zuständig ist, nämlich der textuellen Organisation. Analog zu den von Halliday (2004: 29-31) unterschiedenen Grundfunktionen der Sprache („metafunctions“ oder „modes of meaning“) gehen Kress und van Leeuwen davon aus, dass auch multimodale Kommunikationsbeiträge eine *repräsentative* („ideational“), eine *interpersonelle* und eine *textuelle* Funktion haben (Kress / Leeuwen 1996: 40/1, vgl. ebenso Baldry 2005). Um als vollständiges System der Kommunikation zu fungieren, müssen sich multimodale Kommunikationsangebote auf die – vom Sprecher erfahrene – Welt beziehen, d.h. Objekte und ihre Relationen außerhalb des Zeichensystems repräsentieren, eine soziale Beziehung zwischen einem Produzenten und eine Rezipienten her-

stellen und zur Organisation und Kohärenz eines Kommunikationsbeitrags oder Zeichenkomplexes beitragen. Wie im Falle von Texten die Kohärenz- und Kohäsionsmittel die Struktur und Aufbau der Äußerungen organisieren, so ist das Design im Falle von multimodalen Äußerungen die Ressource, die für deren Komposition zuständig ist.

Nach Kress und van Leeuwen besitzt das Design drei bedeutungschaffende Potentiale, um seine kompositionelle Funktion zu erfüllen: es kann durch Platzierung Informationswerte ausdrücken, Elemente hervorheben („salience“) und Elemente kombinieren („framing“) wobei die letzteren beiden Funktionen graduierbar sind (Kress / Leeuwen 1996: 181-186). Es sind also diese drei Prinzipien, die im Rahmen eines semiotischen Ansatzes die Kompositionalität multimodaler Angebote erklären sollen.

Die Ausformulierung einer Grammatik dieser Kompositionsprinzipien, also die Systematisierung der Mittel, mit denen die Funktionen umgesetzt werden können, hat allerdings normative Züge. Der Grund dafür liegt im Ansatz der sozialen Semiotik begründet, demzufolge kompositionelle Entscheidungen regelhaft bestimmte (ideologische) Einstellungen des Autors repräsentieren. So soll die Position links oben auf einer Seite das Bekannte („given“) und das Ideelle („ideal“) ausdrücken, während eine Platzierung links unten anzeigt, dass der Beitrag Neues und Reales ausdrückt (Kress/van Leeuwen, 1996: 223f.; Bateman 2008: 44f.). Empirisch lässt sich leicht zeigen, dass solche Zusammenhänge nicht bestehen und die Deutung von Designaspekten nicht aus einer Grammatik der visuellen Mittel, sondern aus einer sorgfältigen Interpretation der entsprechenden Kommunikation abgeleitet werden kann (vgl. dazu auch Bateman 2008: 44-53). Die Anwendung auf eine Analyse des Layouts von Tageszeitungen (vgl. Kress / van Leeuwen 1998) macht deutlich, dass die Designauffassung von Kress und van Leeuwen repräsentational und nicht kommunikativ-funktional ist. In einer Passage zu den verschiedenen Platzierungstypen – links, rechts, oben, unten, zentral, peripher – heißt es dementsprechend: „Each zone accords („verleiht“) specific values to the elements placed within it“ (Kress / van Leeuwen 1998: 188). Natürlich verleihen nicht die Zonen dem Beitragselement einen entsprechenden Wert, sondern der Kommunikator tut dies – möglicherweise indem er den Beitrag in entsprechender Weise platziert. Mittels der Platzierung kann er versuchen, die intendierte Wertzuweisung an den Rezipienten zu übermitteln. Voraussetzung dafür ist, dass beide die Zonenbelegungen nach denselben Regeln verstehen, d.h. die Zonenbelegungen konventionalisiert sind.

Hier lassen sich Einwände empirischer Art gegen die Semiotik der Platzierung von Kress / van Leeuwen vorbringen. So stimmt es für deutsche Tageszeitungen eben nicht, dass „the elements on the left are presented as Given, and the elements on the right as New“ (Kress / van Leeuwen 1998: 189; vgl. auch van Leeuwen 2005: Kap. 10). Die meisten deutschen Tageszeitungen haben die Meldungsleiste mit den neuesten Informationen in der linken Spalte der Zeitungsseite platziert. In der FAZ steht beispielsweise „in der ersten Spalte links die so genannte Spitzenmeldung“ (FAZ: Alles über die Zeitung 1994: 17), also nicht das Bekannte, sondern das Top-Aktuelle. Gibt es Kommentarspalten auf der Titelsei-

te, so sind diese in der Regel auf der rechten Seite platziert, wo also laut Kress / van Leeuwen das Neue stehen soll. Da Kommentare aber auf bereits Berichtetes Bezug nehmen, entsprechen sie eindeutig dem Bekannten und bereits Eingeführten.

Auch die Interpretation der eigenen Beispiele von Kress und van Leeuwen wirken beliebig, willkürlich und übergeneralisiert. Als Beleg dafür, dass die Oben-Unten-Einteilung einer Zeitungsseite die semiotische Relation von „Ideal“ und „Real“ ausdrückt, verweisen Kress / van Leeuwen auf die Inhaltsleiste am Seitenkopf des Guardian: Sie soll das „Ideale“ ausdrücken, da in ihr die „Essenz“ der Zeitung enthalten ist (vgl. Kress / van Leeuwen 1998: 194f.). Man könnte das genauso gut andersherum sehen, nämlich dass die Inhaltsangabe gerade das Reale, die „more specific information“ oder „more down to earth information“ ausdrückt, zumal sie noch über dem Titelkopf des Guardian steht und gewissermaßen diese Ausgabe der entsprechenden Zeitung konkretisiert. Die semiotischen Interpretationen von Zeitungsseiten lassen auch außer Acht, dass jede Seite im Kontext der Gesamtausgabe gelesen wird. So konstituiert sich der Sinn der Titelseite auch aus ihrer Funktion als Eröffnungs- oder Schaufensterseite für die gesamte Ausgabe (vgl. Gracia 1993: Kap. 6, insb. 149). In diesem Funktionszusammenhang erhalten auch einzelne Elemente der Titelseite ihre je spezifische, seitenübergreifende Bedeutung – beispielsweise die Promoboxen als Inhaltsankündigungen – eine Bedeutungsschicht, die bei isolierten Betrachtung dieser Elemente gar nicht ins Blickfeld geraten kann.

Der Ausgangspunkt der willkürlichen und normativen Bedeutungszuweisung an einzelne Designelemente ist der repräsentationale Charakter einer semiotischen Zeichentheorie, die allen Elementen der Kommunikation feste Bedeutungen zuweisen muss, für die diese Elemente dann stehen. Aspekte der Verwendung der Designelemente wie Kontextfaktoren oder das gemeinsame Wissen der Kommunikationspartner werden für die Erklärung ihrer Bedeutung überhaupt nicht in Betracht gezogen. Dabei spielt gerade das gemeinsame Wissen die entscheidende Rolle, wenn entschieden werden soll, was in der Kommunikation neu und was bereits bekannt ist. Allein aufgrund der Platzierung eines Beitrags lässt sich nicht entscheiden, ob die Information alt oder neu, der Inhalt ideal oder real ist. Die Konsequenz dieser repräsentationalen Zeichentheorie ist eine Hypostasierung der Bedeutung gegenüber dem, was in der Kommunikation mit der Verwendung eines bestimmten Designs gemeint sein könnte.

Die Leistung der sozialen Semiotik liegt darin, die Bandbreite der semiotischen Ressourcen aufgezeigt und dem Design eine herausragende Stellung in der Beschreibung multimodaler Kommunikationsformen zugewiesen zu haben. Man kann die von Kress und van Leeuwen vorgeschlagenen Kompositionsprinzipien der Platzierung, Hervorhebung und Rahmung als produktive Hinweise darauf betrachten, welche Aspekte des Designs für die Sinnkonstitution eines multimodalen Angebotes genutzt werden können. Die Systematisierung der organisierenden Funktionen des Designs zeigt insofern auf, welche Kommunikationsressourcen in diesem Kommunikationsmodus vorliegen. Entscheidend für das Verständnis eines multimodalen Angebotes ist aber, welcher Gebrauch von diesen Designele-

menten gemacht wird. Dieser lässt sich jedoch nicht aus der Form ableiten, sondern erfordert eine kontextorientierte Interpretation, die sowohl die externen Zusammenhänge eines multimodalen Kommunikationsbeitrags als auch seine internen berücksichtigt (ausführlicher dazu Bucher 2007).

2.3. Dynamisch-pragmatische Ansätze: Transkription und Interaktion

Begriffe wie „Sehlesen“ (Schmitz 2007) oder „Mit Worten sehen“ (Holly 2006) aber auch „Intersemiosis“ oder „semantic multiplication“ (Fei 2004) verweisen auf den dynamischen Aspekt der Bedeutungszuweisung für multimodale Kommunikationsangebote: kommunikative Einheiten mit verschiedenen Modi konstituieren wechselseitig den Gesamtsinn eines multimodalen Kommunikates. Diese Art der Zugangsweise zum Problem der Kompositionalität stellt eine Verbindung her zum Problembereich des Verstehens und der Rezeption. Es ist zwar richtig, dass „Verstehen“ kein Verlaufs- sondern ein Erfolgsverb ist, das nicht in Kombination mit Durationsangaben wie „drei Stunden lang“ gebraucht werden kann. Allerdings setzt sich das Verstehen eines multimodalen Angebotes aus verschiedenen Teilverständnissen, vorläufigen Verständnissen, Umdeutungen und Neudeutungen zusammen, die in Kombination einen Prozess des Interpretierens bilden. Auch die Idee des hermeneutischen Zirkels beruht auf einer solchen prozessorientierten Sichtweise des Verstehens. Ulrich Schmitz sieht multimodale Angebote als „Text-Bild-Konglomerate, in denen Text und Bild einander kontextualisieren und monosemieren“ (Schmitz 2004). Im integrativen Modell von Fei wird ein „space of integration“ definiert, für „the dynamics of interaction between language and visual images for meaning making in a multi-semiotic text“ (Fei 2004 oder 2006: 238f.).

Neben dem Begriff der Interaktion, der später noch ausführlich zu klären sein wird, wird in dieser Theoriefamilie seit neuesten der Begriff der „Transkription“ verwendet, um den dynamischen Prozess der Bedeutungszuweisung („meaning-making“) zu erklären. Es lassen sich zwei Verwendungsweisen dieses Begriffs unterscheiden. Mit „Transkription“ kann erstens eine Darstellungstechnik und Form der Analyse bezeichnet werden: „Transcription is a way of revealing both the co-deployment of semiotic resources and their dynamic unfolding in time along textually constrained and enabled pathways or trajectories“ (Baldry / Thibault 2005). Ausgangspunkt für die zweite Verwendungsweise des Begriffs der Transkription sind Fälle, in denen der Sinn einer Kombination von Elementen in verschiedenen Modi – beispielsweise Text und Bild – nur aus ihrem wechselseitigen Bezug verständlich werden. Erst durch die gegenseitige Kontextualisierung können Bedeutungen entstehen, die den jeweils einzelnen Elementen nicht zu entnehmen sind. Ein von Holly verwendetes Beispiel besteht aus einer Filmsequenz, die den neuen thüringischen Ministerpräsidenten Dieter Althaus mit seiner Frau am Frühstückstisch zeigt, und dazu folgenden aus dem Off gesprochenen Text enthält: „jung – dynamisch – Gartenspinne“. Mit dem letzten Ausdruck, wird auf einen entsprechenden Gegenstand Bezug genommen, der in der Früh-

stücksszene beim Blick aus dem Fenster erkennbar ist. Entsprechend der Analyse von Holly, wird der Gegenstand der Gartenspinne erst durch seine entsprechende Erwähnung im Off-Text fokussiert, während der Text mit seiner überraschenden Reihung „jung, dynamisch, Gartenspinne“ ohne das Bild unverständlich bliebe. Text und Bild leisten als audiovisuelle Bedeutungskonstitution eine „soziale Stilisierung“ des neuen Ministerpräsidenten als „ziemlich bieder, wenn nicht sogar spießig“ (Holly 2006: 4), wobei die Wäschespinnne als negatives Statussymbol genutzt wird. Die theoretische Auswertung dieses Beispiels dient der Einführung der Transkriptionstheorie, derzufolge Bedeutungsgenerierung ein Prozess ist, „bei dem ein Präskript zum Skript wird, indem es transkribiert wird“ (Holly 2009: 9).

Die Sichtweise, dass in multimodalen Angeboten die Zusammenhänge zwischen verschiedenen Modi als Transkriptionen zu beschreiben sind, führt zu einer Hypostasierung der kommunikativen Elemente, bei der die Unterscheidung von Bedeutung und Gebrauch kommunikativer Einheiten nicht mehr bedacht wird. So wird bei Holly „das Bild vom Sprechertext transkribiert“, der zudem „dem Bild eine entsprechende Bedeutung zuweist“ (Holly 2006: 5). Und er fragt umgekehrt, „wie die Bilder diesen Sprachtext transkribieren“ (Holly 2007: 404). Die Handlungen des Zuweisens und Transkribierens werden nicht mehr Akteuren zugewiesen, sondern Gegenständen der Kommunikation, die Mittel der Kommunikation werden zum Agens des Vermittlungsprozesses. Damit wird eine Semantik eingeführt, die mit einer verwendungsorientierten und handlungstheoretischen nicht mehr vereinbar ist.

Die Hypostasierung der sprachlichen Elemente folgt aus der Logik des Begriffs „Transkribieren“: Transkribieren setzt voraus, dass ein Gegenstand zum Transkribieren vorhanden ist. Insofern muss der Text oder das Bild bereits Bedeutung haben, bevor es dem Prozess des Transkribierens unterworfen wird, so wie ein aufgezeichneter Dialog gegeben sein muss, bevor er transkribiert werden kann. Gerade diese Annahme aber ist nicht nur bedeutungstheoretisch schädlich, sondern widerspricht auch den empirischen Tatsachen: Verständnisse von komplexen Kommunikationsangeboten entstehen erst aus der wechselseitigen Deutung ihrer Einzelelemente und lassen sich nicht aus diesen ableiten und dann zusammensetzen. Die Fallanalysen von Holly belegen im Übrigen, dass er bei seinen Deutungen kontextorientiert verfährt, so dass man den Begriff der Transkription weglassen kann, ohne dass die Interpretationen dadurch verständlicher oder weniger gehaltvoll wären. Die Bedeutung sprachlicher und nichtsprachlicher Äußerungen folgt aus deren Gebrauch und nicht umgekehrt, wie es eine repräsentationale Semantik annimmt. Wie die mangelnde Unterscheidung von Bedeutung und Gebrauch zeigt, fällt die Transkriptionstheorie aufgrund der Logik des Skriptbegriffs in eine repräsentationale Semantik zurück, wie es bereits für die funktions-grammatikalischen Ansätze gezeigt wurde.

Mit dem Begriff der Implikatur steht ein erprobtes und gut begründetes Erklärungsmodell für die Analyse des erwähnten Beispiels von Holly zu Verfügung: Die scheinbar fehlende Kohärenz des Ausdrucks „Gartenspinne“ im vorliegenden Kommunikationszusammenhang, löst ein Gricesches Raisonement

aus, bei dem rekonstruiert wird, unter welchen Voraussetzungen der Ausdruck relevant, informativ, verständlich und angemessen sein kann. Die Eingabedaten für dieses Raisonement werden dem visuellen Teil des multimodalen Kommunikationsbeitrags – den begleitenden Filmbildern – entnommen, wo ein entsprechender Gegenstand gezeigt wird, so dass das entsprechende Referenzobjekt des sprachlichen Ausdrucks identifiziert werden kann. Die weitere Deutung des Beispiels unterscheidet sich nicht von der Deutung einer Textsequenz zur Charakterisierung von Dieter Althaus, in der es heißen könnte: „Dieter Althaus ist glücklich verheiratet und in seinem Garten steht eine Wäschespinne“.

Das Beispiel weist auch auf zwei Aspekte der Erklärung multimodalen Verstehens hin, die bislang vernachlässigt blieben, die *Dimension des kommunikativen Kontextes* oder *Verwendungszusammenhangs* und der *Aspekt des Wissens*. So ist die Erwähnung und das Zeigen der Gartenspinne zwar relevant im Kontext eines Filmporträts, nicht aber im Kontext eines Nachrichtenfils, in dem die analysierte Sequenz offensichtlich ursprünglich vorkam, bevor sie in einem Jahresrückblick neu betextet wurde. Eine befriedigende und umfassende Erklärung multimodaler Bedeutungszuweisung kann sich also nicht auf die entsprechenden Einzelelemente beschränken, sondern muss deren Einbettung in den vorliegenden Kommunikationszusammenhang mit berücksichtigen. Und nur aufgrund des gemeinsam geteilten Wissens zwischen dem Autor und dem Rezipienten, dass Gartenspinnen Symbole für Spießigkeit sein können, wird der Sinn des multimodalen Kommunikationsbeitrages für den Fernsehzuschauer erkennbar.

Die gegenüber einer Transkriptionstheorie für multimodales Verstehen formulierten Einwände müssen den von Ludwig Jäger (2008) entwickelte Transkriptionsbegriff nicht notwendig tangieren, sondern bezweifeln den Sinn seiner Übertragbarkeit auf die Beschreibung multimodaler Kommunikationsformen. Der Sinn der Jägerschen Transkriptivitätstheorie besteht darin, dass sie ein Phänomen unter eine einheitliche Betrachtungsweise stellt, wie es für Mediengesellschaften typisch ist – das Phänomen der permanenten Übertragung und Weiterverarbeitung von Kommunikation in anderen Medien und Kommunikationskontexten: aus dem Film wird ein Trailer, aus den Quellen eine historische Darstellung, aus dem Lehrbuch ein digitaler Hypertext, aus dem linearen Fernsehprogramm eine non-lineare Mediathek und aus der gedruckten Zeitung eine Online-Zeitung (vgl. dazu Jäger et al. 2008). Auch wenn der Begriff der Transkription für diese Medienphänomene als Metapher verwendet wird, so ist das in diesem Zusammenhang theoretisch unschädlich. Im Rahmen einer Theorie multimodaler Kommunikation führt seine Nutzung jedoch zu fragwürdigen bedeutungstheoretischen Konsequenzen. Da alternative Analyseverfahren verfügbar sind, ist der Transkriptionsbegriff nicht nur theoretisch unproduktiv, sondern auch überflüssig und wird demzufolge ein Opfer des Ockhamschen Rasiermessers: *Entia non sunt multiplicanda sine necessitate*.

Eine Alternative mit theoretisch etablierten Begriffen besteht darin, anknüpfend an die Theorie der Implikaturen, den Begriff der Interaktion als Grundlage einer Verstehentheorie für multimodale Kommunikationsangebote heranzuziehen: Der Prozess des Deutens und Verstehens wird als kontrafaktische Interakti-

on zwischen den Rezipienten und dem entsprechenden Kommunikationsangebot modelliert, wie das in verschiedenen Theorietraditionen implizit geschieht (Fei 2004; Krippendorff 2006, Baldry / Thibault 2005). Wenn die Bedeutung von Kommunikationselementen nicht repräsentational darin gesehen wird, wofür sie stehen, sondern in ihrem Gebrauch in der Kommunikation, so impliziert dies eine interaktionale Sichtweise der Bedeutungszuweisung (vgl. Krippendorff 2006). Im folgenden Abschnitt soll eine interaktive Theorie des multimodalen Verstehens skizziert und mit Beispielen der empirischen Rezeptionsforschung belegt werden. Eine solche interaktionale Theorie des Verstehens und der Rezeption soll auch dazu beitragen, die folgenden Problembereiche der bisherigen Theoriegruppen adäquater zu lösen:

1. Das Problem der Dynamik des multimodalen Verstehens: Multimodale Angebote werden nicht auf einen Schlag verstanden, sondern in einem Deutungsprozess erschlossen, dessen Systematik und innere Struktur im Rahmen einer interaktionalen Theorie rekonstruierbar wird.
2. Das Problem des Status der intra- und intermodalen Relationen: Im Zentrum der diskutierten Ansätzen steht das Problem der Kohärenz zwischen einzelnen kommunikativen Basiseinheiten multimodaler Kommunikationsangebote. Der Status dieser Relationen bleibt aber ungeklärt. Im Rahmen der funktionsgrammatischen Theorien sowie der Transkriptionstheorie werden die Relationen nach dem Modell einer repräsentationalen Semantik ontologisiert. Die Soziale Semiotik sieht die Relationen in einer normativen Deutung des Designs ausgedrückt.
3. Das Problem der diskontinuierlichen Struktur non-linearer Kommunikationsformen: Diese Kommunikationsformen zeichnen sich dadurch aus, dass Einzelelemente von den Rezipienten in unterschiedliche Zusammenhänge mit unterschiedlichen Partnerelementen gestellt werden können. Dementsprechend müssen Elemente, die in einen Sinnzusammenhang gebracht werden auch nicht räumlich benachbart sein. Für diese Diskontinuität des Verstehens multimodaler Angebote bieten die beschriebenen Ansätze keine befriedigende Lösung, da sie entweder auf Paarsequenzen beschränkt bleiben oder aber normative Lesepfade vorausgesetzt werden (Bateman 2008:158).

3. Rezeption als Interaktion

Multimodalität kann verstanden werden als Sonderform non-linearer Kommunikation. Multimodales Verstehen zu erklären und damit das Problem der Kompositionalität aufzulösen, setzt deshalb voraus, das Verstehen non-linearer Kommunikation zu erklären. Die Nonlinearität der Kommunikation beginnt im Grunde genommen mit dem Übergang von der Mündlichkeit zur Schriftlichkeit, durch die die zeitliche Dimension der Kommunikation um eine räumliche erweitert wurde. Die Anordnung von Zeichen im Raum war bereits durch alle Formen der handschriftlichen Fixierungen gegeben, aber erst durch die Erfindung des Drucks

mit beweglichen Zeichen wurden Anordnungen im Raum reproduzierbar. Ong beschreibt diesen Übergang von der Handschrift zum Druck folgendermaßen:

Writing moves the sound world to a world of visual space, but print locks words into a position in this space. [...] Print develops far more sophisticated use of space for visual organisation and for effective retrieval. [...] Because print controlled not only what words were put down to form a text but also the exact situation of the words on a page and their spatial relationship to one another, the space itself on a printed sheet took on high significance. (Ong 2002: 119, 122, 126)

Mit dieser Transformation der Kommunikation in die Dimension des Raumes wird die Kohärenz der Kommunikation nicht mehr alleine durch eine sequentielle Ordnung sichergestellt sondern ebenso durch eine räumliche Anordnung („Komposition“). Die temporale Logik linearer Kommunikation wird durch eine räumliche Logik ersetzt, in der die Zusammenhänge zwischen Elementen nicht mehr eindimensional fortlaufend, sondern zweidimensional in der Fläche bestehen. Ein multimodales Angebot zu verstehen heißt deshalb immer auch, zu erkennen, wie ein simultan präsentiertes Kommunikationsangebot aus mehreren kommunikativen Elementen räumlich strukturiert ist: welche Elemente gehören enger zusammen? Welche Zusammenhänge bestehen zwischen ihnen? Wie sind sie hierarchisiert? Für die Beantwortung dieser Fragen wurde verschiedentlich auf die Prinzipien der Gestalttheorie als Mechanismus zur Reduktion von Komplexität zurückgegriffen (Shriver 1996; Smith 2005). Die Idee, multimodales Verstehen in einem Schichtmodell zu modellieren und gegenüber linearen Texten zusätzliche Erschließungsebenen einzuführen – die Layoutstruktur und die Navigationsstruktur (Bateman 2008: 19), das Design (Kress / van Leeuwen 2001: 5f. und Kap. 3) oder das Textdesign (Bucher 2007: 58-67) – ist auf diesem Hintergrund naheliegend. Multimodale Formen der Kommunikation haben aus dieser Perspektive hypertext-ähnliche Strukturen und werfen dementsprechend analoge Probleme des Verstehens auf (vgl. Bucher 1998).

Baldry / Thibault haben darauf hingewiesen, dass die Dimension des Raumes und der damit verbundene Aspekt der Nonlinearität auch auf audiovisuelle Filmbeiträge übertragbar ist: Der Film wird als Text betrachtet, der aus einzelnen Phasen („Phases“) zusammengesetzt ist, die ihrerseits eine non-lineare, multimodale Struktur aufweisen. Filmbeiträge weisen dementsprechend beide Strukturmerkmale gleichermaßen auf: sie sind linear strukturiert in der Abfolge von Phasen und non-linear in der multimodalen Konzeption der Phase selbst, die aus verschiedenen visuellen, textlichen, lautlichen, musikalischen, sprachlichen Elementen bestehen kann. „A phase may be defined as a set of co-patterned semiotic selections that are co-deployed in a consistent way over a give stretch of text“ (Baldry / Thibault 2005: 47). Die folgenden Ausführungen zum multimodalen Verstehen sind dementsprechend auch auf die Analyse von Filmsequenzen anwendbar, auch wenn die empirischen Belege aus Print- und Online-Medien stammen.

Eine der Besonderheiten nonlinearer Kommunikation besteht darin, dass Anordnung und Zusammenhang der Elemente in vielen Fällen diskontinuierlich ist. Batman hat dieses Problem erkannt und schlägt eine entsprechende Erweiterung der für lineare, sequentielle Kommunikationsformen entwickelten Rhetorical Structure Theory (RST) vor: „we restrict RST relations to pairs of document parts, which are adjacent in any direction. (...) Spatial contiguity is present across area model cells when they are adjacent and within cells when elements are spatial neighbours” (Bateman 2008: 158). Dieser Vorschlag, das Sequenzprinzip durch ein räumliches Nachbarschaftsprinzip zu erweitern, greift allerdings aufgrund der Diskontinuität multimodaler Kommunikationsformen zu kurz. Erstens gibt es, gerade auf Titelseiten von Zeitungen, Elemente – z.B. Teaser und Promos –, deren Funktion gerade darin besteht, auf nicht-benachbarte Elemente zu verweisen, die meistens sogar auf anderen Seiten zu finden sind. Zweitens können Leser und Zuschauer Zusammenhänge zwischen nicht benachbarten Elementen herstellen, beispielsweise unter einem thematischen Gesichtspunkt, hinsichtlich Textsortenidentität oder auf der Grundlage optischer Hervorhebungen. Gerade die diversen Blickaufzeichnungsstudien zeigen, wie diskontinuierlich ein Rezeptionsverlauf sein kann. (vgl. Holsanova et al. 2006; Bucher / Schumacher 2007; Stark / Quinn 2007). Ein dritter Fall diskontinuierlicher Rezeption liegt dann vor, wenn dasselbe Element von verschiedenen anderen Elementen in jeweils unterschiedlichen Zusammenhängen rezipiert wird. Das kann der Fall sein, wenn eine Abbildung oder eine Grafik von verschiedenen Stellen eines Textes aus betrachtet wird, und so unterschiedliche Ebenen der Erschließung dieser Abbildung erarbeitet werden. Ein vierter Fall von Diskontinuität liegt vor im Falle der Tickermeldungen auf dem Fernsehschirm, die mit dem audiovisuellen Kernangebot in der Regel in keinem Zusammenhang stehen und dementsprechend auch separat rezipiert werden können.

Nutzt man den Begriff der Non-Linearität zur Klärung des Begriffs der Multimodalität, so werden zwei Strukturebenen des multimodalen Verstehens erkennbar, die den für Hypertexte unterschiedenen Aneignungsebenen entsprechen: die operationale Ebene und die Kommunikationsebene

1. Die Ebene der *Identifizierung und Gruppierung bedeutungstragender Elemente* im Kommunikationsraum. Diese Dimension kann aus der Sicht einer Rezeptionstheorie auch als *Selektionsproblem* beschrieben werden.
2. Die Ebene der *Kohärenz* und der Erschließungspfade im Kommunikationsraum. Auf dieser Ebene ist die *Dynamik des Erschließungsprozesses* anzusiedeln.

Diese beiden Strukturebenen des multimodalen Verstehens werden in den folgenden Abschnitten näher ausgeleuchtet und mit empirischen Befunden unterlegt. Theoretischer Grundgedanke ist dabei, dass die Rezeption eines multimodalen Angebotes nach dem Modell eines Interaktionsprozesses verstanden werden kann: Der Leser, Zuschauer, Nutzer interagiert mit dem Angebot, indem er es Schritt für Schritt auf der Basis seines Vorwissens, seiner Absichten, seiner An-

nahmen und seines aktualisierten Wissensbestandes erschließt. Dadurch wird berücksichtigt, dass multimodale Angebote nicht auf einen Schlag erfasst, sondern in einem Prozess erschlossen werden. Diese interaktionale Sichtweise der Rezeption erlaubt es dementsprechend, einen dynamischen Kontextbegriff zugrunde zulegen, wie ihn van Dijk vorschlägt: „It is not the social situation that influences discourse, but the way the participants define such a situation“ (van Dijk 2008: X). In einem Erschließungsprozess ist diese Situationsdefinition zusätzlich dynamisiert und verändert sich von Schritt zu Schritt mit der Verschiebung der Aufmerksamkeit und der Veränderung des jeweiligen Wissensstandes.

3.1. Selektion und Interaktion

In ihrem Beitrag „Understanding is identifying a structure“ hat Paulina Lindström auf die enge Beziehung zwischen Verstehen und Problemlösen hingewiesen, die darauf basiert, dass in beiden Fällen relevante Zusammenhänge erkannt werden müssen (Lindström 2006: 2). Betrachtet man multimodale Kommunikationsformen als non-lineare Formen der Kommunikation, so kann man das Problem, relevante Zusammenhänge zu erkennen, in Teilprobleme zerlegen, wie sie auch für die Erschließung von Hypertexten typisch sind (Bucher 2004b: Kap. 3):

1. *Das Identifizierungs- oder Lokalisierungsproblem:* Wer ist der Kommunikationspartner oder Autor des multimodalen Angebotes? Mit welcher Art von Kommunikationsangebot habe ich es zu tun?
Typische Mittel: Senderkennungen, Logos, Titel, aber auch Signalfarben oder eine bestimmte Typografie.
2. *Das Orientierungsproblem:* Was ist der aktuelle Kommunikationsstand? An welcher Stelle im Kommunikationsangebot befinde ich mich?
Typische Mittel: Inhaltsverzeichnisse, Fortsetzungsindikatoren, Sitemaps, Zeit- und Sendungseinblendungen, Ressorttitel, Seitenzahlen, Logos, Farbleitsysteme, Trennlinien und -farben.
3. *Das Hierarchisierungsproblem:* Welche Unter-, Über- und Nebenordnungen bestehen zwischen den verschiedenen Elementen der Kommunikation?
Typische Mittel: Typografie, Platzierung, Bebilderung, druckgrafische Auszeichnungen, Weißraum, Textsorten- und Rubrikenbezeichnungen.
4. *Das Navigationsproblem:* Wie gelange ich von einem Element zu einem damit zusammenhängenden anderen? Wo liegen die übergangsrelevanten Stellen?
Typische Mittel: Verweise und Seitenhinweise, Links, Programmhinweise, Navigationsleisten, Links, Inhaltsverzeichnisse, Sitemaps, Suchhilfen.
5. *Das Rahmungsproblem:* Was gehört – enger oder weiter – zusammen? Was steht im Vordergrund. Was im Hintergrund? Was ist aktuell relevant, was ist latent?
Typische Mittel: Seitenlayout, Platzierung, Rubriken, Übergangs- und Abtrennungsmarkieren (optisch, verbal, oder akustisch).

6. *Das Sequenzierungs- und Einordnungsproblem:* Wie hängen die verschiedenen – als relevant identifizierten – Elemente einer Kommunikation untereinander zusammen (funktional, strategisch, thematisch).

Typische Mittel: Links mit Tags, Clustervorspanntext, Textsortenangaben, explizite deiktische (sprachliche oder gestische) Verweise auf weiterführende, simultane oder vergangene Beiträge, Seiten- und Beitragsteile.

Die Problemliste ist so formuliert, dass sie auf unterschiedliche multimodale Kommunikationsformen wie Fernsehsendungen, Zeitungs- oder Zeitschriften(seiten), Online-Angebote, Lehrbücher aber auch auf Vorträge mit Präsentationen anwendbar ist. Voraussetzung für die Lösung dieser Probleme ist die *Selektion* der relevanten Kommunikationselemente. So muss für die Lösung des Identifikationsproblems die entsprechende Anbieterkennzeichnung – etwa der Zeitungstitel oder das Anbieterlogo – gefunden werden, für die Lösung des Identifikationsproblems die entsprechenden Links, für die Lösung des Einordnungsproblems die relevanten Kommunikationsinhalte und für die Lösung des Rahmungsproblems die relevanten Merkmale für die Zuordnung der Elemente zu einem entsprechenden Cluster. Blickaufzeichnungsdaten zeigen, dass die Lösung dieser Aufgaben in unterschiedlichen Phasen der Aneignung geschieht: So werden das Identifikationsproblem und das Orientierungsproblem bereits in einer frühen Phase aufgegriffen, während das Rahmungs- und das Sequenzierungsproblem erst später angegangen werden.

Die Erklärungsansätze für die Selektion lassen sich in die beiden bereits eingangs genannten Modelle einteilen: die Saliencetheorien, die davon ausgehen, dass die Selektion vom Angebot selbst gesteuert wird – beispielsweise durch optisch auffallende Elemente (Jarvenpaa 1990); und die Schematheorien, die die Selektion entweder als intentionale Handlung des Rezipienten oder als Resultat seiner kognitiven Schemata betrachten (Schnotz / Bannert 2003). Im Falle der Angebotssteuerung wird von einem Bottom-up-Prozess gesprochen, in dem das Angebot selbst die Relevanz der entsprechenden Elemente signalisiert, im Falle der Nutzersteuerung von einem Top-down-Prozess, bei dem der Rezipient darüber entscheidet, was für ihn relevant ist. Linguistische Theorien der Multimodalität, wie sie im Kapitel 2 beschrieben sind, basieren in der Regel auf einer stillschweigend vorausgesetzten Saliencetheorie: in den funktionsgrammatischen, den semiotischen und den dynamischen Theorien der Multimodalität werden Angebotsstrukturen beschrieben und von diesen auf das Verstehen und die Interpretationsprozesse geschlossen. Das wird deutlich an der Annahme bei Kress/van Leeuwen, dass der Platzierung eines Elements im Kommunikationsraum eine ideologische Bedeutung zukommt und sie eine entsprechende Sichtweise – Bekannt / Neu oder Zentral/ Marginal – auf das Dargestellte ausdrücken soll. Von der Präsentationsform wird somit direkt auf das Verständnis geschlossen, ohne den dynamischen Aspekt des Verstehens zu berücksichtigen (Kress / Leeuwen 1996: Kap. 6, insb. 187). Dasselbe Denkmuster findet sich in den Ansätzen, die den Transkriptionsbegriff zur Erklärung multimodalen Verstehens heranziehen.

Der Zusammenhang von Stimulus, Wahrnehmung, Selektion und Verstehen ist allerdings komplexerer Art. Eine der intervenierenden Variablen ist die Aufmerksamkeit: Die Klärung des Problems der Selektion setzt eine Klärung des Aufmerksamkeitsproblems voraus (Neumann 1992; Neumann 1996). Aufmerksamkeit ist selbst ein mehrschichtiger Begriff, der zumindest zwei Verwendungsweisen hat: man kann Aufmerksamkeit (intentional) schenken und Aufmerksamkeit kann (nicht-intentional) geweckt werden (Bucher/Schumacher 2006). Moderne Aufmerksamkeitstheorien zeichnen sich dadurch aus, dass sie eine intentionale und eine nicht-intentionale Form der Aufmerksamkeit annehmen, die jeweils die Extreme auf einem Kontinuum von Mischformen sind (Neumann 1992: 88). „Visual attention cannot simply be explained through the sole consideration of visual features. There are higher-level intentional factors involved” (Duchowski 2003: 5). Der Begriff der Aufmerksamkeit legt insofern selbst eine Integration von Bottom-up und Top-down-Prozessen nahe.

Diese differenzierte Sichtweise des multimodalen Verstehens wird auch durch empirische Befunde der Trierer Blickaufzeichnungen bestätigt (Bucher / Schumacher 2006; Bucher et al. 2007). Sowohl die Angebots- als auch die Nutzersteuerung der Selektion ist nachweisbar. So zeigen Laborexperimente, bei denen die Probanden Seiten in Online-Angeboten mit unterschiedlichen Aufgaben betrachten sollen, auch unterschiedliche Selektionsmuster. Während bei der Aufgabe, die Seite im so genannten „activity mode“ ohne genaue Vorgaben zu explorieren, zu einer breiten und unspezifischen Distribution der Aufmerksamkeit führt, verhält es sich im sogenannten „goal mode“ mit genauen Zielvorgaben ganz anders: es werden diejenigen Teile des Angebotes fokussiert, die für die Lösung der Aufgabe relevant sind. Die Intention und die Kompetenz zu ihrer Umsetzung liefern also in diesem Fall die entscheidenden Kriterien, was relevant ist und was nicht. Das multimodale Angebot hat deshalb keine feststehende „bottom-up“ Bedeutung, sondern diese wird dem Angebot vom Rezipienten im Rahmen seiner Aneignungsabsichten „top-down“ zugewiesen. Auch die Tatsache, dass verschiedene Rezipienten dasselbe Angebot unterschiedlich explorieren, macht deutlich, dass der „page-flow“ (Bateman 2008: 159) nicht – oder zumindest nicht ausschließlich – von der Struktur des Angebotes gesteuert wird. Vergleicht man die Blickaufzeichnungsdaten verschiedener Probanden, so wird deutlich, dass man verschiedene Selektionsmuster unterscheiden kann, die offensichtlich auf verschiedenen Erschließungsstrategien beruhen. Im Falle der Erschließung der Einstiegsseite von BILD-online in Abbildung 1 sind ganz deutlich zwei Strategien zu unterscheiden: eine inhaltsorientierte Strategie, die hauptsächlich Überschriften und Orientierungstexte auswählt, und eine strukturorientierte Strategie, die schwerpunktmäßig Navigationselemente, Links, und Angebotskennzeichnungen fokussiert.

Kulturelle Unterschiede in der Rezeption derselben Angebote belegen ebenfalls, dass die Selektion der Elemente und Aspekte, die relevant für das Verstehen sind, von den jeweiligen Akteuren und ihren kulturspezifischen Schemata abhängen (vgl. Bucher 2004a). Die Steuerung durch das Angebot wird erkenn-

bar, wenn Angebotsmerkmale modifiziert werden und sich dadurch die in den Blickverläufen manifestierten Aneignungsprozesse ändern.



Abb. 1: Zwei unterschiedliche Erschließungsstrategien der Einstiegseite von BILD-Online vom 15.07.2005: die inhaltsorientierte Strategie und die strukturorientierte Strategie. (Das entsprechende Video befindet sich wie auch sämtliche weiteren Abbildungen (in Farbe) auf der beiliegenden CD)

Holsanova et al. (2008) zeigen am Beispiel systematischer Modifikationen von Text-Grafik-Clustern, dass das Clusterdesign die Rezeption von Text und Grafik erheblich beeinflusst: eine in den Text integrierte Sequenz von Grafiken führt zu einer wechselseitigen Rezeption von Text und Bild, wobei allerdings die Grafiken im Durchschnitt weit weniger Betrachtungszeit auf sich ziehen können, als in einer separierten Präsentation. Werden Text und Grafik-Sequenz in zwei räumlich getrennten Einheiten platziert, so werden sie auch als zwei separate Elemente zeitversetzt rezipiert, wobei der Text kaum noch gelesen wird. Die integrierte Präsentation führt auch zu einer integrierten Rezeption, bei der eine zeitliche Splitting der Aufmerksamkeit für die beiden Elemente vermieden wird. Neben diesem Nachbarschaftsprinzip (Contiguity-Principle) belegen Holsanova et al die Wirksamkeit eines zweiten Prinzips, das ebenfalls die Steuerungsleistung des Angebotes hervorhebt – das sogenannte „dual scripting principle“ (Prinzip der doppelten Auszeichnung): eine konzentrisch angeordnete Grafiksequenz wird in eine serielle umgewandelt, wobei die Grafiken in eine logische Abfolge vom

Allgemeinen zum Besonderen gebracht werden. Ein Vergleich der Blickdaten zeigt, dass in der seriellen und logischen Anordnung die Probanden die Grafiken auch sequentiell rezipieren, sie länger betrachten und die Lektüre des Grafiktextes stärker in den Rezeptionsprozess integrieren.

Ein anderer Typus von Steuerungsleistung des medialen Stimulus wird von Folker/Sichelschmidt (2005) nachgewiesen: Sie zeigen in einem vergleichenden Experiment, dass die Hervorhebung von Begriffen in einem Text durch Farbmarkierung, die sich mit den entsprechenden Grafikbeschriftungen deckt, die Rezeptionszeit eines multimodalen Clusters verkürzt, die Anzahl der Blickwechsel zwischen Text und Grafik erhöht und damit offensichtlich multimodales Verstehen erleichtert. (Folker / Sichelschmidt 2005).

Als Fazit aus diesen experimentellen Befunden mit Intentions- und Akteursvarianten einerseits und Angebotsvarianten andererseits lässt sich festhalten: Blickaufzeichnungsdaten belegen empirisch, dass zur Erklärung der Aufmerksamkeitsverteilung und damit des multimodalen Verstehens Top-Down und Bottom-up-Prozesse, Angebots- und Nutzersteuerung, Intentionalität und Reaktivität gleichermaßen herangezogen werden müssen. Damit ist nachgewiesen, dass zur Erklärung des multimodalen Verstehens eine Theorie erforderlich ist, die zwei Kriterien erfüllt: sie muss die beiden Seiten der medialen Aneignung integrieren und sie muss dynamisch sein, um den Aufmerksamkeitswechsel, die Rekursivität des Verstehens zu erklären. Eine solche Theorie kann als interaktiv bezeichnet werden. Eine interaktive Theorie der Aneignung kann auch die alte Streitfrage zwischen Sensualisten und Realisten einerseits und den Idealisten oder Kognitivisten andererseits lösen: beide Seiten betrachten nur jeweils die eine Hälfte des interaktiven Aneignungsprozesses.

3.2. Interpretation als Interaktion

Es charakterisiert die Ansätze zum multimodalen Verstehen, die auf den Transkriptionsbegriff zurückgreifen, dass sie die Dynamik des Aneignungsprozesses hervorheben, die von den zeichentheoretisch ausgerichteten Theorien gerade nicht erfasst wird. Die Dynamik wird darin gesehen, dass sich modal unterschiedliche Elemente wie Text und Bild gegenseitig „transkribieren“. So beschreibt Holly die „Audiovisualität als Dynamik“ nach dem Muster eines „Wort-Bild-Reißverschlusses“: „Produktiv wie rezeptiv geht die Bedeutungsgenerierung oszillierend zwischen Sprach- und Bildseite hin und her, den fluiden Parcours der Sprachlaute und Filmbilder mäandernd durchlaufend“. Die beiden Zeichenarten des „audiovisuellen Bandes“ – Text und Bild – sollen dabei „wie bei einem Reißverschluss“ wechselseitig ineinander greifen (Holly 2009: 12). Beschreibungen dieser Art rufen unweigerlich die Diagnose Wittgensteins ins Gedächtnis, dass „philosophische Probleme entstehen, wenn die Sprache feiert“ (Wittgenstein 1977: § 38). Dass die Beschreibung multimodalen Verstehens derart poetisch und metaphorisch ausfällt, liegt bereits in der Metaphorik des Transkriptionsbegriffs selbst begründet, dessen terminologische Verwendungsweise eher ungeklärt

bleibt. Da die Zeichen selbst als Subjekt und Agens der Transkription angenommen werden, entzieht sich der Begriff einer verwendungsorientierten, gebrauch- oder handlungstheoretischen Auflösung. Stattdessen kommt es zu einer Hypostasierung der Zeichen, derzufolge Bild und Text sich gegenseitig transkribieren und damit „die Lesbarkeit unlesbar gewordener Ausschnitte wiederherstellt“ (Jäger 2004: 72, zitiert in Holly 2006: 13). Die Folge dieser Betrachtungsweise ist die in Kapitel 2 bereits beschriebene repräsentationale Bedeutungsauffassung, derzufolge die „Unlesbarkeit“ und Unverständlichkeit der Kommunikationselemente in den verschiedenen Modi als Ausgangspunkt gesetzt wird. Ganz im Gegensatz dazu stehen die Beschreibungen exemplarischer Beispiele durch die Vertreter einer Transkriptionstheorie: Beschreibungen wie „mit Bildern authentisieren, autorisieren, sensualisieren, adressieren, grundieren oder übermalen“ sind eindeutig handlungsbezogen und instrumentell (Holly 2007).

Es ist eine logische Konsequenz aus dieser Unverträglichkeit zwischen Theorie und Methode, den Begriff der Transkription zur Beschreibung der Deutungsdynamik multimodaler Angebote fallen zu lassen und durch einen anderen zu ersetzen. Dafür bietet sich der Begriff der Interaktion an, der seit vielen Jahrzehnten als Zentralbegriff einer rezipientenorientierten Verstehenstheorie in Umlauf ist und auch handlungstheoretisch und akteursbezogen verwendet werden kann (Iser 1980; Bucher 2004a). Nicht Elemente eines Kommunikationsangebotes agieren miteinander, sondern der Rezipient agiert mit den Elementen eines Angebotes auf der Basis seiner in dialogischen Kommunikationen erworbenen Interaktionskompetenz. Natürlich kann auch der Transkriptionsbegriff handlungstheoretisch verwendet werden, ohne dass das begriffliche Verwirrung stiften muss. Das ist zum Beispiel dann der Fall, wenn ein Akteur die aufgeführte Musik in Noten transkribiert, gesprochene Sprache in geschriebene oder aus einem Spielfilm einen Trailer macht. Im Folgenden soll erstens gezeigt werden, dass Blickaufzeichnungsdaten als Rezeptionsindikatoren Indizien für eine interaktionale Theorie des multimodalen Verstehens liefern können. Zweitens sollen diese Befunde genutzt werden, um den Interaktionsprozess zwischen medialem Angebot und Rezipient zu erhellen.

Zerlegt man den Blickverlauf bei der Rezeption eines multimodalen Angebotes wie Zeitungsseiten, Bildschirmseiten eines Online-Angebotes oder eines Fernsehbeitrages oder Werbeanzeigen in einzelne Zeitabschnitte – z.B. in Sekundenabschnitte –, so wird deutlich, dass in unterschiedlichen Phasen auch jeweils spezifische Elemente des Angebotes betrachtet werden. Orientierungselemente wie der Seitenkopf oder das Medienlogo erhalten in der Anfangsphase der Rezeption einer Zeitung oder einer Online-Zeitung hohe Aufmerksamkeit. Ab einem bestimmten Zeitpunkt werden sie aber nicht mehr wahrgenommen, was darauf hinweist, dass sie offensichtlich ihre Funktion als Orientierungshilfen erfüllt haben. Andere Elemente, wie die Navigationsleiste einer Online-Zeitung werden über eine längere Zeit regelmäßig und immer wieder in die Rezeption einbezogen. Ein anderes Muster ist die wechselweise Fixierung bestimmter Elemente, wie z.B. benachbarter Text- und Bildelemente. Die Abbildungen 2 und 3 zeigen zwei Beispiele, in denen dieser Prozess der wechselseitigen Fixierung von Text und

Bild erkennbar ist. Im ersten Fall werden in den ersten 20 Sekunden der Betrachtung der Titelseite der Rheinzeitung vom 11.05.05 das Aufmacherphoto und die Aufmacherüberschrift wechselseitig fixiert. Im zweiten Beispiel, den Blickdaten aus der Rezeption der entsprechenden Online-Ausgabe, werden Bild und Teasertext abwechselnd betrachtet.

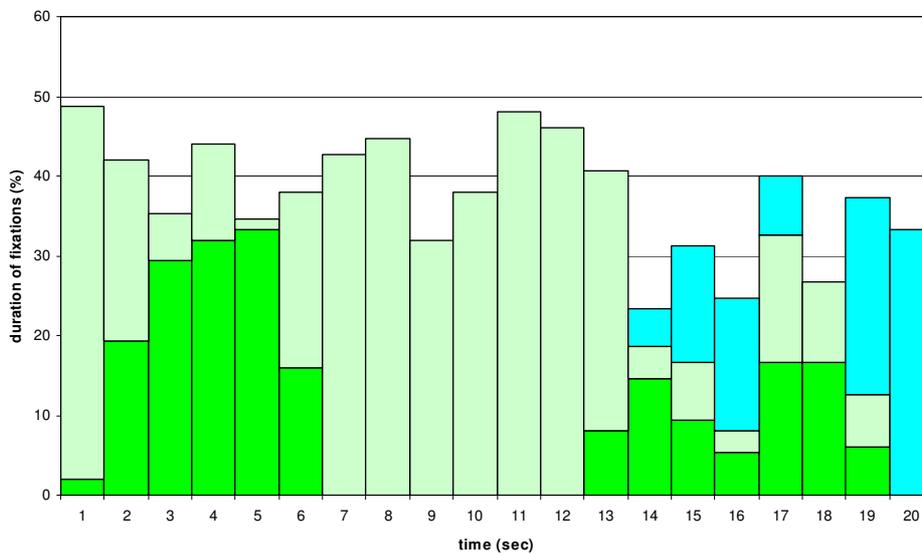


Abb. 2: Die interaktive Erschließung des Aufmachers der Rheinzeitung vom 11.05.2005 in Sekundenschritten: Aufmacherfoto und Aufmacherüberschrift werden in den ersten 20 Sekunden jeweils wechselseitig rezipiert. Die Farbsäulen in der Verlaufsgrafik entsprechen den gleichfarbigen „Areas of Interest“ (AOIs) auf der Zeitungsseite. In der ersten Sekunde wird beispielsweise nahezu ausschließlich das Aufmacherfoto wahrgenommen, während die Lektüre des Aufmachertextes erst in Sekunde 14 einsetzt. Die übrigen AOIs der Seiten wurden in der Blickaufzeichnung zwar miterfasst, sind hier aber aus Gründen der Übersichtlichkeit ausgeblendet.

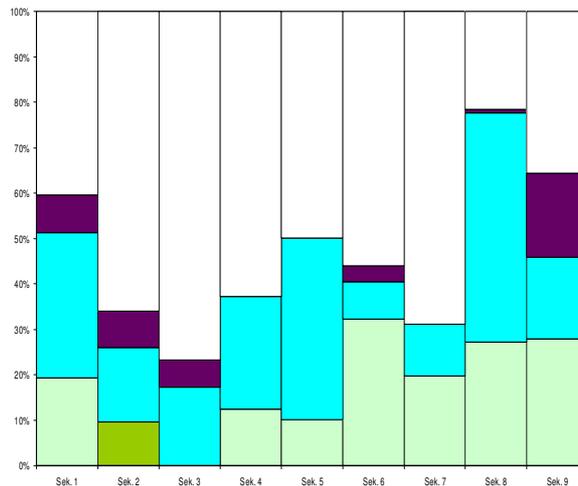


Abb. 3: Die interaktive Erschließung des Aufmachers in der Online-Ausgabe der Rheinzeitung vom 11.05.2005 in Sekundenschritten: Aufmacherfoto und Teasertext werden in den ersten 10 Sekunden jeweils wechselseitig rezipiert.

Da sich systematische Sequenzen solcher alternierender Fixationen von zwei Elementen auch im Falle anderer Angebote wiederfinden, kann man sie als Indikatoren für eine rekursive Deutung auffassen: Für das Verstehensproblem, das eines der beiden Elemente aufwirft, werden deutungsrelevante Hinweise in einem benachbarten Element gesucht, das im Falle der Abbildung 2 interessanterweise nicht der Bildtext, sondern das optisch auffallendere Element, die Schlagzeile des Aufmacherartikels, ist. Nachdem das Aufmacherfoto angeschaut wurde, wird die Überschrift gelesen:

Denkmal für 6 000 000 Opfer. Holocaust-Gedenkstädte im Berlin eingeweiht. KZ-Überlebende hielt bewegende Ansprache.

Nach der Lektüre wird nochmals das Bild betrachtet, offensichtlich jetzt mit einer anderen Wissensbasis. Erst nachdem dieser Deutungsprozess abgeschlossen ist, beginnt die Lektüre des Aufmacher-Textes. Im Falle der Online-Zeitung werden wechselweise das Foto und der daneben platzierte Teasertext rezipiert.

Sowohl in der gedruckten Ausgabe als auch in der Online-Ausgabe ist das jeweils verwendete Foto isoliert betrachtet nur schwer zu verstehen. Die Steinstele der Holocaust-Gedenkstätte sind in beiden Fällen in einem eher grafischen Stil abgebildet, so dass nur schwer erkennbar ist, was die Fotos zeigen und sich deshalb ein dokumentarisches Verständnis des Fotos nicht ohne weiteres einstellt. Das Foto der Online-Ausgabe unterscheidet sich insofern vom Foto der gedruckten Ausgabe, als letzteres mit der Rose im Vordergrund eine zweite Deutungsebene enthält. Der Leser muss nicht nur verstehen, was das Foto zeigt – die neue Holocaust-Gedenkstätte – sondern auch, was diese Rose bedeutet. Damit wird deutlich, dass ein Verständnis des Fotos zwei Ebenen umfasst: die referentielle Ebene und eine szenische, die den Entstehungszusammenhang des Fotos umfasst. Das gilt im Übrigen auch für das Foto der Online-Ausgabe, wo diese zweite Ebene allerdings keine Markierung im Foto selbst aufweist. Mit dem im jeweiligen Textelement enthaltenen Hinweis auf den Einweihungsakt wird der Entstehungszusammenhang des Fotos ausgedrückt. Das Foto selbst wird erst durch diese Information als aktuelles, dokumentarisches Foto verstehbar und nicht – beispielsweise – als nicht-aktuelles Schmuckbild aus dem Archiv. Die Relevanz des Entstehungs- und Herstellungszusammenhangs von Abbildungen für deren Verständnis wird dann besonders deutlich, wenn in manipulativer Absicht inszenierte oder gefälschte Fotos verwendet werden. (Muckenhaupt 1986: Kap. 4.3.5). Ein Bild verstehen bedeutet deshalb nicht nur zu verstehen was das Bild zeigt, sondern auch, was mit dem Bild gezeigt werden soll. Im vorliegenden Fall soll mit dem Foto gezeigt werden, dass die Holocaust-Gedenkstätte eingeweiht wurde, was voraussetzt, dass das Bild die Gedenkstätte zeigt. Der wiederholte Wechsel in der Rezeption von Bild und Text im vorliegenden Beispiel könnte erklären, dass diese Ebenen des Bildverstehens auch in Schritten phasenweise abgearbeitet werden: es wird schrittweise geklärt, was das Foto zeigt (die Holocaust-Gedenkstätte), wie und warum es entstanden ist (anlässlich der Einweihung) und wie es verwendet wird (als Dokumentation der Einweihung). In welcher Reihenfolge diese Ebenen des Verstehens erschlossen werden, kann variieren und dementsprechend offen bleiben.

Das Verständnis der Text-Bild-Kombination wird stufenweise erweitert, indem die zusätzlich gewonnene Information aus dem vorausgegangenen Deutungsschritt herangezogen wird. Dieser Prozess ist insofern rekursiv, als jede Deutung als Grundlage einer weiteren Deutung fungiert. Insofern kann die Blickaufzeichnung in diesem Falle auch als Visualisierung des sogenannten hermeneutischen Zirkels gesehen werden. Von einer interaktionalen Aneignung zu sprechen ist deshalb gerechtfertigt, weil es sich nicht um eine reaktive oder additive Aneinanderreihung von Deutungen handelt, sondern um systematisch verschränkte, bei denen die Elemente wechselseitig kontextualisiert werden. Der Rezeptionsprozess folgt einem Relevanzprinzip, demzufolge derjenige Aspekt

des Fotos oder des Textes für die Deutung herangezogen wird, der für den jeweils neu entstandenen Kommunikationsstand relevant ist. Der Deutungsprozess folgt dem Muster einer antizipierten kontrafaktisch unterstellten Interaktion, in der sich sowohl der Kontext des zu deutenden Elements als auch der Wissensstand des Rezipienten fortlaufend verändern. Analog zu einer dialogischen Kommunikation „erhält“ er vom Kommunikationsangebot jeweils die Information, die dem jeweiligen Kommunikationsstand entspricht. Mit dem Wissen aus dem Text, dass die Holocaust-Gedenkstätte eingeweiht wurde, sieht der Zeitungsleser das Aufmacherbild anders, beispielsweise als Darstellung dieser Einweihung.

Für Text-Bild-Kombinationen lässt sich diese Beschreibung multimodalen Verstehens folgendermaßen verallgemeinern: Das Verstehen selbst ist ein mehrstufiger Prozess, der sich entlang der verschiedenen Aspekte des Verstehens systematisieren lässt. Für Textbeiträge – geschriebene und gesprochene – sind diese Aspekte oder Erschließungsebenen mehrfach beschrieben worden und umfassen beispielsweise die Funktion einer Kommunikationseinheit, ihre Inhalte (Propositionen), ihr Thema, die mit ihr verfolgten strategischen Absichten, ihre sprachliche Form, die Festlegungen, die jeweils eingegangen werden und das vorausgesetzte Wissen (Fritz 1994; Bucher 1999). Im Falle multimodaler Angebote ist diese Liste um die Aspekte des Designs und der Anordnung und Hierarchisierung im Kommunikationsraum zu erweitern. Für bildliche Darstellungen hat Muckenhaupt eine spezifische Aspektliste vorgelegt (Muckenhaupt 1986: 199f.). So gehört es zum Verstehen eines Bildes, dass wir verstehen,

- welche Art von Handlung mit dem Bild vollzogen wird,
- welchen Inhalt die kommunikative Handlung aufweist,
- was mit dem Bild gezeigt wird,
- auf welche Weise etwas mit dem Bild gezeigt wird,
- welche Voraussetzungen derjenige macht, der die kommunikative Handlung mit dem Bild vollzieht,
- auf welche Folgerungsbeziehungen er sich mit Verwendung des Bildes festlegt,
- in welchem Handlungszusammenhang das Bild verwendet wird,
- wie das Bild entstanden ist oder hergestellt wurde.

Die interaktionale Struktur multimodalen Verstehens besteht darin, dass für jedes der einzelnen Elemente in einem Kommunikationsraum die jeweils spezifischen Ebenen der Erschließung durchlaufen werden um so schrittweise das Verständnis systematisch zu erweitern. Eine Systematisierung der einzelnen Deutungsschritte kann dabei aus den modusspezifischen Aspekten des Verstehens der einzelnen sinntragenden Einheiten eines multimodalen Angebotes abgeleitet werden. Insofern besteht der Vorteil des Interaktionsbegriffs gegenüber dem Transkriptionsbegriff darin, dass er eine produktive Nutzung modispezifischer Kommunikations- und Verstehenstheorien eröffnet, während letzterer begrifflich ungeklärt bleibt.

Die Interaktionstheorie lässt sich über Text-Bild-Konstellationen hinaus auch auf andere Modi ausweiten: der Rezipient erkennt ein musikalisches Motiv oder hört die Textzeile eines Songs und versteht den Zusammenhang von Musik und Szene; er identifiziert ein Geräusch, und versteht, an welchem Schauplatz sich die Handlung des Audiofeatures abspielt; er liest eine entsprechende Textpassage und versteht die Funktion der Abbildung. Im vorliegenden Beispielfall (siehe Abbildung 2) gehört es auch zum Verständnis des multimodalen Clusters aus Bild und verschiedenen Textelemente, dass diese aufgrund von Design und Layout als Aufmacher der Titelseite oder der Einstiegsseite erkannt werden. Offensichtlich sind es die optisch auffallenden Gestaltungselemente des Aufmachers, die beim Leser diesen Deutungsprozess auslösen und die übrigen Elemente des Kommunikationsangebotes in den Hintergrund treten lassen. Auch das Design wird in multimodalen Angeboten als Form des kommunikativen Handelns aufgefasst und in den interaktiven Prozess des Verstehens einbezogen. Es *zeigt an*, wie wichtig ein Beitrag ist, welche Beiträge zusammengehören und welche nicht. Eine interaktionale Theorie des multimodalen Verstehens hat einen weiteren Vorteil: ihre Rekursivität kann die Bidirektionalität des Deutungsprozesses erklären, derzufolge jedes Element im Lichte der anderen Elemente gedeutet wird. Man kann dementsprechend die Struktur multimodaler Handlungen, wie sie beispielsweise in der Tageszeitung oder in einer Nachrichtensendung vorkommen, folgendermaßen beschreiben:

| | | |
|-------------------------------------|---|----------------------------|
| A informiert B über X, indem | A berichtet , wie sich X ereignet hat | Text |
| | A zeigt , wo/wie sich X ereignet hat | Foto, Graphik, Filmbeitrag |
| | A anzeigt , welchen Stellenwert der Bericht über X hat | Design |
| | A anzeigt , was zum Bericht über X gehört | Design |
| und dann | das berichtete Ereignis X kommentiert | Fortsetzungstext |

Tab. 1: Struktur multimodaler Handlungen

Diese Struktur multimodaler Handlungen ist auf andere Handlungsmuster wie „Für ein Produkt/ eine Partei werben“ oder „Erklären, wie X funktioniert“ übertragbar. Für eine Theorie des multimodalen Verstehens zeigt die Strukturbeschreibung, welche Ebenen der Kommunikation jeweils relevant sind – das Geäußerte, das Gezeigte und das Angezeigte –, mit welchen Mittel die jeweiligen Teilhandlungen realisiert werden und wie diese Teilhandlungen zusammenhängen.

4. Konsequenzen oder: *Entia non sunt multiplicanda sine necessitate*

Der hier gemachte Vorschlag, multimodales Verstehen im Rahmen einer interaktionalen Rezeptionstheorie zu klären, stützt sich auf empirische Befunde von Blickaufzeichnungsstudien. Hierbei darf nicht übersehen werden, dass Blickaufzeichnungen zwar darüber Aufschluss geben, was die Probanden anschauen, nicht aber darüber, was sie sehen. Diese interpretatorische Lücke habe ich versucht auf drei Wegen theoretisch auszufüllen: Aus der kritischen Diskussion verschiedener Theorien und Ansätze multimodalen Verstehens wurde *erstens* eine interaktionale Theorie entwickelt, die die Blickaufzeichnungsbefunde als Rezeptionsindikatoren integriert. Basis dieser Theorie sind einerseits Ansätze aus Rezeptions- und Verstehenstheorien, die dem sogenannten interaktionalen Paradigma zugeordnet werden können und sich dementsprechend von einem wirkungsorientierten Paradigma abgrenzen (vgl. Bucher 2008). *Zweitens* wurden die zentralen Elemente einer handlungstheoretischen Kommunikations- und Sprachtheorie genutzt, um wesentliche Differenzierungen für das multimodale Verstehen einzuführen: die Organisationsebenen von Kommunikation wie Funktionen, Themen, Inhalten, Festlegungen, Wissensvoraussetzungen und Handlungsprinzipien und die daraus ableitbaren Aspekte des Verstehens. Ein *dritter* Weg zur Theoriebildung war die Rückführung der Multimodalität auf den Begriff der Non-Linearität. Er eröffnet die Möglichkeit, Erkenntnisse zum Verständnis von Hypertexten auf das multimodale Verstehen zu übertragen. Das wichtigste Ergebnis dieser Übertragung besteht in der Konkretisierung des Begriffs der Interaktivität zur Beschreibung der spezifischen Aneignungsdynamik multimodaler Kommunikationsangebote. Multimodales Verstehen wird dadurch modellierbar als eine Form des Problemlösens, das sich in für Hypertexte spezifische Teilprobleme systematisieren lässt. Der Begriff der Non-Linearität bietet auch die Basis, um die spezifische räumliche Logik multimodaler Kommunikation – ihre konstellative Struktur – sowie deren Markierung im Design als weitere Ebene des Verstehens in eine umfassende Theorie zu integrieren.

Als eine besondere Herausforderung für eine Erklärung des multimodalen Verstehens gilt der sogenannte Multiplikationseffekt. Er stellt eine Theorie vor die Aufgabe „to capture the way in which different semiotic modalities co-contextualise each other in ways that are not predictable on the basis of the different semiotic resources seen as separate modalities“ (Baldry / Thibault 2005: 18). In Kapitel 1 habe ich diese Herausforderung als das Kompositionalitätsproblem eingeführt und um das Rezeptionsproblem ergänzt. Beide Probleme sind nicht erst mit den multimodalen Kommunikationsformen aufgekommen, sondern beschäftigen Sprach-, Kommunikations- und Medienwissenschaft schon seit längerer Zeit. Mit einer interaktionalen Auffassung des multimodalen Verstehens kann der Multiplikationseffekt auf der Grundlage einer handlungstheoretisch fundierten Verstehenstheorie für multimodale Kommunikationsformen beschrieben werden. Die Frage, ob dafür herkömmliche Sprach- und Kommunikationstheorien ausreichen, habe ich in diesem Beitrag praktisch zu beantworten

versucht: Es kommt darauf an, die Potentiale richtig zu nutzen. Das jedenfalls ist erfolgsversprechender, als mit neuen Begriffen neue Probleme einzuführen.

Literatur

- Baldry, Anthony / Thibault, Paul, J. (2005): *Multimodal Transcription and text analysis. A multimedia toolkit and coursebook with associated online-course*. London/Oakville: Equinox.
- Barnhurst, Kevin G. / Nerone, John C. (2001): *The Form of News. A History*. New York/ London: The Guildford Press.
- Barthes, Roland (1990): *Rhetorik des Bildes*. in: Roland Barth (Hg.), *Der entgegenkommende und der stumpfe Sinn*. Frankfurt: Suhrkamp, 28-46 (Orig. 1982).
- Bateman, John (2008): *Multimodality and genre: A foundation for the systematic analysis of multimodal documents*. London: Palgrave Macmillan.
- Bateman, John et al. (2007): *Mapping the Multimodal Genres of Traditional and Electronic Newspapers*. in: Terry D. Royce / Wendy L. Bowcher (Hgg.), *New Directions in the Analysis of Multimodal Discourse*. New Jersey/London: Lawrence Erlbaum, 147-172.
- Blum, Joachim/Bucher, Hans-Jürgen (1998): *Die Zeitung: Ein Multimedium. Textdesign – ein Gestaltungskonzept für Text, Bild und Grafik*. Konstanz: UVK.
- Bucher, Hans-Jürgen (1998): *Vom Textdesign zum Hypertext. Gedruckte und elektronische Zeitungen als nicht-lineare Medien*. in: Werner Holly / Bernd-Ulrich Biere (Hgg.), *Medien im Wandel*. Wiesbaden: Westdeutscher Verlag, 63-102.
- Bucher, Hans-Jürgen (1999): *Sprachwissenschaftliche Methoden der Medienanalyse*. in: Joachim-Felix Leonhard et al. (Hgg.), *Medienwissenschaft. Ein Handbuch zur Entwicklung der Medien und Kommunikationsformen*. Berlin/New York: de Gruyter, 213-231.
- Bucher, Hans-Jürgen (2006): *Gedrucktes im Internet. Online-Zeitungen und Online-Magazine auf dem Weg zu einer eigenständigen Mediengattung*. in: Peter Schlobinski (Hg.), *Von *hdl* bis *cul8r**. Sprache und Kommunikation in den neuen Medien. Mannheim: Duden Verlag, 210-232.
- Bucher, Hans-Jürgen (2007): *Textdesign und Multimodalität. Zur Semantik und Pragmatik medialer Gestaltungsformen*. in: Kersten Sven Roth / Jürgen Spitzmüller (Hgg.), *Textdesign und Textwirkung in der massenmedialen Kommunikation*. Konstanz: UVK, 49-76.
- Bucher, Hans-Jürgen et al. (Hgg.) (2007): *Mit den Augen der Leser: Broadsheet und Kompakt-Format im Vergleich. Eine Blickaufzeichnungsstudie zur Leser-Blatt-Interaktion*. Ifra Special Report. Darmstadt: Ifra.
- Bucher, Hans Jürgen (2004a): *Online-Interaktivität - ein hybrider Begriff für eine hybride Kommunikationsform. Begriffliche Klärungen und empirische Rezeptionsbefunde*. in: Christoph Bieber / Claus Leggewie (Hgg.), *Interaktivität. Ein transdisziplinärer Schlüsselbegriff*. FrankfurtMain: Campus, 132-167.
- Bucher, Hans Jürgen (2004b): *Is there a Chinese Internet? Intercultural investigation on the Internet in the People's Republic of China: Theoretical considerations and empirical results*. in: Fay Sudweek/Charles Ess (Hgg.), *Forth International Confer-*

- ence: Cultural Attitudes towards Technology and Communication 2004 (Proceedings). Murdoch WA 6150: Murdoch University (Australia). 416-428.
- Bucher, Hans Jürgen (2008): Vergleichende Rezeptionsforschung: Theorien - Methoden - Befunde. in: Josef Seethaler / Jürgen Wilke (Hgg.), Medien- und Kommunikationsforschung im Vergleich. Grundlagen, Gegenstandsbereiche, Verfahrensweisen. Wiesbaden: VS Verlag für Sozialwissenschaften, 309-340.
- Bucher, Hans Jürgen / Schumacher, Peter (2006): The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print- and online media. in: Communications. The European Journal of Communications Research 31/3/2006: 347-368.
- Bucher, Hans-Jürgen et al. (Hgg.) (2007): Mit den Augen der Leser: Broadsheet und Kompakt-Format im Vergleich. Eine Blickaufzeichnungsstudie zur Leser-Blatt-Interaktion. Ifra Special Report. Darmstadt: Ifra.
- Duchowski, Andrew T. (2003): Eye Tracking Methodology: Theory and Practice. London: Springer.
- Fei, Victor Lim (2004): Developing an integrative multi-semiotic model. in: Kay L. O'Halloran (Hg.), Multimodal Discourse Analysis. Systemic Functional Perspectives. London/New York: Continuum.
- Folker, Sonja et al. (2005): Processing and Integrating Multimodal Material - The Influence of Color-Coding. in: Bruno G. Bara / Lawrence Barsalou / Monica Bucciarelli (Hgg.), Proceedings of the 27th Annual Conference of the Cognitive Science Society. Mahwah/NJ: Lawrence Erlbaum Associates, 690-695.
- Fritz, Gerd (1994): Grundlagen der Dialogorganisation. in: Gerd Fritz / Franz Hundsnurscher, (Hgg.), Handbuch der Dialoganalyse. Tübingen: Niemeyer, 177-201.
- Fritz, Gerd (2008): Rhetorical Structure Theory. Gießen (unveröffentlichtes Manuskript).
- Garcia, Mario (1993): Contemporary Newspaper Design. A Structural Approach. Englewood Cliffs/New York: Prentice Hall.
- Halliday, M.A.K. (2004): An introduction to functional grammar. London: Hodder Arnold.
- Holly, Werner (2005): Zum Zusammenspiel von Sprache und Bildern im audiovisuellen Verstehen. in: Dietrich Busse / Thomas Niehr / Martin Wengeler (Hgg.), Brisante Semantik. Neuere Konzepte und Forschungsergebnisse einer kulturwissenschaftlichen Linguistik. Tübingen: Niemeyer, 337-353.
- Holly, Werner (2006): Mit Worten sehen: Audiovisuelle Bedeutungskonstitution und Muster „transkriptiver Logik“ in der Fernsehberichterstattung. in: Deutsche Sprache 1-2/2006, 135-150.
- Holly, Werner (2007): Audiovisuelle Hermeneutik. Die Zeitgebundenheit des Bild-Verstehens am Beispiel der Medienberichterstattung. in: Fritz Hermanns / Werner Holly (Hgg.), Linguistische Hermeneutik. Theorie und Praxis des Verstehens und Interpretierens. Tübingen: Niemeyer, 357-386.
- Holly, Werner (2009): Der Wort-Bild-Reißverschluss: Über die performative Dynamik audiovisueller Transkriptivität. in: Helmuth Feilke / Angelika Linke (Hgg.), Oberfläche und Performanz (in Druck).
- Holsanova, Jana et al. (2008): Reading information graphics: The role of spatial contiguity and dual attentional guidance. Lund: Cognitive Science Department, 1-25.

- Holsanova, Jana et al. (2006): Entry points and reading paths on newspaper spreads: comparing a semiotic analysis with eye-tracking measurements. in: *Visual Communication* 5/1/2006, 65-93.
- Iedema, Rick (2003): Multimodality, Resemiotization: Extending the Analysis of Discourse as Multi-Semiotic Practice. in: *Visual Communication* 2/1/2003, 29-57.
- Iser, Wolfgang (1980): Interaction between text and reader. in: Susan R. Suleiman / Inge Crosman (Hgg.), *The reader in the text. Essays on audience and interpretation.* Princeton/NJ: Princeton University Press, 106-119.
- Itti, Laurent / Koch, Christof (2000): A salience-based search mechanism for overt and covert shifts of visual attention. in: *Vision Research* 40/2000, 1489-1509.
- Jäger, Ludwig et al. (2008): Transkriptivität: Operative Medientheorien als Grundlage von Informationssystemen für die Kulturwissenschaften. in: *Informatik_Spektrum_31_1_2008*, 21-29.
- Jarvenpaa, Sirkka L. (1990): Graphic displays in Decision Making - The Visual Salience Effect. in: *Journal of Behavioral Decision Making* 3/1990, 247-262.
- Kress, Gunther / van Leeuwen, Theo (1996): *Reading Images. The Grammar of Visual Design.* New York: Routledge.
- Kress, Gunther / van Leeuwen, Theo (1998): *Front Pages. The Critical Analysis of Newspaper Layout.* in: Alan Bell / Peter Garrett (Hgg.), *Approaches to media discourse.* Oxford: Wiley Blackwell.
- Kress, Gunther / van Leeuwen, Theo (2001): *Multimodal Discourse. The Modes and Media of Contemporary Communication.* London: Arnold.
- Kress, Gunther / van Leeuwen, Theo (2002): Colour as a Semiotic Mode: Notes for a Grammar of Colour. in: *Visual Communication* 1/3/2002, 343-368.
- Krippendorff, Klaus (2006): *The Semantic Turn: A New Foundation for Design.* Boca Raton, Florida: CRC Press.
- Lemke, Jay L. (2002): Travels in Hypermodality. in: *Visual Communication* 1/3/2002, 299-325.
- Lindström, Paulina (2006): Understanding is identifying a structure. in: *Lund University Cognitive Studies* 131/2006 [<http://www.lu.se/LUCS/131/LUCS.131.pdf>].
- Martinec, Radan/Salway, Andrew (2005): A system for image-text relations in new (and old) media. in: *Visual Communication* 4/3/2005, 337-371.
- Mirzoeff, Nicholas (Hg.) (1998): *The Visual Culture Reader.* London/New York: Routledge.
- Mitchell, William J. Thomas (1995): *Picture Theory. Essays on Verbal and Visual Representation.* Chicago u.a.: University of Chicago Press.
- Muckenaupt, Manfred (1986): *Text und Bild. Grundfragen der Beschreibung von Text-Bild-Kommunikationen aus sprachwissenschaftlicher Sicht.* Tübingen: Niemeyer.
- Neumann, Odmар (1992): Theorien der Aufmerksamkeit: von Metaphern zu Mechanismen. in: *Psychologische Rundschau* 43/1992, 83-101.
- Neumann, Odmар (1996): Theorien der Aufmerksamkeit. in: Odmар Neumann / Andries F. Sanders (Hgg.), *Enzyklopädie der Psychologie. Themenbereich C Theorie und Forschung. Serie II Kognition. Band 2 Aufmerksamkeit.* Göttingen u.a.: Hogrefe, 559-643.

- Ong, Walter J. (2002): *Orality and Literacy. The technologizing of the world.* London/ New York: Routledge (Orig. 1982).
- Sachs-Hombach, Klaus (2003): *Das Bild als kommunikatives Medium. Elemente einer allgemeinen Bildwissenschaft.* Köln: Herbert von Halem Verlag.
- Schmitz, Ulrich (2004): *Sprache in modernen Medien. Einführung in Tatsachen, Theorien, Themen und Thesen.* Berlin: Erich Schmidt.
- Schmitz, Ulrich (2007): *Bildakte? How to do things with pictures.* in: *Zeitschrift für Germanistische Linguistik (ZGL)* 35/2007, 419-433.
- Schnotz, Wolfgang/Bannert, Maria (2003): *Construction and interference in learning from multiple representations.* in: *Learning and Instruction* 13/2003, 141-156.
- Schrivver, Karen A. (1997): *Dynamics in Document Design: Creating Text for Readers.* New York: Wiley Computer Publishing.
- Shriver, Karen (1996): *Dynamics of Document Design. Creating Texts for Readers.* New York u.a.: John Wiley & Sons Inc.
- Smith, Ken (2005): *Perception and the Newspaper Page: A Critical Analysis.* in: Ken Smith et al. (Hgg.), *Handbook of Visual Communication. Theory, Methods and Media.* Mahwah/London: LEA, 81-95.
- Stark, Peggy et al. (2007): *Eyetracking the News. A study of print and online reading.* St. Petersburg FL: Poynter.
- Stöckl, Hartmut (2004): *Die Sprache im Bild - Das Bild in der Sprache. Zur Verknüpfung von Sprache und Bild im massenmedialen Text. Konzepte. Theorien. Analysemethoden.* Berlin: de Gruyter.
- Stöckl, Hartmut (2004): *Typographie: Körper und Gewand des Textes. Linguistische Überlegungen zu typographischer Gestaltung.* in: *Zeitschrift für Angewandte Linguistik (ZfAL)* 41/2004, 5-48.
- Stöckl, Hartmut (2005): *Typography: body and dress of a text – a signing mode between language and image.* in: *Visual Communication. Special Issue “The New Typography”* 4/2/2005, 204-214.
- Stöckl, Hartmut (2007): *Hörfunkwerbung – „Kino für das Ohr“. Medienspezifika, Ko- deverknüpfungen und Textmuster einer vernachlässigten Werbeform.* in: Kersten Sven Roth / Jürgen Spitzmüller (Hgg.), *Textdesign und Textwirkung in der massenmedialen Kommunikation.* Konstanz: UVK, 177–202.
- Stöckl, Hartmut (2008): *Werbetypographie. Formen und Funktionen.* in: Sylvia Bendel / Gudrun Held (Hgg.), *Werbung – grenzenlos. Interkultureller Blick auf multimodale Gestaltungsstrategien aktueller Werbetexte (= Sprache im Kontext).* Frankfurt/Wien: Lang, 13–36.
- Van Dijk, Teun A. (2008): *Discourse and Context. A Socio-Cognitive Approach.* Cambridge: Cambridge University Press.
- Wittgenstein, Ludwig (1977): *Philosophische Untersuchungen.* Frankfurt: Suhrkamp (Original: 1958).