Werner Holly (Chemnitz)
Sprache, Bild, Text

Visualität und Intermedialität von Sprache

Abstract: In den letzten Jahrzehnten hat sich die Perspektive auf den Gegenstand der Sprachwissenschaft immer wieder verändert, vor allem gegen überkommene Reduktionismen erweitert. Hier soll auf Entwicklungen im Zusammenhang mit der Herausbildung einer „Medienlinguistik“ eingegangen werden, die (allgemeiner) die Medialität von Sprache behandelt, auch im Zusammenspiel mit anderen Zeichenarten, (spezieller) die Rolle von Sprache in (technischen) Medien. Von den sehr zahlreichen Varianten der Kombination von Modalitäten und Kodalitäten sollen hier zwei sehr unterschiedliche Schwerpunkte betrachtet werden: Visualität von Sprache, *in* und *von* Texten, und sekundäre Audiovisualität.

1 Zur Einführung: ein medial komplexes Poster

Zur Illustration und Legitimation meines Themas möchte ich mit einem „Paratext“ dieses Bandes einsteigen, einem Poster, das die Tagung des Instituts für Deutsche Sprache, die diesem Band zugrunde liegt, ankündigte (siehe Abb. 1).

Entscheidend ist, dass wir hier etwas *sehen*. Was sehen wir? Schrift, (im Original) Farben, alles in einer bestimmten Anordnung, grafisch gestaltet. Natürlich kann man, muss man Geschriebenes immer sehen, in einem trivialen Sinn, aber wir sagen dennoch nicht, dass wir einen Text *sehen*, wir *lesen* ihn. Hier kann man sagen, dass wir das Geschriebene nicht nur lesen, wir sollen es auch in einem nicht trivialen Sinn sehen, wir sollen sehen, dass bestimmte Schriftzeichen Versalien zeigen, andere nicht, dass die Buchstaben verschiedene Größen haben, verschiedene Farben und an sehr verschiedenen Stellen stehen, die nach der üblichen Leserichtung bei uns von links oben nach rechts unten auch eine bestimmte Reihenfolge nahe legen, aber auch nicht sehr strikt, wahrscheinlich springen wir – eher von der Größe geführt – herum und setzen Gewichtungen. Wir sehen auch, dass die Schrifttypen variieren, es gibt zwei so genannte Logos mit verschiedenen Schriften und wir verstehen, dass rechts unten nicht Leibniz als Autor dieses Textes signiert hat, sondern dass wir es hier mit einer renommierten Forschungseinrichtung der Leibniz-Gemeinschaft zu tun haben, die jetzt 50 Jahre besteht.

Abb. 1: Poster zur IDS-Tagung

Wir sehen außerdem (im Original) farbige grafische Elemente, die verbildlichen, dass hier etwas *fokussiert* wird, weil sie sich in einer Bewegung konzentrisch auf etwas richten, die Sprachwissenschaft eben. Das verstehen wir, weil uns der Sprachtext mit der Schlagzeile „IM FOKUS“ überdeutlich sagt, was wir sehen sollen, danach sehen wir es – oder ahnen es eher – auch farbig verbildlicht (was der Schwarz-Weiß-Druck nicht zeigt). Das ist aber etwas anderes, als wenn wir es nur gelesen hätten, es ist viel schöner, nicht nur im Sinn des Ornamentalen, man könnte es sogar deuten, es hat das „Blau des Geistes“, das „Rot der Leidenschaft“, der wissenschaftlichen natürlich, und das „Grün der Hoffnung“, verknüpft mit

dem Gelbton, der zur Corporate Identity des Instituts gehört und entsprechend repräsentiert ist. Aber hier breche ich die wilden Deutungen besser ab. Es soll ja nur gezeigt werden, wie Sprache, hier in Form von Schrift, mit anderen Zeichen wie selbstverständlich kombiniert und verstanden wird, weil alle „cross-kodal“ zusammenspielen.

Und wenn wir genau hinsehen, erkennen wir am Schluss dieser kleinen Bildbetrachtung links unten ein schwarz gepixeltes Quadrat. Es handelt sich hierbei um einen QR-Code (QR für *quick response*). Wenn man nun ein Smartphone, das mit einer entsprechenden App ausgestattet ist, darauf richtet, landet man auf einer Website des IDS, die die Tagung ausführlicher ankündigt.

Mit diesem kleinen unauffälligen Textelement ist aus dem guten alten Plakat ein regelrechter Hypertext geworden, jedenfalls für die Kenner und Smartphone-User. Ohne die Einbeziehung solcher medialer Strukturen haben wir nur eine sehr simple und reduzierte Vorstellung von Sprache und ihrem Gebrauch. Die logische Folgerung ist, dass Sprache ohne Medialitäten nicht angemessen beschrieben werden kann.

Damit möchte ich andeuten, worum es im Folgenden gehen soll, nämlich um die These, dass sich die Sprachwissenschaft nicht nur um tradierte Pfeiler wie Grammatik, Lexik, Pragmatik kümmern muss, sondern auch um sehr elementare Fragen der medialen Bedingungen, unter denen z.B. Grammatik, Lexik und Pragmatik funktionieren. Dabei konzentriere ich mich auf das Gebiet der Visualität und der Intermedialität von Sprache. Wie sich solche Fragen in den letzten 50 Jahren entwickelt haben, kann ich hier nicht im detaillierten Verlauf und im vollen Umfang nachzeichnen, höchstens blitzlichtartig illustrieren.

Das exemplarische Vorgehen soll hier in zwei Schritten erfolgen, denen zwei gleichermaßen wichtige Aspekte einer Medienlinguistik entsprechen: erstens die grundlegende Frage, wie es mit der Visualität von Sprache selbst aussieht, der Sprache insgesamt, der Lautsprache und der Sprache in Texten (siehe Abschnitt 3.1); zweitens die Frage, wie Intermedialität von Sprache in und mit technischen Medien prozessiert wird, am Beispiel der Herstellung sekundärer Audiovisualität in einer Polit-Talkshow (Abschnitt 3.2).

Zuvor will ich aber kurz andeuten, wie sich im Zuge der pragmatischen Neuorientierung der Sprachwissenschaft zunehmend Medienfragen geradezu aufgedrängt haben, so dass sich allmählich so etwas wie eine Medienlinguistik herausgebildet hat (Abschnitt 2). Dabei werde ich keinen Gesamtüberblick geben, sondern mich auf die Perspektive des IDS fokussieren, das immerhin gelegentlich Mediales aufgegriffen hat, in den letzten Jahren sogar verstärkt, wovon nicht nur die beiden Tagungen 1999 (siehe Kallmeyer (Hg.) 2000) und 2009 (Deppermann/Linke (Hg.) 2010) mit expliziten und ausführlichen Thematisierungen zeugen, aber auch zuvor schon und daneben zahlreiche Aktivitäten.

2 Von der Pragma- zur Medienlinguistik

Parallel zur Weiterführung der Publizistik zu „Medienwissenschaften“ bzw. zu Kommunikationswissenschaften, entwickelt sich auch in der Sprachwissenschaft ein Interesse für Medien: zunächst im Zuge der Erweiterung der linguistischen Perspektive auf Kommunikation und kommunikatives Handeln, vor allem als Interesse an *Massenmedien*. Noch im Register des Lexikons der germanistischen Linguistik (Althaus/Henne/Wiegand (Hg.) 1980) verweist der Ausdruck *Medium* ausschließlich auf das indogermanische grammatische Genus verbi zwischen Aktiv und Passiv. Es gibt darin allerdings schon den Artikel von Erich Straßner zur „Sprache in Massenmedien“ (Straßner 1980), 1984 erscheint der Götschen-Band von Harald Burger „Sprache der Massenmedien“ (Burger 1984), 1987 der Artikel von Ulrich Schmitz im HSK-Band zur Soziolinguistik unter der Überschrift „Sprache und Massenkommunikation“ (Schmitz 1987). Medien, d.h. Massenmedien, waren allerdings höchstens in zweiter Linie wegen ihrer spezifischen Medialität interessant, eher als ein ‚Kommunikationsbereich‘ wie Politik, Sport oder Wirtschaft, der zweifellos von großer gesellschaftlicher Bedeutung war, vor allem das Leitmedium Fernsehen. Andere Massenmedien wie Buch (etwas für Literaturwissenschaftler) oder Kino (zu bildlastig) waren im Abseits.

Das Hauptaugenmerk in dieser Phase lag einerseits auf Fragen der Sprachentwicklung, also auf einer sprachhistorischen Perspektive, zum andern auf Gattungsfragen, die zur sich ebenfalls gerade herausbildenden Text- und Gesprächslinguistik passten. Entsprechend kamen erste Aktivitäten am IDS aus der „Kommission für Fragen der Sprachentwicklung“ in zwei Kolloquien der Jahre 1990 und 1991 mit den bezeichnenden Themen „Mediensprache nach 1945“ und „Medientexte – gestern und heute“, zusammengefasst im Sammelband „Sprache in den Medien nach 1945“ (Biere/Henne (Hg.) 1993).

Aus einer linguistischen Binnensicht war die Frage der Medialität manchmal auf die Differenz ‚gesprochen/geschrieben‘ bzw. ‚Mündlichkeit/Schriftlichkeit‘ reduziert worden und damit auf Aspekte, die mit dem neu erwachten Interesse für gesprochene Sprache und Gespräche und mit der Erweiterung von der Satz- auf die Textebene in die schon erwähnten Teildisziplinen ‚Text- und Gesprächslinguistik‘ überführbar waren. Damit waren zwei Schubladen parat, die so unterschiedlich erschienen, dass man sich im Weiteren um Medialitätsfragen nicht mehr allzu sehr zu kümmern brauchte, zumal man beides nach Prototypikalität konzeptualisierte: Mündlichkeit nach dem Vorbild des informellen *face-to-face*-Gesprächs als ‚Sprache der Nähe‘, Schriftlichkeit als Verdauerung und auch räumliche Zerdehnung nach dem Vorbild formeller Schreiben als ‚Sprache der Distanz‘ (so die berühmte Unterscheidung von Koch/Österreicher 1985). Damit hat man

zwar Wesentliches erfasst, aber man wird mit einer solchen Polarisierung der medialen Vielfalt in tatsächlicher Kommunikation in keiner Weise gerecht. Es gibt eine formelle Mündlichkeit (Versteigerungen, Litaneien, Gerichtsverhandlungen usw.), die keineswegs konzeptionell schriftlich ist, wie eine informelle Schriftlichkeit, die mit sehr eigenen Schriftmitteln arbeitet, wie z.B. Listen, Abkürzungen, Diagrammen, Emoticons u.ä.

In pragmatischer Perspektive war Medialität für die Textlinguistik ohnehin nur ein Teil des so genannten ‚pragmatischen Rahmens‘, also der ‚Situation‘, die nicht zum relevanten Kern des Untersuchenswerten gehörte. Mehr Beachtung fand die Kategorie der ‚sprachlichen Handlung‘, die durch die paradigmaprägende Sprechakttheorie im Zentrum des Interesses war. Entsprechend waren auch Textsortenbegriffe primär funktional definiert, während situative und thematische Faktoren in Mehrebenentypologien lediglich binnendifferenzierend herangezogen wurden. Es gab zwar auch Typologisierungen, die als „Situationsmodelle“ gelten, weil sie nicht handlungszentriert waren, aber – wie Heinemann (2000, S. 540 f.) resümierte –: „Weitaus wichtiger für situativ geprägte Texttypisierungen ist das Bezugnehmen auf die Rahmen-Einordnungsinstanz ‚Kommunikationsbereich‘.“ Entsprechend finden wir im HSK-Band zur Textlinguistik eine Gliederung von „Kommunikationsbereichen und ihren konstitutiven Textsorten“ (Brinker et al. (Hg.) 2000, S. VIII), während das Mediale wieder einmal aus dem Blick gerät.

Am ehesten fand sich textlinguistisches Interesse für Mediales im Begriff der ‚Kommunikationsform‘, den Ermert (1979) in seiner Untersuchung zu Briefsorten geprägt hat und der dann von Brinker in seine Einführung in die Textlinguistik übernommen wurde, allerdings in einer sehr schmalbrüstigen Version mit insgesamt 6 Kommunikationsformen. Demgegenüber kommt Schmitz (2004) der wirklichen Vielfalt mit 13 Kommunikationsformen schon etwas näher.

Inzwischen hat sich der Begriff insofern bewährt, als mit seiner Hilfe klar wurde, dass die neuen computerbasierten Internetdienste kaum wie anfänglich als Textsorten zu fassen waren. Dem trugen Ziegler/Dürscheid (Hg.) (2002) mit dem Buchtitel „Kommunikationsform E-Mail“ ausdrücklich Rechnung. Ebenso ist in der rasch reagierenden Erforschung solcher neuer Kommunikationsformen wie E-Mail, Chat und SMS eine andere Differenz deutlich geworden, indem man sprachliche Gebrauchsweisen bestimmter Nutzergruppen, die anfänglich in der Überzahl waren, vorschnell als Spezifika der Kommunikationsformen ausgab. Genauere Betrachtungen konnten aber nur ergeben, dass sich soziolinguistisch zu verortende Sprachgebrauchstypika nicht den kommunikationsstrukturell geprägten Kommunikationsformen zuschreiben lassen.

Dennoch kann man nicht darüber hinwegsehen, dass die Kommunikationsformen nicht allein durch die technischen Apparate, derer sie sich bedienen,

bestimmt sind. Auf der Basis der technischen Limitierungen und Potenziale bilden sich kulturelle Praktiken heraus, die historisch und sozial verankert sind und entsprechend variieren können.

Solcherlei Beobachtungen münden in ein Plädoyer für die umfassende Beschreibung des Kommunikationsformen-Haushalts moderner Mediengesellschaften und ihrer Vorläufer in der Geschichte. Während man sich (mit vielleicht zu großem Hochdruck) der Beschreibung neuer und neuester Medien zuwendet, hat man viele überkommene Kommunikationsformen in ihrer Komplexität noch keineswegs erfasst. Dies gilt für die hochkomplexe Audiovisualität von *face-to-face*-Kommunikationen, die immer noch ziemlich am Anfang ihrer Beschreibung steht, ebenso wie für die zahlreichen Formen der „Mesokommunikation“ in Veranstaltungen unterschiedlicher Art, wo wir teils körpergebunden, teils mit medialer Unterstützung kommunizieren und so hochkomplexe Wechselwirkungen hervorbringen. Dies zeigt sich nicht nur in den aktuellen Powerpoint-Präsentationen (z.B. Lobin 2009), sondern schon in jedem traditionellen Vortrag oder Unterricht mit Tafeln, Zeigestöcken, Dias, Karten, Overhead-Projektoren, Mikrofonen usw.

Ein anderer, immer noch wenig erforschter Bereich ist die Mesokommunikation im öffentlichen Raum, wo wir ortsgebundene und orts(un)gebundene Kommunikate in großer Vielfalt vorfinden (Domke 2010), die wiederum anderen Regeln ihrer jeweiligen und sehr spezifischen Medialität unterliegen.

Dabei ist auch zu berücksichtigen, dass mit der immer weiter um sich greifenden Digitalisierung einerseits traditionelle Medien selbst ihren Charakter verändern, wie der Übergang vom Festnetztelefon zu mobiler und dann digitaler Telefonie zeigt, andererseits dass Übergänge von traditionellen Medien in neue Medien (Zeitung, Fernsehen oder Radio im Internet) jene und deren Nutzung verändern, wie schließlich auch schon die neuen Medien wie Chat oder E-Mail sich unter den Bedingungen ihrer Einbindung in Social Networks wie „Facebook“ verschieben. Dabei kommt es immer darauf an, nicht nur die veränderten technischen Strukturen, sondern auch die damit verbundenen kulturellen Praktiken ihrer Nutzung zu beschreiben.

3 Medienlinguistik

Inzwischen ist die Medienlinguistik ein breit ausgebautes und durchaus etabliertes Feld, das sich von Anfang an nicht auf Text- oder Gesprächsanalysen (also in der Terminologie der Medienwissenschaft auf Produktanalysen) beschränkt hat, sondern schon früh auch Produktions- und Rezeptionsgesichtspunkte einbezo-

gen und vorangetrieben hat. Hier kann deshalb nicht einmal im Ansatz ein Überblick über die medienlinguistische Forschung gegeben werden. Ein grober Verlauf gewisser Entwicklungslinien lässt sich aber durchaus aus den z.T. schon erwähnten IDS-Publikationen ablesen.

Nach dem schon erwähnten Band von Biere/Henne (Hg.) (1993) erschienen noch zwei weitere Bände, die auf Kolloquien der IDS-Kommission für Fragen der Sprachentwicklung aus den Jahren 1994–1996 zurückgehen: Biere/Hoberg (Hg.) (1996) „Mündlichkeit und Schriftlichkeit im Fernsehen“ und Holly/Biere (Hg.) (1998) „Medien im Wandel“. Man mag an diesen Bänden ablesen, wie man sich in diesem Rahmen nach einer eher sprachhistorischen Perspektive mit einem Schwerpunkt auf Textsorten dann zunächst stärker auf Sprachmedialität fokussierte, immer noch in einem Massenmedium, um schließlich den Blick medienhistorisch zu weiten und auch andere Medien wie Telefon und Computer einzubeziehen. Nebenbei: Inzwischen hat das IDS das Medienthema auch international unterstützt durch die Mitwirkung an einer Konferenz in Moskau 2008, mit einer daraus hervorgegangenen Publikation (Volodina 2013).

Ich möchte hier den sehr bruchstückhaften Überblicksansatz abbrechen und mich stattdessen dem ersten der beiden angekündigten exemplarisch zu behandelnden Stränge zuwenden, der unhintergehbaren Medialität der Sprache selbst, die mit einem technischen Medienbegriff gelegentlich aus dem Blick geraten ist. Deshalb soll hier zunächst grundsätzlich und eher referierend auf die Medialität von Sprache selbst hingewiesen werden, vor allem in der Perspektive des Zusammenhangs von Sprache und Visualität. Danach soll es um die Intermedialität von Sprache in technischen Medien gehen, und zwar am Beispiel der Audiovisualität in einer TV-Talkshow.

3.1 (Inter-)Medialität I: Sprache und Visualität

Bisher war überwiegend von dem zweiten – von der Dynamik der Entwicklung technischer Medien angeregten – Strang der Medienlinguistik die Rede, wo entsprechend technische Medien im Fokus sind, weniger die Medialität der Sprache selbst. In deren Zusammenhang ist Medialität – wie schon erwähnt – zumeist in der Dichotomie von ‚gesprochen‘ und ‚geschrieben‘ gefasst worden, und zwar anhand der Modi ‚auditiv‘ vs. ‚visuell‘. Demgegenüber ist schon seit einiger Zeit darauf hingewiesen worden, dass Sprache in dreierlei Form existiert. So spricht z.B. Jäger (2001) von „Sprache als audiovisuelle(m) Dispositiv des Medialen“ und sammelt Argumente für die These, dass nicht die gesprochene Sprache, sondern die Gebärdensprache primär sei. Schrift ist also nicht die einzige visuelle Sprachform.

Darüber hinaus hat Ellen Fricke in ihrer Habilarbeit gezeigt, dass die Lautsprache „primär audiovisuell“ ist, und argumentiert, dass man bei Gesten nicht mehr von „nonverbaler Kommunikation“ sprechen sollte; sie hält die Klassifizierung von Gesten als „nichtsprachlich“ für falsch (Fricke 2012, S. 38). Stattdessen reklamiert sie für die Lautsprache „sprachliche Multimodalität“, weil für redebegleitende Gesten gilt: „Sie sind mit der begleiteten Rede eng verbunden im Hinblick auf die zeitliche Organisation und teilen mit ihr gemeinsame semantische und pragmatische Funktionen“ (ebd., S. 46). In einem englischsprachigen Handbuchartikel (Fricke 2013) wird entsprechend der „spoken language“ nicht nur ein „verbal level“ zugeordnet, der „auditory“ ist, sondern auch ein „gestural level“, der „visual“ operiert (siehe Abb. 2).

Abb. 2: Mediality and linguistic multimodality (aus: Fricke 2013, S. 783)

Übrigens muss auch auf diesem Feld, das einerseits aus der Gebärdensprach- und Gestenforschung, andererseits aus der Konversationsanalyse hervorgegangen ist, IDS-Forschung erwähnt werden: Ich nenne nur den Sammelband von Reinhold Schmitt (Hg.) (2007) „Koordination. Analysen zur multimodalen Interaktion“, der auf ein IDS-Kolloquium von 2005 zurückgeht.

Dass Schrift visuell ist und dementsprechend auch Texte visuell sind, ist – wie am Anfang schon bemerkt – trivial. Dennoch ist die Visualität von Texten eine weitere Betrachtung wert (siehe auch Holly (Hg.) 2013). Und zwar deshalb,

weil es darum geht, Schrift und Text nicht einfach als sichtbares Äquivalent der Lautsprache zu verstehen, sondern den medialen Eigensinn zu verstehen, den diese Sichtbarkeit mit sich bringt und den Sybille Krämer (2003, 2010; Krämer/Cancik-Kirschbaum/Totzke (Hg.) 2012) mit dem Begriff der ‚Schriftbildlichkeit‘ belegt. Sie beschreibt Schrift als „intermediales Phänomen“, das „eigenständig [...] gegenüber der gesprochenen Sprache wie gegenüber dem gewöhnlichen Bild“ sei (Krämer 2010, S. 19).

In diesem Zusammenhang ist der Peircesche Begriff der ‚Diagrammatik‘ wieder aufgegriffen worden, der dem gesamten Bereich von Schriften, Graphen, Diagrammen, Tabellen, Listen usw. ein semiotisches Potenzial zuschreibt, das sowohl ikonisch wie symbolisch ist und das für die Beschreibung von Texten selbst, aber auch in der Analyse der Funktion von Bildern und Diagrammen in Texten expliziert werden kann (siehe Bauer/Ernst 2010).

Statt linearer Ketten von Schriftzeichen, die akustischen Sequenzen von Lautzeichen zu entsprechen scheinen, sieht man in Texten Sehflächen (Gross 1994; Schmitz 2011), die nicht nur nach und nach gelesen werden, sondern wie Bilder auch ganzheitlich und wahrnehmungsnah erscheinen. Steinseiffer (2013) fragt deshalb nach Möglichkeiten, Texte als Diagramme aufzufassen. Man erinnert sich an den Ansatz der Social Semiotics, in dessen Zusammenhang Kress/van Leeuwen schon 1996 mit „Reading Images“ wichtige Interpretationskategorien entwickelt haben, nicht zuletzt auch sehr mutige Hypothesen zur Gestaltung von Zeitungsfrontseiten. Bei uns hat sich unter dem Stichwort ‚Textdesign‘ eine Forschungsrichtung entwickelt, die auch auf die Orientierungsfunktion der Textgestaltung achtet (z.B. Bucher 1996).

Weitergehend noch ist die Pointe der Diagrammatologie, dass mit Diagrammatischem auch ein epistemisches Potenzial von Schrift und Text verbunden ist, das über die Darstellung hinaus eine operationale kognitive Ressource enthält, also eine eigensinnige Art der Semiose und damit auch des Denkens ermöglicht, eine Art von „Verfertigung der Gedanken beim Schreiben bzw. Textgestalten“. Zwischenräume, Anordnungen, all das, was man Mikro- und Makrotypographie nennt, spielen dabei keine geringe Rolle; sie zeigen jedenfalls, dass die Visualität von Texten über die triviale Tatsache, dass sie mit den Augen aufgenommen werden, weit hinausgeht.

Wir können unserem eingangs betrachteten Plakat, selbst wenn wir die Schrift nicht lesen können, eine ganze Menge an Bedeutung entnehmen (siehe Abb. 3a), z.B. was die Gewichtung der Informationen angeht, was nicht allein an den grafischen und farblichen Elementen hängt, sondern auch an der Anordnung, Größe und Gestaltung der Schriftzeichen. Und selbst einer unscharfen Bleiwüste wie der daneben (siehe Abb. 3b) sehen wir an, dass es sich um ein Inhaltsverzeichnis handelt (Steinseiffer 2013).

Abb. 3a und 3b: Poster und Inhaltsverzeichnis verwischt (letzteres aus Steinseifer 2013, S. 10)

Was zu zeigen war: Sprache, ob gesprochen oder geschrieben, ist – vielleicht von wenigen Sonderverwendungen abgesehen – intermedial, mit Visuellem, das immer auch ein ikonisches Potenzial hat, verbunden.

3.2 (Inter-)Medialität II: Vom „Machen“ der Audiovisualität

In technischen Medien ist der Zusammenhang von Sprache und Visualität vor allem unter zwei Aspekten betrachtet worden, als Text-Bild-Zusammenhang in Printmedien, von dem bisher auch schon die Rede war, und als (technische/ sekundäre) Audiovisualität, von der ich jetzt ein Beispiel behandeln werde.

Eine der immer noch aktuellen Fragen der Medienlinguistik, die auch das Leitthema der Jahrestagung 2009 war (Deppermann/Linke (Hg.) 2010), ist also das intermediale Zusammenspiel mit Sprache; wie gehen Stimme und Schrift, Bild und Ton in einen gemeinsamen Prozess der Semiose ein. Dabei sind allgemein immer zweierlei Beziehungen möglich: interkodale, also das Zusammenspiel verschiedener Zeichenarten, z.B. Sprach- und Bildzeichen, und/oder intermodale, d.h. das Zusammenspiel verschiedener Sinne, vor allem des auditiven und des visuellen Sinnes. Während wir die primäre Audiovisualität der *face-to-*

face-Interaktion im Alltag meistens routiniert hervorbringen und verstehen, treffen wir in technischen Medien auf alle Komplexitätsstufen einer artifiziell und in Teilen sehr kontrolliert hervorgebrachten Intermedialität, die aber – mit dem bekannten Phänomen der Transparenz von Medien – von den Rezipienten gleichwohl wie „naturwüchsig“ oder besser „kulturunspezifisch“ aufgenommen und verarbeitet wird, obwohl man sich natürlich prinzipiell über die Gemachtheit des Ganzen im Klaren ist.

Von der möglichen Komplexität solcher intermedialen Verflechtungen möchte ich hier einen Eindruck geben anhand einer kleinen Workplace-Studie, in der ich mich mit der Produktionsseite einer TV-Talkshow beschäftigt habe. Ich habe dazu im Jahre 2011 in der Bildregie des ZDF-Studios in Berlin während der Herstellung einer „Maybrit Illner“-Sendung als teilnehmender Beobachter Tonaufnahmen gemacht, um mehr darüber zu erfahren, wie die Kameraarbeit gesteuert und koordiniert wird. Ich kann hier natürlich nur kleine Eindrücke von diesem Setting geben und werde mich auf die beiden Aspekte konzentrieren, wie (1) generell Kamerabilder gesteuert werden und (2) wie außerdem zusätzliches großformatiges Bildmaterial auf einer Leinwand hinter den Talkshowteilnehmern eingespielt wird. Meine Leitfrage ist dabei, wie und wie weit überhaupt das komplexe semiotische Material kontrolliert werden kann, inwiefern man also wirklich vom ‚Machen‘ sprechen kann, oder ob nicht eigentlich ein semantisches multikodales und multimodales Rhizom wuchert, das niemand so recht überblicken kann, trotz hochprofessioneller vielseitiger und ständiger Kontrollversuche eines aufwändigen Produktionsapparats.

3.2.1 Zur Maybrit-Illner-Sendung vom 26. Mai 2011

Im Vorfeld der angekündigten neuen Talkshow-Struktur in der ARD hatte das ZDF ihrem Talkshow-Flaggschiff, der Donnerstagabend-Sendereihe ‚Maybrit-Illner‘, ein neues Outfit gegeben. Dazu wurde der Schauplatz aus dem Innenhof des Berliner ZDF-Gebäudes in ein Studio zu ebener Erde verlegt, mit zwei flankierenden Publikumstribünen, einem nach vorne einladend offenen Rundtisch (anstelle der früheren Sichelsitzanordnung ohne Tische), und vor allem mit einer großen Bildwand im Hintergrund, die als Projektionsfläche Raum für zusätzliches semiotisches Material bietet (siehe Abb. 4). Zusammen mit insgesamt sieben Kameras (vier fahrende, eine Steadycam, eine am Kran und eine von der Decke hängende für den Schuss auf Maybrit Illners Tablet, von dem aus die Einspieler gestartet werden können) ergibt dies eine gesteigerte Komplexität, die auch gesteigerte Anforderungen an das gesamte Team stellt, besonders natürlich an die Bildregie.

Abb. 4: Setting „Maybrit Illner“

In der dokumentierten Sendung wurde der damals aktuelle Prozess gegen den TV-Moderator Jörg Kachelmann und seine mediale Begleitung thematisiert. In der Abbildung 4 sieht man auch die Sitzverteilung der Protagonisten, von links: Rüdiger Bagger (ehem. Staatsanwalt), Alice Schwarzer („Emma“-Herausgeberin), Gerhard Strate (Staranwalt), Hans-Hermann Tiedje (ehem. „Bild“-Chef), Mark Benecke (Kriminalbiologe); in der ersten Reihe links sitzt (hier nicht sichtbar) außerdem ein Gast, der zwischendurch an einem Stehpult befragt wird: Sabine Hartwig (als Vertreterin der Opfer-Organisation „Weißer Ring“).

3.2.2 Der semiotische Beitrag der Kameraszenierung

In verschiedenen Arbeiten (Holly 2010, 2012, ersch. demn.) habe ich verschiedene Funktionen herausgearbeitet, die als semiotischer Beitrag der Kamerainszenierung in Polit-Talkshows verstanden werden können. Dabei verwende ich den Begriff der ‚Transkriptivität‘, den Ludwig Jäger (2002) zur Theorie der Bedeutungsgenese durch intramediale und intermediale Bezugnahmepraktiken entwickelt hat. So werden Zeichen derselben oder unterschiedlicher Medialität aufeinander bezogen, um „anders lesbar“ (oder gerade unlesbar) gemacht zu werden, durch Paraphrase, Explikation, Erläuterung, Kommentierung, Übersetzung usw. Nach dieser Logik kann man auch die Beziehungen zwischen Wortbeiträgen und Bildkomponenten in solchen Talkshows verstehen, so dass eine These der Bild-Sprach-Transkription lauten könnte: Die Kameraführung in Polit-Talkshows generiert durch die Selektion von Einstellungen und Umschnitten Bedeu-

tungskomponenten, die sprachliche Äußerungen „transkribieren“, d.h. überformen, implizit kommentieren und dadurch „anders lesbar“ machen. Brisant ist vor allem, dass so die alleinige Auktorialität dem Sprecher entzogen wird und er die „performative Letztfassung“ seines Beitrags nicht mehr selbst kontrolliert, sondern sich partiell an Instanzen technischer Medialität ausliefert, die nahezu unmerklich auch auf den Rezipienten wirken können.

Zusammenfassend kann man als die drei wesentlichen Funktionen der Kamerainszenierung ansehen (siehe dazu Holly 2010, 2012, ersch. demn.): (1) die visuelle Profilierung des jeweiligen Sprechers; (2) Abwechslung, Gliederung, Strukturierung des thematischen Ablaufs; (3) die Kontextualisierung von Beteiligungsrollen und Frames, indem Einstellungen bestimmter Personen als implizite Kommentierungen zum gerade Gesagten oder als implizite Hinweise auf bestimmte Themen oder semantische Frames verstanden werden können, für die sie stehen.

3.2.3 Komplexität, Automatismen und Kontrollversuche

Geht man davon aus, dass Bilder und Sprachbeiträge sich *wechselseitig* transkribieren, gilt auch das Umgekehrte. Keine Bildregie kann vollständig vorhersehen, wie das Gesagte die Bilder überschreibt. Alle Beteiligten müssen also damit rechnen, dass es zu ‚Automatismen‘ kommt, so nennen Bublitz et al. (2010, S. 9) „Abläufe, die sich einer bewussten Kontrolle weitgehend entziehen“. Mit solchen Automatismen ist generell jeder Sprecher ständig konfrontiert, erst recht aber im Kontext technischer Medialität (siehe auch Holly 2012). Dies gilt *handlungstheoretisch*, insofern Handlungen zwar prinzipiell kontrollierbar sind, aber nicht immer kontrolliert (so dass wir es gelegentlich mit versehentlichen und fahrlässigen Handlungen zu tun haben). *Sprachtheoretisch* gesehen sind die Sprecher nicht ganz frei, sondern an die sozialen Gestalten einer jeweils vorfindlichen Sprachgemeinschaft gebunden, die Sprache als „invisible-hand“-Phänomen hervorbringt. *Zeichentheoretisch* betrachtet, muss jeder Sprecher immer symptomhafte Anteile seiner Äußerungen bedenken, die z.B. durch die Leiblichkeit der Stimme und der weitgehend routinierten und automatisierten Körpersprache nur begrenzt kontrollierbar sind. *Medientheoretisch* kommt mit Techniken unterschiedlicher Komplexität ein jeweiliges Inszenierungsdispositiv hinzu, das den Protagonisten und Machern ein doppeltes Problem auferlegt, insofern sie nicht nur mit den eigenen, sondern auch mit anderen Unkontrollierbarkeiten konfrontiert werden. Insgesamt kann man also eine derart komplexe Live-TV-Sendung als ein multikodales, multimodales und multiauktoriales Textgeflecht ansehen.

Deshalb gilt grundsätzlich: In Talkshows überlagern sich die körpergebundenen Sprach- und Körpersprachzeichen mit den technisch kodierten Ton- und Bildartefakten, so dass das Gesamtergebnis ein relativ komplexes audiovisuelles Geflecht von transkriptiven Spuren darstellt, das niemand ganz und gar kontrolliert, auch wenn einzelne Elemente mit mehr oder weniger Berechtigung der Autorschaft von jeweiligen Akteuren zugeschrieben werden können. Keiner überblickt je, was der andere gleichzeitig transkriptiv unternimmt, geschweige denn, dass die möglichen Transkriptqualitäten einhellig beurteilt werden. Unmittelbar anschaulich ist aber für den reflektierten Beobachter der ständig wachsende Kontrollversuch der Macher, die eine perfekte Inszenierung anstreben.

Für die dokumentierte Sendung lassen sich vielfältige Elemente einer solchen Inszenierungskontrolle festhalten. Zunächst gibt es von der Redaktion im Vorfeld erarbeitete Ablaufpläne, also ziemlich detaillierte Skripts mit vorformulierten Fragen der Moderatorin, eine ziemlich genaue Profilierung der eingeladenen Gäste, die schon in der Vorstellung und manchmal auch in zusätzlichen ‚Einspielen‘ auf bestimmte inhaltliche Positionen und Argumente zuspitzend festgelegt werden, so dass man sie im Sinne einer bestimmten Dramaturgie an bestimmten Stellen abrufen kann. Es werden sogar mit studentischen Hilfskräften als Stellvertretern der Gäste Probeläufe veranstaltet, die nicht nur der Vorbereitung bestimmter Kamerapositionen und Einstellungskonfigurationen dienen, sondern auch das Skript unter Studiobedingungen zu testen erlauben. Zur Koordination des großen differenziert und spezialisiert beitragenden Teams dient eine diskret funktionierende sogenannte „offene Regie“, bei der alle wichtigen beteiligten Macher im Studio durch „In-Ear-Monitoring“ untereinander und mit der Bildregie verbunden sind. Sie verständigen sich fortlaufend und besonders an strukturell wichtigen Schaltstellen über das Antizipieren möglicher Fortsetzungen, so dass jeder jederzeit optimal vorbereitet ist und eine rasche und flexible Korrektur bei unvorgesehenen Abläufen und erst recht bei dennoch unvermeidlichen kleinen und größeren „Schnitzern“ möglich wird. Zum erhöhten Kontrollaufwand tragen auch die vorbereiteten Einspieler und nun im „neuen“ Setting auch die Bildwandprojektionen bei, die ad hoc aufmontiert werden. Um die Frage, wie dieses Bildwandmaterial eingesetzt und der Einsatz jeweils prozessiert wird, soll es nun im nächsten Abschnitt gehen. Zuvor werde ich noch an einem Beispiel der Arbeit in der Bildregie zeigen, wie die visuelle Profilierung von Sprechern gesteuert wird.

So war dem Regisseur aufgefallen, dass einer der Gäste auffällig tätowierte Hände hat, die zur optischen Charakterisierung bei Gelegenheit ins Bild genommen werden sollen. Deshalb weist er den für diesen Gast üblicherweise zuständigen Kameramann präventiv an:

„F. [Kamera 3] wenn de mal die hände von dem äh Benecke bekommst mit den ganzen tattoos ja wenn er redet genau [...] jetzt nicht [...] wenn er demnächst mal redet [...]

Als es dann so weit ist, lautet die Regieanweisung, die zu der entsprechenden Einstellung (Abb. 5) führt:

„jetzt Benecke hände s hat er se irgendwie grad noch unten [...] und jetzt die – wart ich sags dir genau ---- die drei drei -- zieh auf -- oder hoch – genau“

Abb. 5: Benecke Hände

An diesem Beispiel kann man erkennen, wie die Live-Sendung dadurch lebt, dass man optimal vorbereitete Bildinszenierungen in der Situation exakt prozessiert – und dabei flexibel bleibt. Der Regisseur hatte sich die Auflösung der GroßEinstellung auf die Hände nämlich als Rückwärtszoom vorgestellt („zieh auf“), ist dann aber auch mit der vom Kameramann gewählten Lösung, einem Aufwärtsschwenk einverstanden („oder hoch“).

3.2.4 Die Bildwand

Ein Großteil der Aufmerksamkeit in der Bildregie geht um das Management der adäquaten zusätzlichen Bebilderung auf der noch relativ neuen Projektionswand.

Dafür sind mehrere statische Bilder vorbereitet, die dann an passenden Stellen eingeblendet werden können. In der dokumentierten Sendung sind das:

1. die sogenannte „Fotostrecke“: eine Collage von Zeitschriftentiteln mit Fotos von und Schlagzeilen zu Kachelmann; sie wird als Standard verwendet, wenn gerade kein anderer Hintergrund durch bestimmte Äußerungen „abgerufen“ wird;
2. das Bild einer goldenen Justitia-Statue vor einem blauen Himmel;
3. drei Aufnahmen von Kachelmann aus dem Kontext des laufenden Prozesses: je eine mit den beiden Verteidigern, eine auf einem Flur und eine beim Aussteigen zwischen Gefangenentransportwagen und Gerichtseingang;
4. ein Bild des französischen Politikers Dominique Strauss-Kahn, der wie Kachelmann einer Vergewaltigung beschuldigt worden war.

Die Bilder sind also durch den thematischen Frame der Sendung motiviert und liefern selbst semantische Kontextualisierungen, die sie dem gesprochenen Text hinzufügen, allerdings ohne Wissen und Zutun der Sprecher. Sie funktionieren – wie die Kameraeinstellungen selbst – als optische Kommentare und formieren einen eigenen visuellen Frame. In unserem Fall thematisieren sie in erster Linie die mediale Aufbereitung des Kachelmann-Themas, die Prozess-Rolle Kachelmanns als Beschuldigter/Angeklagter, und zwar im Rahmen der allegorisierten Justiz, die ebenfalls problematisiert wird, und schließlich den Parallellfall des Prominenten Strauss-Kahn als Zentralfigur eines Sexskandals.

Diese visuellen Frame-Elemente werden ihrerseits in der Sendung durch verbale Elemente ausgelöst; für solche Frameaktivierungen unterscheidet Fillmore (1985, S. 232, zitiert nach Ziem 2008, S. 232) zwischen „invoked“ und „evoked“, je nachdem ob es sich um eine Interpretationsleistung des Rezipienten handelt, die gewissermaßen von außen an den Text herangetragen wird, oder einen konventionell mit dem Text assoziierten Frame. In unserem Fall werden die Bilderframes außer dem Standardbild des Medienkontextes vom Sprecher ‚evoziert‘, denn alle Nicht-Standardbilder sind durch Verbaltextstellen motiviert.

Hier ein Beispiel: In einer Äußerung von Alice Schwarzer, die vom Kachelmann-Prozess für die Bild-Zeitung berichtet hat und sich als Frauenrechtlerin dabei eindeutig gegen Kachelmann positioniert hat, wird der Name Strauss-Kahn genannt. Dies führt nahezu reflexartig zu einer entsprechenden Anweisung des Bildregisseurs (+, # für zeitgleiche Einsätze; // für Bildumschnitte):

Sprechertext	Bildregie
jede zweite vergewaltigung passiert innerhalb einer // beziehung also durch den eigenen ehemann oder freund und wie sie eben schon sachten wird dann natürlich // die beweislage ganz schwierig herrn strauss-kahn wird es schwerfallen + äh überzeu-gend darzulegen dass ihn äh die schwarze putzfrau innerhalb einer halben stunde überfallen hatund sich // sein glied inn mund gesteckt hat und dann extra schreiend rausgelaufen ist das wird schwierig aber wenn // eine beziehung is kann # natürlich	+ strauss-kahn dsk vor gericht wechsel jetzt? ja wechsel – okay achtung für die drei und die drei # da isser ja

Hier braucht die Kamerainszenierung zwei Schritte. Nachdem der Regisseur mit deutlich erhobener Stimme den Namen *Strauss-Kahn* wiederholt und mit dem „Codewort“ für das vorbereitete Bild „DSK [für Dominique Strauss-Kahn] vor Gericht“ und der fragend intonierten Vorbereitung „Wechsel jetzt?“ und der entsprechenden Bestätigung „ja Wechsel“ das Bild auf die Bildwand hat montieren lassen, wird in einem zweiten Schritt auf die Kamera umgeschnitten („die drei“), in deren Perspektive die Bildwand auch für den Zuschauer sichtbar ist. Dies wird zufrieden quittiert: „Da isser ja“ – das Manöver hat geklappt (siehe Abb. 6).

Abb. 6: DSK vor Gericht – „Da isser ja“

Nicht immer klappt das komplexe Spiel reibungslos. Es gibt nicht nur die Möglichkeit, vorbereitete Standfotos zu projizieren, sondern auch bewegte Kameraeinstellungen; davon wird allerdings seltener Gebrauch gemacht, so dass es in der folgenden Episode zu einem Missverständnis kommt:

R: *guck mal die quatschen noch mal in der 3 da gemeinsam* – [C: *und die 4*] -- *äh die 2 mal auf die wand -- äh die 2 auf die WAND warum machst du denn auf J. dann in dem moment acht- ---wieder zurück zurück zurück 1 zufahrn 1 --- und maz ab*

Hier will der Regisseur (R) das Bewegtbild der Kamera 2 „auf die Wand“ montiert haben, während die Cutterin (C) den Routine-Umschnitt erledigt; der Kameramann (J) versteht die Äußerung aber als Aufforderung, die Wand ins Bild zu nehmen und macht einen entsprechenden Rückwärtszoom. Das löst eine irritierte Reaktion aus („Warum machst du denn auf?“) und veranlasst eine Korrektur („wieder zurück zurück zurück“). Die Situation wird gerettet, weil ohnehin ein Einspieler vorgesehen war („MAZ ab“). Während nun der Einspieler läuft und alle erst mal etwas entspannen können, gibt es eine kleine klärende Nachverbrennung:

R: *J. warum fährst du w was machst du da wenn ich sage auf die wand*

J: *die 2 auf die wand zu*

R: *die zwei wollten WIR auf die wand doch nicht DU auf die wand sondern wir machen dein bild auf die wand* [lacht]

Wie man sieht, handelt es sich um ein klassisches Missverständnis, ausgelöst durch die Polysemie der Kurzformel „auf die Wand“, die keinen Agens hat und der eine explizite Adressierung fehlt; gemeint war die Anweisung für denjenigen Mitarbeiter, der Bilder auf die Bildwand setzt, nicht für den Kameramann, der dachte, er solle die Wand ins Bild nehmen. Kein Zuschauer konnte diese Panne bemerken, es ist lediglich eine Idee des Regisseurs nicht umgesetzt worden. Die Komplexität der Abläufe und der hohe Zeitdruck machen solche kleinen Pannen durchaus wahrscheinlich. Umso wichtiger ist eine optimale Vorbereitung, die ständige Antizipation des nächsten Schrittes und die dauernde Konzentration des gesamten Teams.

Die Live-Performanz erfordert ein optimal vorbereitetes und gut eingespieltes Team, das einerseits feste Praktiken, andererseits flexible Manöver beherrscht, um nicht vorhersehbare Züge begleiten zu können. Durch die getrennte Hervorbringung und technisch kombinierte Inszenierung des semantisch heterogenen Materials entstehen komplexe Bedeutungsgeflechte als Spuren, teils musterhaft und intendiert, teils musterhafte Automatismen.

Während das verbale „Framing“ weitgehend in der Kontrolle der Protagonisten bleibt, sind die kamera-inszenierten Frameinvokationen und -evokationen von ihnen nicht kontrollierbar. Auch die Inszenierer können nicht beliebiges Bildmaterial einspielen, sondern arbeiten im nur partiell vorhersehbaren Ablauf live/online jeweils reaktionsschnell und nach aktuellen Erfordernissen. Umso mehr sind sie auf allgemeine und leicht nachzuvollziehende Wissensstrukturen angewiesen, können aber Perspektivierungen einbringen, indem sie das transkriptive Potenzial nutzen.

4 Fazit

Der Blick auf die Herausbildung einer Medienlinguistik führt auf die Spur von Intermedialität, vor allem von Sprache und visuellen Kodes. Sowohl Lautsprache als auch Schrifttexte erweisen sich bei genauerer Betrachtung als multimodal bzw. intermedial.

Für die Lautsprache gilt, dass sie üblicherweise mit visuellen Gesten kombiniert auftritt. Schrifttexte werden nicht nur häufig mit Bildern und Diagrammen zusammengebracht, sondern enthalten selbst als ‚Sehflächen‘, durch Makro- und Mikrotypografie nur visuelle Zeichenqualitäten, die deshalb auch nicht vorgelesen werden können.

In technischen Kontexten sekundärer Audiovisualität sind die intermedialen Bezüge zwischen Lautsprache und Visuellem vielfältig, wie das Beispiel einer Polit-Talkshow gezeigt hat. Der Produktionsprozess der visuellen Inszenierung offenbart den Versuch, die multiauktoriale und damit ziemlich komplexe Dynamik von zweierlei Performanz zu überblicken – und durch Steuerung der Kameraperformanz auch die transkriptiven Bezüge zu den Sprachperformanzen der Protagonisten. Dazu haben sich Routinen ausgebildet und Konventionen der Funktionalität (Sprecherprofilierung, Strukturierung, Kontextualisierung von Beteiligungsrollen und Frames). Allgemein gilt: Kameraführung in Polit-Talkshows generiert durch die Selektion von Einstellungen und Umschnitten Bedeutungskomponenten, die sprachliche Äußerungen „transkribieren“, d.h. überformen, implizit kommentieren und dadurch „anders lesbar“ machen. Brisant ist vor allem, dass so die alleinige Auktorialität dem Sprecher entzogen wird und er die „performative Letztfassung“ seiner Rede nicht mehr kontrolliert. Bildwandmaterial ist hierfür ein weiteres prominentes Beispiel, indem es seinerseits nicht nur stereotype Frames aufruft, sondern auch zusätzliche Bedeutungskomponenten und Perspektivierungen einbringt.

Es entstehen hochkomplexe Geflechte oder Rhizome von simultan wahrnehmbarem sprachlichem und bildlichem Material, das trotz intensiver Kontrollanstrengungen der Macher von keinem der Beteiligten ganz kontrollierbar ist.

Im Lichte der unhintergehbaren Medialität und Intermedialität von Sprache, die immer auch Visualität einschließt, erst recht aber im Kontext der mannigfachen intermedialen Phänomene in technisch komplexer Kommunikation kann man für die Entwicklung der letzten 50 Jahre festhalten: Die Sprachwissenschaft hat ihren Gegenstand im Hinblick auf Medialität und Intermedialität neu definiert und erweitert und es erscheint mehr als gerechtfertigt, wenn Ulrich Schmitz in einem Aufsatz-Titel fordert: „Warum sogar Sprachwissenschaftler auch Bilder betrachten müssen“ (Schmitz 2005).

Literatur

- Althaus, Hans-Peter/Henne, Helmut/Wiegand, Herbert E. (Hg.) (1980): Lexikon der germanistischen Linguistik. 2., vollst. neu bearb. u. erw. Aufl. Tübingen.
- Bauer, Matthias/Ernst, Christoph (2010): Diagrammatik. Einführung in ein kultur- und medienwissenschaftliches Forschungsfeld. (= Kultur- und Medientheorie). Bielefeld.
- Biere, Bernd U./Henne, Helmut (Hg.) (1993): Sprache in den Medien nach 1945. (= Reihe Germanistische Linguistik 135). Tübingen.
- Biere, Bernd U./Hoberg, Rudolf (Hg.) (1996): Mündlichkeit und Schriftlichkeit im Fernsehen. (= Studien zur Deutschen Sprache 5). Tübingen.
- Brinker, Klaus et al. (Hg.) (2000): Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 1. Halbbd. (= Handbücher zur Sprach- und Kommunikationswissenschaft 16.1). Berlin/New York.
- Bublitz, Hannelore et al. (Hg.) (2010): Automatismen. (= Schriftenreihe des Graduiertenkollegs „Automatismen“). München.
- Bucher, Hans Jürgen (1996): Textdesign – Zaubermittel der Verständlichkeit? Die Tageszeitung auf dem Weg zum interaktiven Medium. In: Hess-Lüttich, Ernest W. B./Holly, Werner/Püschel, Ulrich (Hg.): Textstrukturen im Medienwandel. (= Forum angewandte Linguistik. Publikationsreihe der Gesellschaft für Angewandte Linguistik 29). Frankfurt a.M. u.a., S. 31–59.
- Burger, Harald (1984): Sprache der Massenmedien. (= Sammlung Götschen 2225). Berlin/New York.
- Deppermann, Arnulf/Linke, Angelika (Hg.) (2010): Sprache intermedial. Stimme und Schrift, Bild und Ton. (= Jahrbuch des Instituts für Deutsche Sprache 2009). Berlin/New York.
- Domke, Christine (2010): Texte im öffentlichen Raum. In: Bucher, Hans-Jürgen/Glönz, Thomas/Lehnen, Katrin (Hg.): Neue Medien – neue Formate. Ausdifferenzierung und Konvergenz in der Medienkommunikation. (= Interaktiva. Schriftenreihe des Zentrums für Medien und Interaktivität 10). Frankfurt a.M./New York, S. 257–281.
- Ermert, Karl (1979): Briefsorten. Untersuchungen zur Theorie und Empirie der Textklassifikation. (= Reihe Germanistische Linguistik 20). Tübingen.

- Fillmore, Charles (1985): Frames and the semantics of understanding. In: Quaderni di Semantica 6, S. 222–254.
- Fricke, Ellen (2012): Grammatik multimodal. Wie Wörter und Gesten zusammenwirken. (= Linguistik – Impulse & Tendenzen 40). Berlin/Boston.
- Fricke, Ellen (2013): Towards a unified grammar of gesture and speech: A multimodal approach. In: Müller, Cornelia et al. (Hg.): Body – Language – Communication. An international handbook on multimodality in human interaction. 1. Halbbd. (= Handbücher zur Sprach- und Kommunikationswissenschaft 38.1). Berlin/Boston, S. 733–754.
- Gross, Sabine (1994): Lese-Zeichen. Kognition, Medium und Materialität im Lese-prozess. Darmstadt.
- Heinemann, Wolfgang (2000): Textsorte – Textmuster – Texttyp. In: Brinker et al. (Hg.), S. 507–523.
- Holly, Werner (2010): Besprochene Bilder – bebildertes Sprechen. Audiovisuelle Transkriptivität in Nachrichtenfilmen und Polit-Talkshows. In: Deppermann/Linke (Hg.), S. 359–382.
- Holly, Werner (2012): Transkriptiv kontrollgemindert: Automatismen und Sprach-Bild-Überschreibungen in Polit-Talkshows. In: Conradi, Tobias et al. (Hg.): Schemata und Praktiken. (= Schriftenreihe des Graduiertenkollegs ‚Automatismen‘). München, S. 161–189.
- Holly, Werner (Hg.) (2013): Themenheft Textualität – Visualität. Zeitschrift für Germanistische Linguistik 41.
- Holly, Werner (ersch. demn.): Bildinszenierungen in Talkshows. Medienlinguistische Anmerkungen zu einer Form von „Bild-Sprach-Transkription“. In: Girth, Heiko/Michel, Sascha (Hg.): Multimodale Kommunikation in Polit-Talkshows. Stuttgart.
- Holly, Werner/Biere, Bernd U. (Hg.) (1998): Medien im Wandel. Opladen/Wiesbaden.
- Jäger, Ludwig (2001): Sprache als Medium. Über die Sprache als audio-visuelles Dispositiv des Medialen. In: Wenzel, Horst/Seipel, Wilfried/Wunberg, Gotthart (Hg.): Audiovisualität vor und nach Gutenberg. Zur Kulturgeschichte der medialen Umbrüche. (= Schriften des Kunsthistorischen Museums 6). Wien, S. 19–42.
- Jäger, Ludwig (2002): Transkriptivität. Zur medialen Logik der kulturellen Semantik. In: Jäger, Ludwig/Stanitzek, Georg (Hg.): Transkribieren. Medien/Lektüre. München, S. 19–41.
- Koch, Peter/Österreicher, Wulf (1985): Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In: Romanistisches Jahrbuch 36, S. 15–43.
- Kallmeyer, Werner (Hg.) (2000): Sprache und neue Medien. (= Jahrbuch des Instituts für Deutsche Sprache 1999). Berlin/New York.
- Krämer, Sybille (2003): ‚Schriftbildlichkeit‘ oder: Über eine (fast) vergessene Dimension der Schrift. In: Krämer, Sybille/Bredenkamp, Horst (Hg.): Bild, Schrift, Zahl. Paderborn/München, S. 207–210.
- Krämer, Sybille (2010): Sprache, Stimme, Schrift. In: Linke/Deppermann (Hg.), S. 13–28.
- Krämer, Sybille/Cancik-Kirschbaum, Eva/Totzke, Rainer (Hg.) (2012): Schriftbildlichkeit. Wahrnehmbarkeit, Materialität und Operativität von Notationen. (= Schriftbildlichkeit 1). Berlin.
- Kress, Gunther/van Leeuwen, Theo (1996): Reading images. The grammar of visual design. London/New York.
- Lobin, Henning (2009): Inszeniertes Reden auf der Medienbühne. Zur Linguistik und Rhetorik der wissenschaftlichen Präsentation. (= Interaktiva 8). Frankfurt a.M.

- Schmitt, Roland (Hg.) (2007): Koordination. Analysen zur multimodalen Interaktion. (= Studien zur Deutschen Sprache 38). Tübingen.
- Schmitz, Ulrich (1987): Sprache und Massenkommunikation. In: Ammon, Ulrich/Dittmar, Norbert/Mattheier, Klaus J. (Hg.): Sociolinguistics. An international handbook of the Science of Language and Society. 1. Halbbd. (= Handbücher zur Sprach- und Kommunikationswissenschaft 3.1). Berlin/New York, S. 820–832.
- Schmitz, Ulrich (2004): Sprache in modernen Medien. Einführung in Tatsachen und Theorien, Themen und Thesen. (= Grundlagen der Germanistik 41). Berlin.
- Schmitz, Ulrich (2005): Blind für Bilder. Warum sogar Sprachwissenschaftler auch Bilder betrachten müssen. In: OBST-Redaktion (Hg.): Paradigms lost. (= Osnabrücker Beiträge zur Sprachtheorie 69). Bremen, S. 187–227.
- Schmitz, Ulrich (2011): Sehflächenforschung. Eine Einführung. In: Diekmannshenke, Hajo/Klemm, Michael/Stöckl, Hartmut (Hg.): Bildlinguistik. Theorien – Methoden – Fallbeispiele. (= Philologische Studien und Quellen 228). Berlin, S. 23–42.
- Steinseifer, Martin (2013): Texte sehen – Diagrammatologische Impulse für die Textlinguistik. In: Zeitschrift für Germanistische Linguistik 41, S. 8–39.
- Straßner, Erich (1980): Sprache in Massenmedien. In: Althaus/Henne/Wiegand (Hg.), S. 328–337.
- Volodina, Maja N. (Hg.) (2013): Mediensprache und Medienkommunikation im interdisziplinären und interkulturellen Vergleich. Mannheim.
- Ziegler, Arne/Dürscheid, Christa (Hg.) (2002): Kommunikationsform E-Mail. (= Textsorten 7). Tübingen.
- Ziem, Alexander (2008): Frames und sprachliches Wissen. Kognitive Aspekte der semantischen Kompetenz. (= Sprache und Wissen 2). Berlin/New York.