

John A. Bateman, Chiao-I Tseng

Linguistik und Multimodalität

Die Integration empirischer, semiotischer und hermeneutischer Methoden durch Triangulation

Abstract: In diesem Beitrag skizzieren wir eine breite Sichtweise auf Multimodalität als vielschichtiges, vor allem sozial orientiertes Phänomen, und einige damit verbundene Herausforderungen für die Durchführung von empirischer Forschung. Wir argumentieren, dass die Triangulation über verschiedene Abstraktionsebenen hinweg eine wichtige methodologische Haltung für die Erforschung diverser semiotisch komplexer Artefakttypen darstellt, und geben einige illustrative Beispiele für empirische Forschung, die wir in diesem Sinne durchgeführt haben. Dabei richten wir einen besonderen Fokus auf unsere Arbeiten zum narrativen Film, zum Comic, zur Graphic Novel und auf Studien zur Blickbewegungsaufzeichnung und zu Neuroimaging-Methoden. Wir veranschaulichen dabei jeweils, wie die Anwendung geeigneter sprachwissenschaftlicher Begriffe als Brücke zwischen Theorie und empirischer Forschung dienen kann, woraus sich Konsequenzen sowohl für die theoretische Reflexion als auch für die praktische Operationalisierung dieser Begriffe im Kontext empirischer Studien ergeben. Rückschlüsse auf die Begriffsbildung in der Linguistik werden dann zum Schluss wieder aufgegriffen.

Keywords: ‚non-verbale Kommunikation‘, semiotische Zeichenmodalitäten, Theorieentwicklung, experimentelle Methoden, Materialität, Diskurssemantik, Eyetracking, Comics, Film

John A. Bateman, Universität Bremen, Fachbereich für Sprach- und Literaturwissenschaft, Bibliotheksstraße 1, 28334 Bremen, GERMANY, bateman@uni-bremen.de

Chiao-I Tseng, Universität Bremen, Fachbereich für Sprach- und Literaturwissenschaft, Bibliotheksstraße 1, 28334 Bremen, GERMANY, tseng@uni-bremen.de

1 Einführung: die Beziehung zwischen Linguistik und Multimodalitätsforschung

Wenn man die wahre Natur der Sprache entdecken will, muss man erst einmal dem Beachtung schenken, was sie mit allen anderen Systemen gleicher Ordnung teilt. (Saussure 2001 [1916]: 22)

Wie in mehreren Beiträgen dieses Bandes angemerkt, beruhen kommunikative Situationen zunehmend auf einem Gebrauch von Sprache (gesprochen sowie schriftlich), der explizit und bewusst andere Zeichenarten mit einschließt (vgl. Schmitz 2004, 2016; Kesselheim 2011; van Leeuwen 2020). Die Tatsache, dass Zeichensysteme gemeinsam eingesetzt werden und stets ineinandergreifen, stellt besondere Herausforderungen an den theoretischen Rahmen, an die richtige Wahl der Terminologie und an die empirischen Methoden, die sich für ihre wissenschaftliche Aufklärung als wirksam erweisen können. Dies gilt darüber hinaus nicht nur für die sprachwissenschaftliche Diskussion, sondern für alle Disziplinen, die sich mit einer oder mehreren der verwendeten Zeichenarten beschäftigen.

Für viele Akteure innerhalb der breitgefächerten Multimodalitätsforschung gilt die Linguistik als Hauptinspirationsquelle. Seit den bahnbrechenden Versuchen von Barthes und anderen in den 1960er Jahren, grundlegende, weitgehend strukturelle Prinzipien der linguistischen Analyse, wie etwa das Aufzeigen von Systemen von Unterscheidungsmerkmalen (die sog. *Komponentenanalyse*), auf andere Zeichensysteme anzuwenden, hat es in der Tat eine ständige Erweiterung und Ausdehnung von ‚linguistisch motivierten‘ Analysen auf eine breite Vielfalt von semiotischen Entitäten gegeben. Wenn wir uns an Saussures Vorstellung erinnern, dass die Linguistik als Musterbeispiel und Hauptvertreter für die Untersuchung anderer Zeichensysteme dienen könnte und vielleicht sollte (vgl. Saussure 1916: 80–81 und die Diskussion bei Krampen 1981: 106–107), wäre auf jeden Fall eine rege theoretische und methodische Interaktion zwischen etablierten Bereichen der Linguistik und Versuchen, linguistische Prinzipien und Methoden auf ‚nonverbale‘ Systeme zu übertragen, zu erwarten. Diese fällt jedoch sehr unterschiedlich aus, je nachdem, um welche Bereiche und Untersuchungsgegenstände es sich handelt. Hier ist nicht der Ort, um einen detaillierten Überblick über den Stand dieser Zusammenhänge zu geben, aber es wird nützlich sein, sowohl die Breite dieser Wechselwirkungen zu skizzieren, als auch einige Konsequenzen zu erläutern.

Es gibt ohne Zweifel eine Reihe von Dimensionen, die man für eine Klassifizierung der verschiedenen Ansätze verwenden könnte. Dies hängt auch vom

Zweck ab, für den eine Klassifizierung vorgenommen wird. Dabei ist es wichtig zu betonen, dass wir uns ausdrücklich auf klare Unterschiede bezüglich des Einsatzes von *empirischen Methoden* zwischen verschiedenen Arten von Sprachtheorien und ihrer Entwicklung in einem multimodalen Kontext beschränken wollen. Dies lässt sich sinnvoll erreichen, indem wir zunächst eine Differenzierung hinsichtlich sprachlicher Abstraktionsebene und Art des betrachteten Phänomens treffen. Wir trennen ‚textlinguistische‘ (oder ‚Textebene‘) Ansätze, grammatikzentrierte Ansätze und Ansätze, die sich mit kommunikativer Interaktion befassen. Darüber hinaus ist es insbesondere für die textlinguistischen Darstellungen sinnvoll, Entwicklungen im deutsch- und englischsprachigen Raum separat zu berücksichtigen, da diese sich auf recht diverse theoretische Positionen stützen und die Ansätze zur Multimodalität unterschiedlich beeinflusst haben. Es wird hier kein Anspruch auf Vollständigkeit erhoben: Unser Ziel ist es, die Breite der Palette der möglichen Beziehungen zwischen multimodalem Theoretisieren und korrespondierender linguistischer Theoriebildung aufzuzeigen. Dies wird zur Diagnose möglicher Lücken und vielversprechender weiterer Entwicklungen genutzt werden.

Zunächst haben sich deutschsprachige textlinguistische Ansätze, die bis in die Multimodalität hineinreichen, auf Text-Bild-Relationen konzentriert. Dabei stützen sie sich im Wesentlichen auf pragmatische und handlungstheoretische Darstellungen einerseits (vgl. Spillner 1982; Stöckl 2004a; Bucher 2011; Holly 2009) und die Entwicklung verschiedener Textsortentypologien andererseits (vgl. Schmitz 2016, Meier 2014). Die Fokussierung auf Gebrauchssituationen führt ziemlich direkt zu Überlegungen zur rhetorischen Organisation, zu Argumentationsstrukturen und Topoi und zum Einfluss des Genres. Die für Multimodalitätsstudien wohl einflussreichste Richtung der englischsprachigen Textlinguistik ist demgegenüber der Diskursansatz der ‚Systemisch-Funktionalen Linguistik‘, wie er für Multimodalität maßgeblich in Kress & van Leeuwen (2006 [1996]) und O’Toole (2011 [1994]) artikuliert wird. Diese Modelle orientieren sich stark an der systemisch-funktionalen Charakterisierung von Grammatik und enthalten infolgedessen viele Begrifflichkeiten, die direkt aus funktionalen Beschreibungen grammatischer Phänomene abgeleitet werden. Hervorzuheben sind hier auch Diskussionen, die auf verschiedenen Beziehungen auf Textebene basieren, wie z.B. Kohäsion (vgl. Royce 2016) und satzübergreifende Beziehungen (Martinec & Salway 2005). Eine völlig anders geprägte Forschungslinie bzgl. der Textebene stützt sich auf Varianten der in den 1980er entwickelten formalen Diskursrepräsentationstheorie und dynamischen Semantik; multimodale Erweiterungen hierzu sind Bateman (2007), Abusch (2013) und Wildfeuer (2014, 2017). Die bisher erwähnten Bereiche stellen im Großen und Ganzen keine expliziten Verbindun-

gen zur Kognition her. Es gibt aber besonders um die Ebene des Satzes und der Grammatik herum mehrere kognitiv orientierte Forschungslinien, die sich mit multimodalen Phänomenen befassen. Neuere Ansätze zu konstruktionsbasierten Zugängen zur Sprache sind auch in den Bereich multimodaler Texte erweitert worden (Bergs & Zima 2017), und formale linguistische Ansätze mit kognitiven Grundlagen erforschen mittlerweile eine Reihe weiterer semiotischer Systeme, darunter Musik, Gestik und Tanz (vgl. Schlenker 2017; Patel-Grosz et al. 2019). Ansätze, die auf der kognitiven Linguistik und insbesondere der dort entwickelten Metaphertheorie fußen, sind auch überwiegend kognitiv affin (vgl. Forceville & Urios-Aparis 2009; Kappelhof & Müller 2011), obwohl sie sich seltener mit experimentellen oder anderen Methoden aus der Psychologie oder den Kognitionswissenschaften beschäftigen. Allerdings wurde eine umfassende kognitive Architektur für Multimodalität, die auf einer Reihe von empirischen Studien zu Bildsequenzen, wie sie in Comics zu finden sind, basiert, von Cohn (2016) in Anlehnung an Jackendoff (2007) dargelegt. Darüber hinaus gibt es in der Psychologie angesiedelte Ansätze, die sich früher auf Themen wie das Diskursverstehen konzentriert haben und nun aktiv das multimodale Diskursverstehen verfolgen (vgl. Magliano et al. 1996; Zacks et al. 2009; Loschky et al. 2015). Schließlich haben in den letzten Jahren Ansätze wie die Gesprächsforschung, die Konversationsanalyse und die Gestenstudien die differenzierten Analysen von Interaktion unter multimodaler Perspektive erforscht (vgl. Fricke 2012; Deppermann 2013; Mondada 2016; Deppermann & Streeck 2018; Norris 2019). Übersichten, Einführungen und kritische Diskussionen zu vieler dieser Arbeiten finden sich beispielsweise in Bateman (2014), Stöckl (2016), Klug (2016), Adami (2017) und Wildfeuer et al. (2020).

Diese Blitztour zeigt, dass mittlerweile umfassende Wechselwirkungen zwischen Linguistik (breit gesehen) und Multimodalitätsforschung bestehen. Dennoch ist die Art der Interaktion, die untersucht wird, je nach Gebiet sehr unterschiedlich. Für die stärker kognitiv orientierten Bereiche und solche, die mit eher formalen Ansätzen der Grammatik zu tun haben, werden unterschiedliche empirische Ansätze eingesetzt. Das Untersuchungsmaterial in diesen Fällen erstreckt sich heute über Sprache hinaus und bezieht andere semiotische Systeme mit ein. Es besteht daher in diesen Fällen nach wie vor eine enge Beziehung zwischen der theoretischen Behandlung von Sprache und ihren multimodalen Pendanten. In anderen Bereichen ist aber diese Verknüpfung schwächer geworden; je weiter man sich von den kognitiven und formalen grammatischen Bereichen entfernt, desto unklarer wird das Verhältnis zur Empirie. Obwohl dies eigentlich auch für die Linguistik gilt, ist die Lage im Bereich der Multimodalität um ein Vielfaches verschärft. Klug (2016) gibt einen nützlichen Überblick über verschiedene metho-

dische Orientierungen auf der Text- und Diskursebene und setzt diese in Bezug zur Multimodalität wie folgt. Als Grundlage stützt sie sich auf eine in der deutschen Textlinguistik übliche Unterscheidung zwischen ‚Text‘ als eher produktorientierter, instanziiertem Einheit der Kommunikation und ‚Diskurs‘ als inhärent transtextuelle Strukturierung kommunikativer Phänomene (vgl. Spitzmüller & Warnke 2011). Obwohl ‚reine‘ Fälle eines dieser beiden Extreme selten sind, gibt es dennoch eine annähernde Angleichung textsemantischer Ansätze mit detaillierter Einzeltextanalyse und diskursemantischer Ansätze mit großangelegten, quantitativen Korpusstudien. Die Textsemantik ist also laut Klug eine primär ‚bottom-up‘, punktuelle Mikroanalyse der Details von Einzeltexten, während die Diskurssemantik eine ‚top-down‘ Erfassung von textübergreifenden kommunikativen Strukturen anstrebt.

Jede individuelle Detailanalyse kann es ermöglichen, interessante und aufschlussreiche Interpretationen über einen einzelnen Text und die Formen des kulturellen Wissens zu erzählen, die hinter dem Text ‚stehen‘. Ein tieferes theoretischen Verständnis erfordert aber auch empirisch abgesicherte Verallgemeinerung, um linguistisch von Interesse zu sein. Eine wichtige Voraussetzung für die Erschließung von Generalisierungen ist es deshalb, die Form der einzelnen Analysen auf allgemeinere Kategorien zu beziehen. In der deutschen textlinguistischen Tradition geschieht dies üblicherweise durch die Zuordnung von Kategorien über verschiedene Beschreibungsebenen hinweg, so genannte Mehrebenenmodelle, die heute in der Regel zumindest einige multimodale Komponenten enthalten (Spitzmüller & Warnke 2011; Gardt 2012). Solche Modelle werden als Eckpunkte für detaillierte Analyse angesehen, als „ein helfender, jedoch lockerer Rahmen der Rekonstruktion von Bedeutungszuschreibungen im angestrebten Sinne“ (Klug 2013: 172). Stöckl etwas ausführlicher dazu:

Jeder multimodale Text realisiert eine Textsorte (z.B. Infografik, Zeitungsnachricht, Werbeanzeige), die sich in einem Mehr-Ebenen-Modell beschreiben lässt. Dabei sind textlinguistische Basiskonzepte wie Kohäsion, Kohärenz oder Textstruktur (thematisch, Handlungsstruktur) etc. hilfreich, um multimodale Integration im Gesamttext zu erklären. Von zentraler Bedeutung ist die Frage, wie stark typisiert die einzelnen Modalitäten verwendet werden und wodurch sich die multimodalen Texturen der einzelnen Textsorten unterscheiden. (Stöckl 2016: 5)

Wir werden besonders diese letztere Frage der Typisierung von ‚einzelnen Modalitäten‘ weiter unten in Detail wieder aufgreifen.

In der Systemisch-Funktionalen linguistischen Tradition wird dies im Prinzip ähnlich gehandhabt, indem Beschreibungen über verschiedene Ebenen (die sogenannten ‚Strata‘ des linguistischen Systems) hinweg in Beziehung zueinander gesetzt werden und zwischen ‚Instanzen‘ und ‚Potentialen‘ unterschieden wird

(Halliday 1992). Das Potential ist eine Beschreibung dessen, welche Kategorien und Kombinationen von Kategorien in einem semiotischen System überhaupt verfügbar sind, und Instanzen sind spezifische Produkte, die im Hinblick auf dieses Potential konstruiert wurden. Natürlich hängt das Ausmaß, in dem von detaillierten Einzelanalysen zu größeren korpusbasierten Studien übergegangen werden kann, erheblich davon ab, wie zuverlässig einzelne Kategorien Textinstanzen zugeordnet werden können und wie gut diese individuellen Kategorien sich in allgemeineren, aber dennoch gut validierten transtextuellen Beschreibungsrahmen verankern lassen.

Dieses Verfahren ist in der Linguistik recht gut verstanden und die meisten Kategorien und organisierenden Strukturierungen, die in den diversen Mehrebenenmodellen aufgelistet werden, sind linguistisch gut motiviert. Dies ist für die multimodalen Erweiterungen allerdings nicht der Fall. Das Hinzufügen von Überlegungen zur Multimodalität zeigt einige wichtige methodologische Unterschiede zwischen den disparaten Ansätzen auf und macht entsprechenden Lücken sichtbar. In der deutschen textlinguistischen Tradition gibt es eine lange Geschichte von gebrauchsbazogenen, pragmatischen Kommunikationsorientierungen; Analysen werden daher in der Regel als Mittel zur Aufdeckung der pragma-semantischen Konstruktion von kognitiv verankertem gesellschaftlichem Wissen angesehen. Spezifische Textphänomene auf jeder der Beschreibungsebenen, die in irgendeinem angenommenen Mehrebenenmodell vorgesehen sind, stehen dann als Anhaltspunkte für die Beschreibung von soziokultureller und kognitiver Wissenskonstruktion zur Verfügung. Es ist daher naheliegend, diese mehrebenen-bezogenen Fragenkataloge beizubehalten, aber mit einer Verschiebung über die rein sprachlichen Aspekte hinaus. Aber wie sollten die Merkmale und Kategorien der nicht-verbalen Ebenen aussehen?

Hier finden sich Mischungen von Elementen wie Linien, Punkte, Konturen usw. für Bilder (manchmal als Bildzeichen benannt), oder Oberflächen-Beschreibungen wie räumliche Anordnungen für Seitenkomposition usw., wobei deren genauer Beitrag zur Bedeutung unklar bleibt. Zugleich wird angemerkt, dass es zum Beispiel bei visuellen semiotischen Systemen erhebliche Probleme gibt, überhaupt sinnvolle Bestandteile bzw. analytische Einheiten zu definieren. Bedeutungstragende Elemente scheinen eher vom Gesamttext abzuhängen, und daher droht jede Analyse in einen Teufelskreis zu geraten. Wenn erkennbare und immanent bedeutungstragende Einheiten sich nicht festlegen lassen, ist die traditionelle linguistische Art der Analyse am Ende. Dies ist aber bei eher ikonisch funktionierenden semiotischen Systemen fast immer der Fall. Eine zentrale Forschungsherausforderung besteht also darin, die herkömmlichen linguistischen Methoden für multimodale Phänomene entsprechend zu erweitern, sowohl, wie

Klug deutlich macht, im Sinne differenzierter qualitativer Analysen als auch in Form von breit angelegten quantitativen Korpusstudien:

Dabei ist methodisch die Frage nach der Anwendbarkeit weiterer qualitativen Methoden auf multimodale Textkorpora ebenso zu stellen wie die Möglichkeiten quantitativer, computergestützter semantischer Analysetools auf multimodale Gegenstände auszuloten sind: Wie können typographische Charakteristika von Texten, die ihnen zugehörigen bildlichen und grafischen Anteile auch in digitalen Korpora abgebildet werden? Wie lassen sich beispielsweise Bilder semantisch annotieren? Diese und andere Fragen sollten vermehrt ins Forschungsinteresse text- und diskurssemantischer Analysen rücken. (Klug 2016: 185)

Leider steht die Antwort auf diese Fragen noch aus. Klug (2013) scheint sogar bereit zu sein, den Anspruch auf entsprechende Methoden für visuelle Zeichensysteme aufzugeben:

Wo dem Analytiker zur Beschreibung sprachlicher Diskursbeiträge also qualitative und quantitative Analyseverfahren zur Verfügung stehen [...] ist eine semiotisch erweiterte Diskursanalyse – zumindest mit Blick auf ihre bildlichen Anteile – immer eine qualitative Analyse, die den Einzeltext in den Blick nimmt. (Klug 2013: 183)

Kurz zusammengefasst: Innerhalb der deutschen textlinguistischen Tradition kommt die Beschäftigung mit den verschiedenen Formen multimodaler Phänomene zu kurz. Beschreibungen bleiben weitgehend intuitiv. Begriffe wie Text, Bild usw. werden vorausgesetzt und sind schwach theoretisiert. Daher operieren die meisten multimodalen Analysen immer noch mit alltäglichen Begrifflichkeiten (Stöckl 2016). Obwohl dies nicht bedeutet, dass solche Kategorien und Differenzierungen automatisch falsch sind, ist dies für ein solch zentrales Konzept wie die semiotische Modalität eine auf die Dauer fragwürdige Forschungsstrategie. Es fällt sicher leicht, sagen zu können, dass dieses Buch Text und Bilder hat, oder dass ein Film Geräusche, gesprochene Sprache und bewegte Bilder einsetzt. Dies mag für einige Fragestellungen schon ausreichend sein, wenigstens um die Forschung in Gang zu bringen. Dass das aus der Perspektive der tieferen Grundlage einer Multimodalitätstheorie nicht ausreicht, wird offensichtlich, wenn man beobachtet, wie schnell Versuche dieser Tradition, Zeichenmodalitäten mit einer präziseren Definition zu bestimmen, ins Schleudern kommen (vgl. Stöckl 2004b; Schmitz 2016 u.v.a.m.). Noch häufiger werden die verwendeten Kategorien ohne genauere Definition einfach hingenommen. Mischkategorien werden dann zur Norm, wodurch es zu heterogenen Gruppierungen von Phänomenen kommt, die empirische Forschung jeglicher Art stark erschweren. Mischkategorien können aus zwei Gründe zustande kommen: Entweder weil die Phänomene selbst unklare Grenzen haben oder weil die angewandten Kategorien unangemessen sind. Diskussionen von Multimodalität schlagen häufig das erstere vor, hier wird dage-

gen argumentiert, dass der zweite Grund eine wichtigere Rolle spielt. Auch deswegen wird der Einsatz breiter empirischer Methoden kritisch sein.

Die multimodale Erweiterung des Systemisch-Funktionalen Ansatzes im englischsprachigen Raum hat dagegen genau diese Fragen der multimodalen Phänomene als solche eingehender behandelt, mit detaillierten Vorschlägen für die Binnenstruktur diverser Zeichenmodalitäten. Da die Theorie sozial orientiert ist, ist es für natürlich angesehen worden, neben den sprachlichen Phänomenen auch multimodale Formen zu untersuchen. Sprachgebrauch, und damit multimodale Kommunikation im Allgemeinen, ist als Reflexion, Träger und Gestalter sozialer Konfigurationen zu sehen. Außerdem hat dieser Ansatz die wohl breiteste multimodale Orientierung entwickelt – alle Kommunikationsformen werden unabhängig von ihrer Verwendung oder Nichtverwendung der verbalen Kommunikation als mögliche Untersuchungsgegenstand betrachtet, von der Gestaltung von Lego-Bausteinen über Websites bis hin zu Innenarchitektur, Theater und Filmen u.v.m. Die Breite dieses Forschungsinteresses hat allerdings wahrscheinlich zu der zunehmenden soziokulturellen Ausrichtung dieses Ansatzes wenigstens teilweise beigetragen, da nur wenige der angesprochenen Bereiche Zugang zu etablierten empirisch fundierten Analyseinstrumentarien haben. Eine Auswirkung dieser Entwicklung ist es dann, dass sich die multimodale Forschungstradition, wie sie von Kress, van Leeuwen und anderen unternommen wird, zunehmend von empirischen Anliegen entfernt und viele der Diskurspositionen und Argumentationsweisen der Kulturwissenschaften übernommen hat (z.B. Kress & van Leeuwen 2001; Kress 2010), obwohl auch diese Tradition aus der breit angelegten empirischen Tradition der Linguistik stammt. Diese Richtung ist dann die der Soziosemiotik (vgl. van Leeuwen 2005; Jewitt & Henriksen 2016).

Die in diesem Rahmen erbrachten Analysevorschlge sind entsprechend explorativ, und es ist nicht immer klar, ob sich die angewendeten Analysekategorien verlsslich einsetzen lassen. Dies ist in mehrerer Hinsicht problematisch. Wie Jewitt & Henrikson berichten:

Like many other labour-intensive microanalytical approaches, this approach focuses its analytical gaze on a small collection of texts. [...] The focus on the sign maker, especially when working with texts alone, raises a number of issues including how intentionality can be evaluated and assigned, and the 'right' of interpretation. This approach is inevitably risky – in that everything hinges on how convincing the reader finds the analytical abstraction from the concrete text to the social. (Jewitt & Henrikson 2016: 162)

Jewitt und Henrikson sehen die Rolle der Soziosemiotik folglich darin, Ideen her-
vorzubringen, die durch groere Studien und unterschiedliche Anstze weiter-
entwickelt werden sollten. Der Schritt zu groeren Studien auf der Grundlage von
eher explorativen Analysekategorien ist jedoch alles andere als einfach.

Darüber hinaus gibt es für viele dieser Zielbereiche der multimodalen Analyse bereits starke kulturelle und hermeneutische Traditionen: So ist beispielsweise der Film seit langem ein Gegenstand kultureller, gesellschaftspolitischer und literarischer Interpretation, das Theater wird in den Performance Studies angesprochen, Musik in der Musikwissenschaft, Comics und Graphic Novels werden primär als literarische Gattungen und mit in der Literatur- und Erzähltheorie verankerten Methoden angegangen, und so weiter. Es ist oft recht unklar, wie die mikroanalytischen Beschreibungen einzelner ‚Texte‘, die im sozial-semiotischen Rahmen als Untersuchungsgegenstände genommen werden, mit denen eher hermeneutischen Diskussionen kommunizieren sollen. Wie Ledin & Machin (2017) anmerken, ist die Antwort auf diese Situation allzu oft einfach, sich nicht mit dem bestehenden Werk auseinanderzusetzen. Aber auch signifikante ‚interne‘ theoretische Verschiebungen sind durch diese Tendenz hervorgerufen worden. Was sich gegenwärtig innerhalb der Sozialemiotik abspielt, erinnert häufig an die klassische Zwei-Kulturen-Debatte von Snow (2001 [1959]), bei der auf der Grundlage von Methoden und Forschungsfragen eine unüberwindbare wissenschaftsideologische Trennung in Wissensbereiche vorgenommen wird, mit der Empirie auf der einen Seite und der Hermeneutik (und zunehmend auch der Sozialemiotik) auf der anderen.

Wir halten dies sowohl für Ansätze zum linguistischen Sprachgebrauch als auch für multimodale Kommunikation im Allgemeinen für wenig hilfreich. Generell gilt, dass die möglichen Aussagen umso begrenzter und schwächer werden, je schwächer der Rückbezug auf Methoden ist, die es erlauben, größere Textmengen zu untersuchen. Können wir Muster finden, die wir auf der Grundlage punktueller Analysen erkennen können, die aber mit transtextuellen Eigenschaften korrelieren? Und weiter bis hin zu sozialen Konfigurationen? Und, für die breitere Multimodalität kritisch, sogar bis hin zu anderen Disziplinen? Als eine Strategie sowohl zur Überwindung dieser Fragmentierung als auch zur Wiedergewinnung eines produktiveren Austauschs zwischen Multimodalitätsforschung und linguistischer Theoriebildung legt der gegenwärtige Stand der Technik in der Multimodalitätsforschung nahe, dass eine Neuverpflichtung auf empirische Methoden vorteilhaft wäre (für eine ausführlichere Diskussion vgl. Bateman 2019). Dies benötigt nicht nur groß angelegte korpusbasierte Studien, sondern auch kontrollierte experimentelle Versuche, um die Kategorien für solche größeren Studien vorab zu bestimmen und ihre Relevanz für den Verstehensprozess zu überprüfen bzw. validieren. Dies ist bereits für reine sprachwissenschaftliche Forschung anerkannt, wie z.B. Gries, Hampe & Schönefeld beobachten:

not all kinds of corpus-based evidence are equally well-suited to the purposes to which they are commonly applied. Ideally, therefore, the results of different methodologies should be

brought to bear on our claims and generalisations such that, for example, corpus-based results should be corroborated by experimenting. (Gries, Hampe & Schönefeld 2005: 636)

Es kann zwar für einige Phänomene mehrere mögliche Beschreibungen oder Berechnungsweisen, die von der Korpusforschung unterstützt werden, geben – welche aber davon auch aus der Perspektive der Sprachverarbeitung und -produktion angemessen sind, kann nur durch entsprechende Experimente ermittelt werden. Dies wird noch stärker für die empirische Multimodalitätsforschung gelten.

Ein wesentlicher Bestandteil der nötigen Brücke zwischen qualitativen Interpretationen einerseits und empirischen Arbeiten andererseits sehen wir in der systematischen Anwendung bestimmter sprachwissenschaftlicher Begrifflichkeiten, die bei entsprechender Erweiterung auch für die Strukturierung multimodaler Analysen als vorteilhaft erscheinen. Wir argumentierten, dass vor allem Begriffe aus der Diskurs- und Textanalyse für diese Aufgabe am effektivsten sind und sich als unabdingbar für die empirische Auseinandersetzung mit einem breiten Spektrum multimodaler Kommunikation erweisen. In diesem Beitrag stellen wir einige unserer jüngsten Arbeiten vor, um eine besondere Perspektive darauf zu eröffnen, wie die Multimodalitätsforschung im Allgemeinen zusammen mit empirischen Arbeiten betrieben werden kann. Diese Perspektive betrachtet Multimodalität als ein allgemeines, für alle Arten von Kommunikation und Textualität relevantes Phänomen, das in einer breiten sozialen, qualitativen Orientierung verankert ist, wie sie in Systemisch-Funktionalen und soziosemiotischen Ansätzen angenommen wird, dann aber durch die Anwendung empirischer Methoden, sowohl experimenteller als auch korpusorientierter, enger spezifiziert bzw. ‚trianguliert‘ werden kann und soll.

2 Ein allgemeines Modell der multimodalen Semiotik

Die wohl größte Herausforderung für die Entwicklung von Positionen dieser Art besteht darin, empirische Forschung zu ermöglichen, die einen angemessenen Beitrag zur Theorieentwicklung und zur motivierten Anwendung dieser Theorie leisten kann. Dafür ist eine viel systematischere theoretische Beschreibung der unterschiedlichen Ausdrucksformen, die multimodale Phänomene hervorrufen, erforderlich. Um dieser Anforderung gerecht zu werden, haben wir in den letzten 10 Jahren gemeinsam mit mehreren Kolleg*innen ein detailliertes abstraktes Modell der Multimodalität entwickelt. Dieses Modell ist sowohl semiotisch als auch

diskurstheoretisch motiviert, und ist zudem auch explizit darauf ausgerichtet, empirische Forschung im linguistischen Sinne zu unterstützen und voranzutreiben (vgl. Bateman 2016; Wildfeuer, Bateman & Hiippala 2020). Das Modell ist in dem Sinne abstrakt, dass es ein Raster für alle möglichen multimodalen Kommunikationsformen anbieten soll.

Das Modell baut auf einer grundlegenden Definition des Begriffs der Zeichenmodalität und seiner Einbettung in eine allgemeine Perspektive auf menschliche Kommunikation auf. Die Hauptfunktion unserer Definition ist es, alle Fragen der Existenz und der Differenzierung von Zeichenmodalitäten nicht aus dem alltäglichen Sprachgebrauch, sondern vielmehr als Ergebnis empirischer Forschung zu beantworten. Das heißt konkret, dass wir nicht von ‚Bild‘ (oder sogar von ‚Text‘) reden, als ob wir bereits wüssten, dass diese Termini problemlos Zeichenmodalitäten identifizieren. Um als Zeichenmodalität zu gelten, müssen vorkommende Untersuchungsgegenstände gegenüber unserer Definition geprüft werden. Dies kann nur empirisch erfolgen, indem wir materielle Regelmäßigkeiten bei einem Untersuchungsgegenstand feststellen, die sich dann durch den Einsatz unseres Modells erklären lassen. Bis wir diese Absicherung haben, ist die Annahme einer Zeichenmodalität in einer Beschreibung immer als *Hypothese* zu verstehen. Damit schlagen wir eine Brücke zwischen unseren alltäglichen Ausgangspunkten und immer feineren theoretisch und empirisch motivierten Ausdifferenzierungen. Zugleich wird die häufig anzutreffende Annahme, dass wir bereits ‚wissen‘, welche Zeichenmodalitäten in einem Artefakt oder einer Performance am Werk sind – z.B. Text, Bild, Farbe, Typographie, Musik usw. – werden explizit als zu verfrüht abgelehnt. Genau deswegen ist eine trennschärfere Definition von Zeichenmodalität unabdingbar.

Ein zentraler Teil unserer Definition ist, dass jede Zeichenmodalität notwendigerweise immer in einer *Materialität* verankert ist, die von der Gemeinschaft der Nutzer*innen, die sich mit dieser Zeichenmodalität beschäftigen, verwendet wird. Materialien, die in dieser Art verwendet werden können, sind äußerst vielfältig, erfordern aber mindestens, dass sie ausreichend ‚kontrollierbar‘ sind, um zielgerichtete Artikulationen zuzulassen – sonst wäre es für eine Gemeinschaft von Nutzer*innen nicht möglich, sie als materielle Träger ‚semiotisch aufgeladener‘ Unterscheidungen zu verwenden. Dies umfasst sowohl körperliche Fähigkeiten wie die Verwendung des menschlichen Sprechapparats zur Lautproduktion sowie die Verwendung der Körpermuskulatur zur Produktion von Gestik, Mimik und Körperhaltungen, als auch externalisierte Darstellungen, möglicherweise unterstützt durch Technologie, wie Zeichnung, Druck, Film, computergestützte Interaktion und so weiter. Wir nehmen *Canvas* als allgemeinen Terminus an, der all diese Materialitäten umfasst, wenn man sie speziell aus der Perspektive ihrer

Verwendung für Kommunikation *innerhalb einer spezifischen Zeichenmodalität* betrachtet. Canvases bestimmter semiotischer Modi können mehrere Sinneskanäle einbeziehen bzw. auf diese zurückgreifen und können daher bereits über einzelne Sinnesmodalitäten hinweg operieren, obwohl sie immer noch einzelne semiotische Modi vertreten. Die Materialität bildet somit die erste von drei grundlegenden semiotischen ‚Schichten‘, die definitorisch zu unserem Begriff einer Zeichenmodalität gehören.

In jeder Zeichenmodalität wird dann die Materie zwecks Bedeutungskonstitution geformt. Die Definition dieser Formen und ihrer Aufbaumechanismen (analog dem Begriff einer ‚Grammatik‘) bilden eine zweite, abstraktere semiotische Schicht der expressiven ‚Form‘. Unterschiedliche Disziplinen haben bezüglich ihrer eigenen Untersuchungsgegenstände verschiedene Benennungen für diese Ebene. Sie ist jedoch immer dabei, wenn eine konkrete Ausdrucksform – wie Sprache, Film, bildende Kunst usw. – behandelt wird. Semiotisch gesehen handelt es sich hier um die Hjelmslevsche ‚Form des Ausdrucks‘-Ebene, die wir normalerweise auch mit Hilfe paradigmatischer und syntagmatischer Beschreibungsachsen spezifizieren. Etwas weniger formal sprechen wir in Anlehnung an die Filmtheorie von den ‚technischen Details‘ bzw. ‚Merkmalen‘ einer Zeichenmodalität.

Schließlich kommt eine weitere, noch abstraktere semiotische Schicht dazu – die *Diskurssemantik*. Anders als bei der deutschsprachigen Verwendung von ‚Diskurs‘ handelt es sich hier nicht um eine transtextuelle Beschreibung, sondern um eine dynamische Betrachtung des Bedeutungsaufbaus jedes einzelnen instanziierten Textes. Die Aufgabe der Diskurssemantik innerhalb einer beliebigen Zeichenmodalität besteht darin, bestimmte Bereitstellungen (zweite Schicht) von ‚semiotisch aufgeladenem‘ Material (erste Schicht) mit ihren Nutzungskontexten und den kommunikativen Zwecken, die sie aufnehmen können, in Beziehung zu setzen. Also: Die Diskurssemantik einer Zeichenmodalität stellt die *Interpretationsmechanismen* bereit, die notwendig sind, um die in jeder Zeichenmodalität unterschiedenen besonderen Formen auf ihre Nutzungskontexte zu beziehen und das *beabsichtigte* Deutungsspektrum dieser Formen abzugrenzen. Solche Interpretationen können in Bezug darauf, wie eng sie begrenzt sein sollen, variieren und reichen von sehr spezifischen bis hin zu eher abstrakten ‚Richtlinien‘ für die Interpretationsbildung.

Die drei miteinander verknüpften Schichten einer Zeichenmodalität sind unten in Abbildung 1 dargestellt. Alle drei Komponenten sind zwingend erforderlich, um von einer Zeichenmodalität reden zu können. Dadurch wird die Trennschärfe zwischen Zeichenmodalitäten und semiotisch interpretierbaren Formen im Allgemeinen wesentlich erhöht. Wenn überzeugende Ausformulierungen der

drei Schichten nicht gefunden werden können, spricht dies gegen die Annahme einer Zeichenmodalität. Zeichenmodalitäten können unter sich in Bezug auf den Umfang der Leistung, den sie mit ihrer Diskurssemantik erbringen, unterschiedlich aufgestellt sein, aber die Anwesenheit irgendwelcher Interpretationsrichtungen ist als definitorisch anzusehen. Das heißt: ohne Diskurssemantik gibt es keine Zeichenmodalität.

Welche Zeichenmodalitäten in einem Artefakt oder einer Performance eingesetzt werden, hängt von soziokulturellen Faktoren ab: Bestimmte Kulturen werden es ‚gewohnt sein‘, bestimmte Zeichenmodalitäten statt anderer miteinander zu kombinieren, und sie werden besondere institutionelle, technologische und materielle Kontexte entwickelt haben, innerhalb derer die Kommunikation unter Verwendung dieser Zeichenmodalität stattfindet (vgl. Posner 1991). Wir definieren diese Konfigurationen formell als *Medien*. Unter dieser Sichtweise wird ein Medium am besten als ein historisch stabilisierter Ort für den Einsatz und die Verbreitung einer gewissen Auswahl an Zeichenmodalitäten zur Erreichung eines sozial eingeschränkten und einschränkenden Spektrums an kommunikativen Zwecken gesehen. Zum Beispiel: Bücher sind ein Medium, das traditionell zumindest Regelmäßigkeiten, die wir einer angenommenen Zeichenmodalität des geschriebenen Textes zuschreiben können, und Regelmäßigkeiten, die die Annahme einer Zeichenmodalität des Seitenlayouts plausibel macht, aufweisen. Als Abkürzung können wir dann diese Zeichenmodalitäten hypothetisch für erklärende Beschreibungen des Artefakts aufnehmen: wenn wir einem Buch begegnen, dann nehmen wir an, dass bestimmte Zeichenmodalitäten wahrscheinlich sind, andere weniger wahrscheinlich und andere überhaupt nicht möglich sind. Empirisch lässt sich dies in drei Untersuchungsstufen auseinanderziehen: (i) Hypothesen darüber, welche Zeichenmodalitäten für welche materielle Regelmäßigkeiten zuständig sind, (ii) Hypothesen darüber, welche angenommenen Zeichenmodalitäten durch das Medium kombiniert werden, und (iii) Hypothesen bezüglich der kommunikativen Zwecke, die gerade angestrebt werden. Bisherige Unklarheiten bzw. Schwierigkeiten bei der Einordnung von Zeichenmodalitäten, Textsorten, Medien usw. sind meistens darauf zurückzuführen, dass diese drei Aspekte nicht adäquat voneinander differenziert worden sind. Standardbenennungen von Textsorten, Genres, Medien u.a. bedienen sich ziemlich frei aus Kombinationen dieser Bereiche, was dann leicht für Verwirrung sorgt.

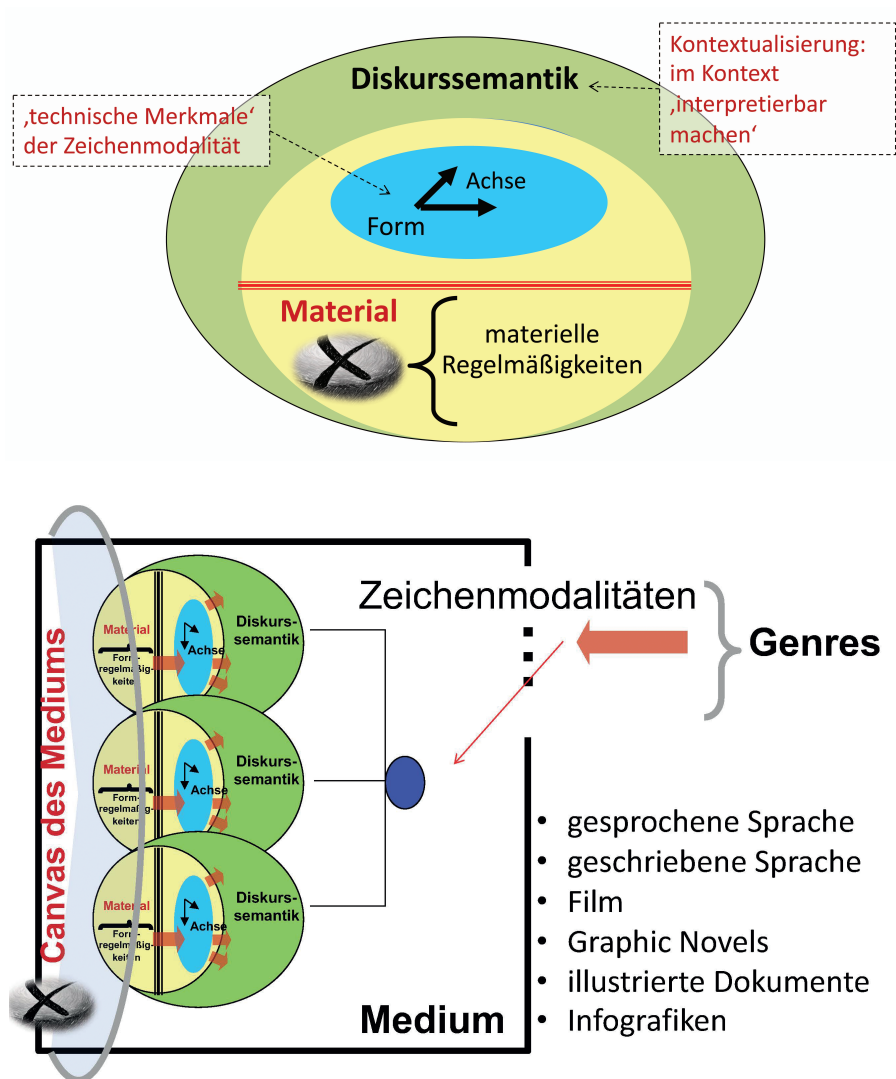


Abbildung 1: Oben: Abstrakte Definition einer Zeichenmodalität. Alle Zeichenmodalitäten kombinieren drei semiotische ‚Schichten‘: materielles Substrat, technische Merkmale (abgekürzt als ‚Form‘) und Diskurssemantik. Unten: Einbettung von Zeichenmodalitäten innerhalb eines Mediums und in Bezug auf Genre (vgl. Wildfeuer et al. 2020: 143, 152)

Dieses Verhältnis zwischen Medien und Zeichenmodalitäten ist in Abbildung 1 grafisch erläutert; in der suggestiven Formulierung von Winkler (2008: 213) betrach-

ten wir folglich jedes Medium als ein *Biotop für Semiose*, weil diverse Zeichenmodalitäten in enge Berührung miteinander gebracht werden. Medien stehen dann als potenzielle Realisierungen oder Ausdrucksstrategien für allgemeine kommunikative Zwecke, die als *Genres* bezeichnet werden. So kann z.B. das Genre einer Erzählung durch die Auswahl eines überwiegend geschriebenen Mediums wie eines Buches oder ebenso als statische visuelle Erzählung im Medium des Comics oder als dynamische visuelle Erzählung im Medium des Films realisiert werden. Die Synchronisation verschiedener Zeichenmodalitäten, die zur Erreichung eines kommunikativen Zwecks beitragen, wird durch Beziehungen zwischen den verschiedenen Diskurssemantiken verwaltet, die von den beitragenden Zeichenmodalitäten bereitgestellt werden. Eine ausführlichere Einführung in diese Ausdifferenzierung der Grundbegriffe der Multimodalität ist Wildfeuer et al. (2020) zu entnehmen.

3 Mehrfache Triangulationen und die Leiter der Abstraktion

Obwohl das im letzten Abschnitt vorgestellte Modell semiotisch gut motiviert ist und zweifellos nützlich ist als allgemeines organisierendes Raster für die Erforschung beliebiger multimodaler Kommunikationssituationen, stellt die Überbrückung der unterschiedlichen Abstraktionsebenen in diesem vielschichtigen Modell doch eine methodische Herausforderung dar. Einerseits sind bei der Untersuchung der multimodalen Kommunikation oft abstrakte soziokulturelle Faktoren involviert, die bei der Erprobung der Verarbeitung von Elementen der unteren wahrnehmungsnäheren Ebenen bestenfalls indirekt zugänglich sind. Beispiele dafür sind etwa konventionalisierte ästhetische Formen und soziopolitische Konstellationen. Eine Reihe empirischer Evidenzen spricht zunehmend dafür, dass solche soziokulturellen Faktoren sogar stärkere Auswirkungen als untergeordnete Formelemente haben können, wenn die ästhetische Interpretation berücksichtigt wird (vgl. Siefkes & Arielli 2018; Van Hedger et al. 2019). Dies stellt eine erhebliche Herausforderung für die Forschung dar, weil bei der Durchführung von Experimenten eine angemessene Vielfalt soziokultureller Variablen und kontrollierte Wahrnehmungsstimuli zu kombinieren sind. Diese können nicht zum Zweck einer einfacheren Untersuchung ausgeklammert werden, ohne die Ergebnisse zu verzerren.

Andererseits muss bei der empirischen Untersuchung darauf geachtet werden, dass Ergebnisse aus einem Bereich nicht verfrüht oder unangemessen in an-

dere übertragen werden. Zum Beispiel kann aus der Tatsache, dass bestimmte empirische Ergebnisse bezüglich des Gesichtsausdrucks in Gemälden erzielt wurden, nicht geschlossen werden, dass diese Ergebnisse für Filme oder Graphic Novels gelten. Umgekehrt gibt es, nur weil eine bestimmte Modellierung der sequentiellen Interpretation für Graphic Novels abgeleitet und empirisch plausibilisiert wurde, keine Garantie dafür, dass diese Modellierung auch für Filme oder andere Formen der sequentiellen Informationsdarstellung gilt. Ebenso lässt sich eine Modellierung der Funktionen von Seitenlayout und Komposition für Zeitungen nicht auf Graphic Novels übertragen, und so weiter. Solche Ergebnisse können als *Hypothesen* für die Organisation anderer Medien verwendet werden, müssen aber dennoch mit Vorsicht behandelt werden, bis die erforderlichen empirischen Untersuchungen durchgeführt worden sind.

Die Notwendigkeit, unterschiedliche Herangehensweisen an die Untersuchung multimodaler Texte zu kombinieren, ist daher offensichtlich. Wir halten es jedoch für unabdingbar, dass solche Trennungen auf einer grundlegenden methodologischen Ebene überwunden werden. Man kann sich vor Augen führen, dass die meisten Probleme der realen Welt sowieso zu komplex sind, um mit einzelnen Perspektiven oder Methoden angemessen angegangen zu werden (vgl. Aguado 2015: 214). Um robuste Methoden zu entwickeln, die die erforderliche Synthese mehrerer Ansätze und Datenquellen in das Forschungsdesign einbeziehen, insbesondere wenn es um die Ästhetik und andere hermeneutische Interpretationsfragen geht, muss daher nach einer grundlegenden Lösung gesucht werden. Ein Ansatz, der verspricht, verschiedene Perspektiven und Dimensionen komplexer Bedeutungs- und ästhetischer Interpretationsprozesse zu erfassen, ist der der *Triangulation*.

Unter „Triangulation“ wird seit mehreren Jahrzehnten eine anerkannte Strategie in den Sozial- und Sprachwissenschaften verstanden (z.B. Denzin 1970; Fine 1985; Mathison 1988; Baker & Egbert 2016), die aber immer noch zu selten angewandt wird, um komplexe Fragen der Ästhetik, des Bedeutungsverständnisses und der Interpretation multimodaler Artefakte und Performances zu untersuchen. Die methodologische Anwendung des Fachterminus „Triangulation“ in der Forschung ist eine als metaphorische Erweiterung ihrer ursprünglichen Anwendung als Messmethode, bei der die Entfernung oder Position unbekannter Punkte in Bezug auf mehrere bereits bekannte Punkte und Entfernungen berechnet werden. Methodologisch ist Triangulation dabei als ein Mittel zur Überwindung von Schwächen gedacht, die jeder einzelnen angewandten Methode innewohnen können: Wenn man mit verschiedenen Methoden konvergierende Beweismittel („converging evidence“) für eine bestimmte theoretische Modellierung eines Gegenstandsbereichs liefern kann, stärkt dies die Beweisführung. Denzin (vgl. 1970:

301) legt eine viel zitierte Typologie von Triangulationsarten vor, darunter Datentriangulation, Forschertriangulation (engl. ‚intercoder reliability‘), Theorietriangulation und Methodentriangulation. Von diesen ist die methodologische Triangulation wahrscheinlich die am häufigsten anzutreffende Form, da sie in den derzeit populären ‚gemischten Methoden‘-Analysestilen durchgeführt wird, die sowohl qualitative als auch quantitative Ansätze umfassen.

Für unsere aktuelle Diskussion ist jedoch die disziplinäre Triangulation die relevanteste, aber auch die anspruchsvollste. Insbesondere in geisteswissenschaftlichen Disziplinen wurde argumentiert, dass Ansichten über die Überwindung von Schwächen durch Triangulation unangemessen sind, da jede Disziplin als ‚vollständige‘ Antwort innerhalb ihres eigenen Diskursuniversums betrachtet werden kann. Dies erfordert, dass die Triangulation über Disziplinen hinweg eher im Sinne der Überwindung möglicher Lücken oder blinder Flecken in den Perspektiven gesehen wird – das heißt die Triangulation liefert eher komplementäre Ansichten als eine Validierung (z.B. Überwindung von Messungenauigkeiten) oder Erweiterungen bereits kompatibler Ansichten (z.B. das Hinzufügen von Winkeln oder Abständen für einen weiteren Punkt, der sich noch im gleichen Gesamttraum befindet) (vgl. Erzberger & Kelle 2003; Hammersley 2008). Dies wurde als ein besonders wichtiges Gegenmittel zu vielleicht einseitigen quantitativen Analysestilen vorgeschlagen (vgl. Janesick 1998: 53). Allerdings ist auch die Triangulation über verschiedene Akte der hermeneutischen Interpretation nicht einfach. In der Tat kann sie leicht als grundlegend fehlgeleiteter Versuch angesehen werden, unvereinbare philosophische Perspektiven auf das Wesen von Wissen und Forschung zu kombinieren (vgl. Mathison 1988; Blaikie 1991). Auch wenn in diesem Fall der Ruf nach multi-, trans- oder pluridisziplinären Herangehensweisen an komplexe soziale Phänomene mitschwingt, muss die Anwendung der Triangulation notwendigerweise auch Bedenken hinsichtlich des Reduktionismus und des mangelnden Bewusstseins aufwerfen, dass es die Disziplinen und ihre Weltsichten sind, die ihre Analyseobjekte überhaupt erst ins Leben rufen. Diese immer wiederkehrende Problematik ist dann stark philosophischer Natur, wie die Skepsis vieler Geisteswissenschaftler*innen zeigt – zum Beispiel als kritische Antwort auf Ansätze in den digitalen Geisteswissenschaften (vgl. Da 2019). Subjektive Bedeutungsinterpretationen und ästhetische Haltungen werden häufig als unvereinbar mit verallgemeinerbaren, objektiven empirischen Belegen interpretiert (vgl. Fish 1980), und infolgedessen lehnen Forscher*innen in eher interpretativen Bereichen nicht selten die Idee ab, dass empirische Studien über größere Datenmengen für ihre eigenen Anliegen und Forschungsfragen relevant sein könnten.

Um diese Diskussion voranzubringen, stützt sich die hier vertretene Sichtweise der disziplinären Triangulation insbesondere auf die Diskussion der ‚ästhetischen Triangulation‘ von Smith (2012), die wiederum auf Flanagan (1992) aufbaut. Smith argumentiert, dass es für das Verständnis komplexer Artefakte, wie z.B. seines primären Beispiels Film, wesentlich ist, ganz unterschiedliche disziplinäre Zugangsweisen zu kombinieren. Smith schlägt vor, dass ein Verständnis von Film (zumindest) neurologische Ansätze zur Funktionsweise des Gehirns bei der Beschäftigung mit Film, phänomenologische Darstellungen der filmischen Erfahrung und kognitiv-psychologische Darstellungen der funktionalen audiovisuellen Verarbeitung erfordert. Darüber hinaus wird dabei jeder Art von ‚Reduktionismus‘ explizit verneint: kein Ansatz ist als ‚grundlegender‘ als irgendein anderer anzusehen. Die Beziehungen zwischen verschiedenen Arten der disziplinären Darstellung werden stattdessen durch die Suche nach ‚Übersetzungen‘ oder Entsprechungen von Vorhersagen über die verschiedenen Bereiche hinweg verfolgt. Erklärungen und offene Fragen innerhalb einer Disziplin werden dann als strukturelle Konfigurationen angesehen, die mit ähnlichen strukturellen Beziehungen in anderen Disziplinen korrespondieren können (oder auch nicht). Eine Triangulation verschiedener Analyseergebnisse auf diese Weise ist dann in der Lage, Interdependenzen zwischen verschiedenen Aspekten des Problems aufzudecken und inhärente Instabilitäten aufzuzeigen, die auftreten würden, wenn einzelne Ansätze auf Kosten anderer privilegiert wären.

Erschwert wird diese Form der Triangulation allerdings dadurch, dass ästhetische oder soziopolitische Interpretationen auf höherer Ebene und Wahrnehmungsdaten auf niedrigerer Ebene kaum direkt miteinander verbunden werden können. Daher fügen wir hier eine weitere vermittelnde Ebene hinzu in Form der Semiotik. Die semiotische Analyse mit Hilfe des oben eingeführten Zeichenmodalitätsmodells stellt eine zwischengeschaltete Ebene von ‚Textstrukturen‘ zur Verfügung, die es erlaubt, Bedingungen zu definieren, unter denen bestimmte Effekte einer Ebene mit Effekten auf anderen Ebenen korrelieren (oder auch nicht). Diese vermittelnde Ebene ist dann zwischen ästhetischen Fragen und Beschreibungen von Materialeigenschaften auf niedriger Ebene angesiedelt. Dies bietet eine Möglichkeit, divergierende Perspektiven in Einklang zu bringen und gleichzeitig produktive Beziehungen zwischen diversen Dimensionen der multimodalen Kommunikation herzustellen. Damit wird eine Tür für die Durchführung von Forschung geöffnet, die auf nachvollziehbare Weise empirische, hermeneutische, interpretative und semiotische Beschreibungsansätze kombiniert.

Diese Triangulationsstrategie wird in Abbildung 2 grafisch verdeutlicht. Wie hier gezeigt, baut die Triangulation auf der Synthese von drei prinzipiellen Analysebereichen auf: (i) hermeneutische Beschreibung in soziokulturellen, ästhe-

tisch orientierten Studien; (ii) korpusbasierte, funktionale semiotische Analyse in Untersuchungen der multimodalen Semiotik; und (iii) experimentelle Ergebnisse in kognitiven Studien. Methoden wie Korpusstudien und experimentelle Methoden sind Techniken, die die Entwicklung eines überprüfbaren Methodenarsenals vorantreiben, das zu intersubjektiv gültigen Ergebnissen führt, und die Kritik, Verfeinerung und Ersetzung bestehender Begriffe und Theorien fördern. Dies funktioniert am besten, wenn es bereits einen Bereich etablierter und allgemein anerkannter Methoden als Ausgangspunkt methodischer Weiterentwicklungen gibt. In vielen Bereichen der Multimodalität gibt es solche Methoden jedoch nicht. Hier wenden wir unser eingeführtes semiotisches Modell direkt an und erweitern die Triangulation, um auch über verschiedene Ebenen der semiotischen Abstraktion zu verfügen. Eine solche Triangulation setzt nicht nur soziokulturelle Studien, Semiotik und Kognitionswissenschaften in Bezug zueinander, sondern auch verschiedene Analyseebenen, die technische Merkmale, Diskurssemantik, narrative Interpretationen, Ästhetik, Affekt und Fragen der Unterschiede verschiedener Genres und Medien umfassen. Damit liefert sie genau jene ‚Fixpunkte‘, die notwendig sind, um mit der Modellierung neuer multimodaler Bereiche zu beginnen und die Analyse von etablierten Bereichen zu verbessern.

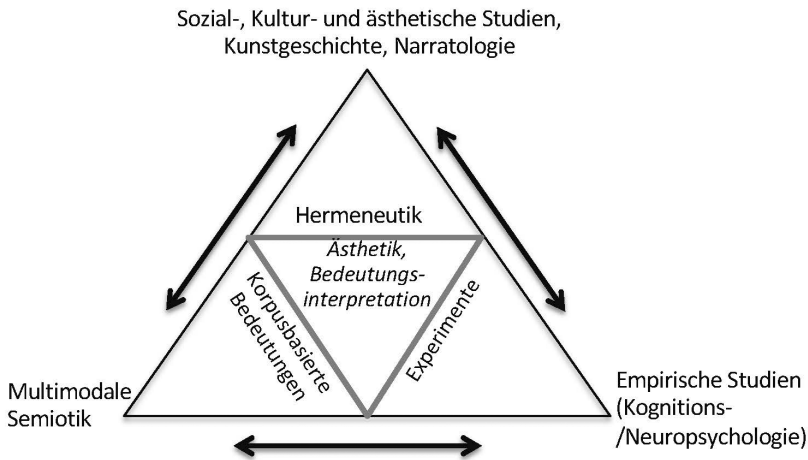


Abbildung 2: Triangulation von Disziplinen zur Behandlung von Problemen der Ästhetik und der Bedeutungsinterpretation

Sowohl in unseren korpusbasierten als auch in unseren experimentellen Untersuchungen und Rezeptionsstudien beziehen wir die verschiedenen Beschreibungsebenen in ähnlicher Weise ein. Die in der multimodalen, linguistisch inspirierten Semiotik verankerten Ansätze werden übernommen, um eine methodische Leiter zur Verbindung von Abstraktionsebenen zu konstruieren, die sonst einfach zu weit voneinander entfernt wären, um sinnvoll in Beziehung gesetzt zu werden. Auf den abstrakteren Ebenen werden hypothetische semiotische Strukturen entworfen, die auf semantische Verbindungen zwischen Elementen auf den unteren Ebenen zurückgreifen, die sowohl für die Wahrnehmung als auch für das Verständnis der Interpretationen relevant sein sollten. Auf den weniger abstrakten Ebenen werden diese Verknüpfungen dazu verwendet, eine Vorauswahl an konkret zu untersuchenden Varianten zu treffen, indem die experimentellen Variablen eingeschränkt werden. Dadurch wird eine Abstraktionsleiter zwischen ästhetischen und anderen hermeneutischen Fragen einerseits und kontrollierten Experimentierstimuli und erwarteten Verteilungen von Phänomenen in größeren Datenkorpora andererseits hergestellt.

4 Empirische Projekte: eine selektive Auswahl

In den letzten Jahren haben wir zusammen mit Kolleg*innen aus unterschiedlichen Disziplinen eine Reihe theoretischer sowie empirischer Studien durchgeführt, die auf die Möglichkeit der Triangulation zwischen Empirie und qualitativer Interpretation fokussieren. Im weiteren Verlauf dieses Beitrags konzentrieren wir uns nun auf diese Untersuchungen. Konkret stellen wir vier Beispiele vor, die sich auf die Triangulation stützen, um eine Synthese aus stilistischer Beschreibung, semiotischer Analyse und empirischen Erkenntnissen zu schaffen. Dabei wird unser erweiterter Ansatz zur Multimodalität eingesetzt, in dem weitgehend *linguistische* Methoden und Begrifflichkeiten aus der Diskursanalyse als Vorbild dienen. Wir charakterisieren die Sichtweise, die wir mit Hilfe der empirischen Analyse im Kontext der Triangulation einnehmen, sowohl theoretisch als auch methodisch, und zeigen, wie die Übernahme sprachwissenschaftlicher Begrifflichkeiten dazu geführt hat, die Funktionsweise unterschiedlicher multimodaler Artefakttypen näher betrachten zu können. Obwohl wir uns mit korpus- und experimentell-basierten Studien gleichermaßen beschäftigen, fokussieren wir im vorliegenden Beitrag lediglich auf experimentelle Studien. Für jedes Beispiel beschreiben wir dementsprechend die experimentellen Methoden und das Design kurz und fassen dann die Hauptergebnisse bezüglich der Triangulation zusam-

men; die vollständigen Details der vorgestellten Studien sind in den referierten Artikeln dargestellt.

4.1 Eyetracking in Comics und Graphic Novels: die Organisation des Diskurses

In unseren ersten zwei Beispielen beschäftigen wir uns mit den Medien Comic und Graphic Novel. Es wird angenommen, dass diese beiden Medien die Mehrzahl ihrer beteiligten Zeichenmodalitäten gemeinsam haben. Die Medien selbst sind äußerst komplex und weisen eine breite Forschungsliteratur auf. Diese Literatur war aber bis vor kurzem von literarischen und mediengeschichtlichen Ansätzen geprägt (vgl. Mikkonen 2017; Aldama 2018; Packard et al. 2019 u.v.a). Dazu hat Cohn (2014) eine kontrovers diskutierte Kritik gebracht, in der er die weitgehende Fehlanzeige an empirischer Comicforschung bemängelte. Diese Situation wird inzwischen explizit angesprochen (vgl. Dunst et al. 2018) und unsere hier berichteten Studien sind auch im Kontext dieses Wandels zu sehen.

Bei genaueren Untersuchungen der semiotischen Ressourcen, die in den Medien Comic und Graphic Novel eingesetzt werden, wird immer deutlicher, dass die bisher übliche Aufmerksamkeit für ‚Bild‘ und ‚Text‘ als die relevanten Zeichenmodalitäten bei weitem nicht ausreicht. Unserer Erfahrung nach führt eine genauere Betrachtung zu nützlicheren Beschreibungen als voreilige Annahmen bezüglich der angewandten Modi. Zur konkreten Veranschaulichung ziehen wir in unseren Beispielen Diskursanalysen heran, die jenseits einer Bild-Sprache-Trennung operieren. Ziel wird es sein, aufzuzeigen, wie wir das volle Spektrum von empirisch durchgeführten Experimenten bis hin zu möglichen Auswirkungen bei den Interpretationen, inklusive Überlegungen zu Narrativ und Gattungen, produktiv in die Diskussion einbringen können.

Als Ausgangspunkt benötigen wir plausible Hypothesen bezüglich der Wirkung der untersuchten Medien durch die angewandten Zeichenmodalitäten. In Frage kommende Modalitätsmöglichkeiten wären aus der Schriftsprache, Seitenlayout, schematisierter piktographischer Darstellungen, Typografie usw. zu erwarten. Auf noch höherer Ebene kommen dann Konventionen bestimmter Comics-Stile und -Gattungen hinzu, die auch Auswirkungen auf experimentelle Ergebnisse haben können. Für narrative Gattungen spielen insbesondere Charaktere, Objekte und Umstände, sowie Sequenzen von Ereignissen diverser Art eine wichtige Rolle. Um darzustellen, wie solche narrativen Einheiten multimodal aufgeführt werden, brauchen wir entsprechende multimodale diskursanalytische Beschreibungen. Hier bietet sich das auf multimodale Kohäsion abzielende semantische Diskursmodell von Tseng (2013) an. Dieses Modell stellt einen dif-

ferenzierten analytischen Rahmen zur Verfügung, um zu untersuchen, wie einzelne Charaktere, Objekte und Umstände in verbalen, visuellen und akustischen Modalitäten durch die Rezipienten identifiziert und nachverfolgt werden können. Um aber sicher zu gehen, dass eine semiotische Beschreibung mit Hilfe der Kohäsionsanalyse tatsächlich für die Zeichenmodalitäten des Mediums relevant ist, müssen wir zeigen können, dass Variationen in der kohäsiven Organisation eines Textes (auch multimodal) mit messbaren Unterschieden in Wahrnehmung und Interpretation korrelieren. Erst dann hätten wir empirischen Belege für eine Ausformulierung unserer Diskurssemantik mit Hilfe von Kohäsionsanalysen.

In Tseng, Laubrock & Pflaeging (2018) berichten wir über eine Versuchsreihe zu Comics und Graphic Novels mit genau diesem Ziel. Als experimentelles Material nahmen wir die ersten Seiten von Paul Karasik und David Mazzuchellis (2004) Graphic-Novel-Adaption von Paul Austers *City of Glass* (1985). Um Variationen der Wahrnehmung festzustellen, verwendeten wir Eyetracking-Techniken und für Variationen in Interpretationen Fragebögen. Die kohäsive Struktur in *City of Glass* zeigt einen markanten Charakter, der eine Vielzahl von Aktionstypen auslöst, die von dynamischem Engagement, Verhalten, Interaktionen mit anderen Charakteren und Gegenständen bis hin zum Ausdruck von Gefühlen reichen (Tseng & Bateman 2018). Um diese Diskursanalyse und die vorgeschlagenen Interpretationen empirisch zu evaluieren, manipulierten wir eine Reihe kohäsiver Ketten in der Graphic Novel, indem wir alle Spuren der visuellen Repräsentation des Charakters durch Leere oder neutrale Gegenstände ersetzten, während wir die zahlreichen sprachbasierten Kohäsionsbindungen auf ihn intakt ließen. Damit sollte getestet werden, welche Auswirkungen das Vorhandensein von visuellen Zeichensignalen auf das narrative Verarbeitungsverhalten und damit auf das Verständnis haben könnte. Wir stellten die Hypothese auf, dass eine solche Manipulation der multimodalen kohäsiven Ketten messbare Auswirkungen haben würde.

Unsere Ergebnisse zeigten in der Tat signifikante Unterschiede im Blickverhalten der Teilnehmer*innen je nachdem, ob sie das Original oder die manipulierte Version gelesen hatten. Damit kann als erwiesen gelten, dass Variationen in der Kohäsion auch mit Variationen im Rezeptionsverhalten zusammenfallen. Insbesondere zeigte sich, dass die Manipulation einen zuverlässigen Effekt auf die Betrachtungszeiten der Panels der Graphic Novel hatte: Die Betrachtungszeiten im Original waren im Durchschnitt 138 ms signifikant länger als im bearbeiteten Zustand. Wir interpretierten dies so, dass die Betrachter*innen erhöhten kognitiven Aufwand betreiben, wenn sie *narrativ-relevante* visuelle Information haben. Dies unterstützt weiter die Annahme, dass die Kohäsion eine funktionelle Rolle beim Verstehen des multimodalen Textes aufnimmt. Die Daten des Frage-

bogens führten zu ähnlichen Schlussfolgerungen. Dies legt nahe, dass die Frage, ob eine *visuell-verbale, cross-modale Kohäsionskette* konstruiert wird oder nicht, Auswirkungen auf das Leseverhalten der Teilnehmer*innen hat. Die Teilnehmer*innen schienen die Hinweise auf eine visuell dargestellte Figur in den Panels zu verfolgen, wenn diese Figur innerhalb der gleichen kohäsiven Kette auffällig und kohärent mit dem Text verbunden war. Dies unterstützt dann weiter die Untersuchung von Kohäsion mit dem Ziel, Vorhersagen darüber zu treffen und zu testen, welche technischen Merkmale in einem Artefakt oder einer Ausführung die Interpretation beeinflussen und wann. Für produktive experimentelle Stimuli, die zwecks Variabilität in der Rezeption und Interpretation eingesetzt werden sollten, scheint ihre kohäsive Zuordnung sicherlich wichtig zu sein.

4.2 Eyetracking in Comics und Graphic Novels: Seitenaufbau/Komposition

Im zweiten Beispiel erweitern wir die untersuchten Phänomene, um möglichen narrativen Auswirkungen von der *Seitengestaltung* (oder *Layout*) in Comics und Graphic Novels näher zu kommen. Seitengestaltung wird im Allgemeinen in der Comicforschung als eine höchst variable und kreativ eingesetzte Gestaltungsdimension angesehen, die sich kaum empirisch untersuchen lässt. Seitengestalterische Unterscheidungen werden trotzdem in der Forschungsliteratur häufig mit Auswirkungen für die Narrative assoziiert. Wenn aber diese Auswirkungen systematisch sind, sollte es möglich sein, Layout als operative Zeichenmodalität in diesen Medien zu identifizieren und spezifizieren. Als Vorarbeit haben wir in Bateman et al. (2017) eine detaillierte Klassifikation möglicher Layout-Strategien in Comics entworfen, die selbst auf der Grundlage einer Korpusanalyse abgeleitet wurde. Korpusanalyse ist hier in dem Sinne gemeint, dass eine etwas größere Auswahl von Comicseiten (d.h., mehrere hundert Seiten) von mehreren Forscher*innen auf mögliche Gestaltungsprinzipien untersucht wurden; diese Prinzipien wurden wiederholt verglichen und verfeinert, bis ein weitgehend stabiler Satz übrig zu bleiben schien. Die Klassifikation diente dann als mögliche paradigmatische Achse der Form-Ebene einer entsprechenden Zeichenmodalität. Wenn das so ist, muss es möglich sein, empirisch zu belegen, dass die Auswahl kontrastierender Strategien aus dieser Klassifikation erstens zu kontrastierenden Interpretationen auf der höheren narrativen Ebene und zweitens zu systematisch unterschiedlichem Verhalten bei den Rezipient*innen führen soll. Um diese Fragen experimentell nachzugehen, haben wir wiederum Eyetracking-Techniken angewandt.

In Bateman et al. (2018) beschreiben wir ein Experiment dieser Art, in dem ein ziemlich subtiler Kontrast aus unserer Klassifikation von Layout-Strategien untersucht wird. Dieser Kontrast entsteht dadurch, dass der Rahmen eines Panels (oder mehrerer) auf einer Seite entfernt wird und die Farbgebung seines Hintergrunds als Hintergrund für die ganze Seite ausgeweitet wird, vor dem die nicht veränderten Panels mit ihren Rahmen stehen. Da die Seite dann den Anschein hat, dass das sich ausbreitende Panel in gewisser Weise in einer sonst durch die übrigen Panels auf der Seite gebildeten gitterartigen Struktur fehlt, entsteht visuell eine Lücke oder ‚Gap‘. Die Diskussion solcher Seitenlayouts in der eher hermeneutischen Literatur führt häufig zu der Vermutung, dass die Einführung von Lücken und abgeschnittenen Panels *erzählerische* Konsequenzen haben kann, da die Panels gruppiert und Unterscheidungen zwischen Vorder- und Hintergrund eingeführt werden. Eine systematische empirische Überprüfung solcher Effekte wurde jedoch bisher nicht durchgeführt, vermutlich weil es ohne stärkeres semiotisches Gerüst methodologisch schwierig ist, verlässlich festzustellen, ob solche Layouts die Rezeption beeinflussen oder nicht.

Um die Frage, ob dieser Kontrast zwischen dem ursprünglichen Layout und dem manipulierten Layout mit der Lücke als Beitrag zu einer Zeichenmodalität des Layouts in Comics dienen könnte, untersuchten wir die Auswirkungen des Kontrasts experimentell wie folgt. 12 Originalseiten aus Comics, die eindeutig die gewünschte Layout-Strategie aufwiesen, wurden aus unserem Korpus ausgewählt. Alle Seiten waren dem Superhelden-Genre zuzuordnen und wurden seit dem Jahr 2000 veröffentlicht. Anschließend wurden 12 modifizierte Versionen dieser Originalseiten erstellt, bei denen der Hintergrund entfernt und ein Rahmen entsprechend der Seite um das zuvor ungerahmte Panel herum wieder eingefügt wurde. Von jedem Paar Original-/modifizierter Seiten war also nur ein Panel betroffen; alle anderen Panels blieben unverändert. Durch diese Manipulation haben wir minimal kontrastierende Paare für das interessierende Layout-Merkmal produziert. Wenn dieser Kontrast für die Zeichenmodalität relevant ist, dann sollten die Rezipient*innen entsprechende Unterschiede in ihrem Verhalten und ihrer Interpretation aufweisen.

Die Studie bestand dann aus zwei Teilen. Da wir insbesondere feststellen wollten, inwieweit die untersuchten Designvarianten mit abstrakteren Fragen der Erzählrezeption korreliert sein könnten, versuchten wir zunächst, die naiven Reaktionen der Leser*innen auf unsere Manipulationen unabhängig vom Blickverfolgungsverhalten näher zu betrachten. Zu diesem Zweck baten wir Teilnehmer*innen einer separaten Pilotstudie, die Paare von Original- und modifizierten Seiten zu betrachten und diese Paare nach potentiell erzählrelevanten Dimensionen (Sequentialität, Kausalität usw.) zu bewerten. Dann wurden im Eyetrack-

ing-Teil der Studie 60 Teilnehmer*innen gemischte Sammlungen von Original- und modifizierten Seiten gezeigt, wobei sie entweder die ursprüngliche oder die modifizierte Seite sahen, nie jedoch zwei Seiten aus demselben Kontrastpaar. Zur Auswertung der Blickverhaltensdaten wurde untersucht, ob die Modifikation als zuverlässiger Prädiktor für Variationen in der Betrachtungsdauer einzelner Panels verwendet werden könnte.

Die Ergebnisse waren zuerst gemischt. Einige manipulierten Seiten verursachten längere Betrachtungszeiten als die nicht manipulierten Versionen, bei anderen war es umgekehrt. Es zeigte sich jedoch, dass es eine Interaktion zwischen den *narrativen Konsequenzen* und der Blickverhaltensvariation gab. Einerseits hatten die fünf Seitenpaare, die in der Pilotstudie als die mit den am wenigsten erkennbaren narrativen Unterschieden beurteilt wurden, auch keine signifikanten Auswirkungen in den Vorhersagen der Panelbesuchsrates. Und andererseits gehörten sechs Seiten, die doch statistische Signifikanz für ihre Interaktion mit dem Modifikationsstatus (d.h. mit ‚Lücke‘ vs. ohne ‚Lücke‘) für die Panelbesuche zeigten, alle zu den Seiten, bei denen die Manipulation in der Pilotstudie am stärksten als erzählrelevant eingeschätzt wurde. Dies bot dann einen weiteren Triangulationspunkt in Bezug auf die Forschungsfrage. Eine statistisch signifikante Interaktion zwischen dem Modifikationsstatus und den Seiten schien gerade in den Fällen wahrscheinlicher zu sein, in denen die Manipulation von den Rezipient*innen als stärker narrativ beurteilt wurde. Ohne Berücksichtigung der interpretativen Ebene der erzählerischen Wirkung wäre es also nicht möglich gewesen, eine Erklärung für die bei einigen der Stimuli beobachtete Variation der Reaktionen zu geben. Dies lässt sich als Hinweis darauf interpretieren, dass wir möglicherweise eine sinnvolle Triangulation über mehrere Abstraktionsebenen gefunden hatten.

4.3 Kohäsion und Blickverfolgung im Film

Für die letzten beiden Beispiele wenden wir uns dem Medium Film zu, um die Methodologie weiter zu beleuchten. Beim ersten Beispiel kombinierten wir Eye-tracking-Methoden und multimodale narrative Kohäsionsanalyse, ähnlich wie bei der oben geschilderten Studie zur Kohäsion in Comics. Hier wollten wir Belege dafür sammeln, dass die kohäsive Analyse eine vergleichbare Rolle für die Diskursinterpretation beim Film spielt wie bei anderen Medien. In Zusammenarbeit mit Kolleg*innen der Universität Potsdam (Tseng et al. 2021) untersuchten wir empirisch, wie sich visuelle und auditive Kohäsionsmarker im Film auf die Aufmerksamkeit und das Verständnis der Geschichte auswirken können. Die Integration eines Blickverfolgungsansatzes und einer Kohäsionsanalyse ermög-

lichte es uns auf diese Weise, zwei Hauptforschungsrichtungen in der filmischen Erzählinterpretation auszubalancieren: (i) die Frage, wie die Lenkung der Aufmerksamkeit eines Filmbetrachters das Verständnis von Rezipienten beeinflusst, und (ii) die rezeptionsbezogene Frage, wie narrative Elemente im Film die Interpretation der Bedeutung lenken.

Das Material für dieses Experiment bestand aus zwei Versionen der Eröffnungsszene von Alfred Hitchcocks berühmtem Film *The Birds* (1963), in der die Protagonistin Melanie eine Straße in San Francisco entlang schlendert und dann eine Tierhandlung betritt. Eine detaillierte Kohäsionsanalyse der Szene zeigte, wie dieselben Schauplätze, Charaktere und Objekte durch verbale, akustische und visuelle Modalitäten miteinander verbunden werden und damit eine kohärente Interpretation ermöglichen. Anschließend wurde eine manipulierte Version der Szene erstellt, indem bestimmte kohäsive Hinweise entfernt wurden, um die Identität des Schauplatzes der Tierhandlung zu verschleiern. Dabei wurden auf subtile Weise verbale und akustische Hinweise auf die Umgebung entfernt, nämlich ein schriftliches Zeichen, das die Tierhandlung identifiziert, und Vogelgezwitscher innerhalb des Geschäfts. Dies bedeutete, dass die Hinweise auf die Tierhandlung in den ersten Einstellungen der Eröffnungsszene außerhalb des Geschäfts aus der Aufnahme entfernt wurden, während der darauf folgende Teil der Eröffnungsszene innerhalb des Geschäfts visuell identisch war. Die Modifikation der Szenen fiel beim normalen Zuschauen nicht störend auf. Da die kohäsive Modellierung jedoch nahelegt, dass eine Störung der kohäsiven Verbindungen auf diese Weise Konsequenzen für die Interpretation haben sollte, wurden dennoch ein unterschiedliches Blickverhalten sowie Unsicherheiten bei der Interpretation vorhergesagt, falls die Kohäsion für die mitwirkenden Zeichenmodalitäten eine Rolle spielte.

Die beiden Versionen wurden zwei Gruppen von Zuschauer*innen präsentiert, und es wurden sowohl Eyetracking- als auch Fragebogendaten erhoben, um ihr Verständnis der gesehenen Szene zu messen. Wir testeten unter anderem die folgenden zwei Hypothesen: (i) Die Zuschauer sind sich über den Ort des Geschehens innerhalb des Ladens weniger sicher und zeigen mehr Orientierungsverhalten; (ii) dadurch wird die Aufmerksamkeit von Melanie abgelenkt. Die Ergebnisse des Fragebogens zeigten, dass die Zuschauer*innen der manipulierten Version sich in der Tat ziemlich unsicher über die Art des Geschäfts waren, obwohl entsprechende visuelle Elemente immer noch leicht zugänglich im Blickfeld waren. Im Gegensatz dazu waren sich die Zuschauer des Originals sehr wohl bewusst, dass sich die Szene in einem Tiergeschäft abspielt. Darüber hinaus legten die Daten aus dem Fragebogen nahe, dass die Geräusche (Vogelgezwitschern) im Geschäft ebenfalls signifikante Hinweise zur Identifizierung der Kulisse waren; die Erset-

zung dieser Geräusche durch allgemeine leise Hintergrundmusik führte dazu, dass die meisten Zuschauer*innen der manipulierten Version glaubten, dass die Szene in einem Kaufhaus stattfand, obwohl die Vögel in Käfigen noch immer explizit auf dem Bildschirm zu sehen waren.

Die Augenbewegungsmuster lieferten weitere Unterstützung für beide Hypothesen. Die Variabilität der Fixierungspunkte im ersten Teil der Szene im Inneren des Ladens war in der manipulierten Version höher, was darauf hindeutet, dass die Betrachter*innen das Innere des Ladens in größerem Maße erkunden, vermutlich um seine Identität festzustellen und die entsprechenden narrativen Kohäsionslücken zu schließen. Im Gegensatz dazu konzentrierten sich die Betrachter*innen des Originals fast ausschließlich auf Melanie, die im Laden eine Treppe hinaufgeht und weitere Aktionen durchführt.

Zusammenfassend lässt sich sagen, dass die Verwendung von Kohäsionsanalysen zur Ableitung empirisch überprüfbarer Hypothesen für die hier beschriebenen Blickverfolgungsexperimente in der Tat neue Erkenntnisse darüber liefert, wie die Aufmerksamkeit der Betrachter*innen und ihre übergeordnete Interpretation von Charakter- und Raumverhältnissen in empirischen Studien sowohl theoretisch als auch methodisch in Beziehung gesetzt werden können. Ausgehend von unserem analytischen Ergebnis auf der formalen Ebene zeigte diese Studie auch, wie subtile digitale Manipulationen einer Filmszene in Kombination mit Eyetracking-Daten genutzt werden können, um die Aufmerksamkeitslenkung zu untersuchen, die von Filmemacher*innen üblicherweise für die Erzählkontinuität eingesetzt wird.

4.4 Kohäsion und Neurocinematics

Mit dem letzten Beispiel wurde eine Forschungsstrategie vorgestellt, mit der die Relation zwischen technischen Details und Erzählverständnis untersucht werden kann. Ausgehend vom gleichen Prinzip der filmischen Kohäsionsanalyse setzten wir mit Kolleg*innen an der Aalto-Universität in Finnland neurowissenschaftliche Methoden ein, um zu untersuchen, ob Variationen in Kohäsion auch mit Variationen der Gehirnaktivität systematisch korrelieren. Als Daten für die Studie dienten fMRI-Gehirnscans, die während der Betrachtung von Christopher Nolans Film *Memento* (2000) vorgenommen worden waren (Tseng et al. 2017). Dieser Film ist bekanntlich eine Herausforderung für Zuschauer*innen, da er eine umgekehrte zeitliche narrative Organisation aufweist.

Die schiere Menge an Daten, die in solchen Hirnbildgebungsstudien gesammelt werden, ist immens, und es ist oft keineswegs klar, wo in diesen Datenmengen interessante Verallgemeinerungen zu suchen sind. Dafür haben wir die Ana-

lyse der Kohäsion verwendet. Auf der Basis der Kohäsionsanalyse lassen sich kontrastierende Diskurssituationen definieren, normalerweise bezüglich der Fortsetzung bzw. Unterbrechung von interagierenden kohäsiven Ketten. Für die aktuelle Studie wurden zwei Diskurssituationen, bei denen kontrastierende Mechanismen der Erzählkonstruktion mobilisiert wurden, auf Basis der Kohäsionsmuster definiert. Ihre Vorkommen im Film wurden dann durch die Zeitintervalle, während denen sie auftraten, dokumentiert. Dafür wurde der ganze zeitliche Ablauf der Gehirndaten mit ‚Ereignisfenstern‘ unterlegt, während zugleich spezifizierte Kohäsionsmuster für die betrachtete Filmsequenz definiert wurden. Auf dieser Basis konnten gezielte Mengen von Gehirndaten daraufhin untersucht werden, ob systematisch wiederkehrende Gehirnaktivitätsmuster gefunden werden konnten.

Zwei zentrale Hypothesen wurden von der Studie ins Visier genommen. Erstens untersuchten wir, ob differenzierte kohäsionsbasierte Ereignisbeschreibungen des Films für messbare neurologische Reaktionen prädiktiv waren. Wichtig war dabei, dass die Ausformulierung der Beschreibungen auf der Diskursebene als eine Methode untersucht wurde, um über die Ergebnisse auf weniger abstrakten Wahrnehmungsebenen, wie z.B. Reaktionen auf audiovisuelle Stimuli, hinauszugehen. Es wäre keine Überraschung, wenn beim Betrachten von visuell identischen Szenen Regelmäßigkeiten im Hirnverhalten der Proband*innen in eher front-end (d.h. stimulus-nahen) visuellen Verarbeitungsbereichen des Gehirns gefunden würden. Wenn aber auf der Grundlage der *Diskursstrukturen* Regelmäßigkeiten in Gehirnbereichen gefunden werden könnten, die für tiefere Verarbeitung bekannt sind, wäre das ein starkes Indiz dafür, dass eine weitere Quelle für produktive Triangulationen festgestellt worden ist. Das heißt, anstatt zu sehen, wie Schnitttechniken oder visuelle Stimuli mit verschiedenen Gehirnreaktionen zusammenhängen, weisen die Analyse und die Ergebnisse darauf hin, dass es möglich ist, das Abstraktionsniveau auch in solchen Untersuchungen durch die Herstellung einer Verbindung mit der Diskurssemantik zu erhöhen.

Dieses zentrale Ziel, Korrelationen zwischen Diskursstrukturen, die als kohäsiv zusammenhängend abgeleitet wurden, und tatsächlichen Rezeptionsprozessen im Gehirn zu finden, wurde erreicht. Die unterschiedlichen Sets von Diskursrealisierungen schienen in der Tat mit signifikanten Unterschieden in den Hirndaten zu korrelieren. Die Verlässlichkeit solcher Ergebnisse bietet dann einen Studienzweig parallel zu den traditionelleren Anliegen, die neurale Aktivität mit der Ästhetik in Beziehung zu setzen (vgl. Changeux 1994; Aldama 2015). Statt ästhetischer Antworten im weiteren Sinne fokussiert sich die Aufmerksamkeit hier eher auf differenzierte Strategien, die von den Filmmacher*innen im Dienste

bestimmter, gewollter Rezeptionseffekte eingesetzt zu werden scheinen. Diese Strategien ließen sich einerseits als Kohäsionsmuster darstellen und andererseits als wiederkehrende Muster von Gehirnaktivität erkennen. Eine logische Fortsetzung dieser Arbeit ist dann die Suche nach Korrelationen zwischen den Datenmustern des Gehirns und weitere abstrakteren, narrativ-relevanten Charakterisierungen der Filmrezeption. Eine entsprechende Studie wurde bereits für verbale Texte vorgelegt (vgl. Ferstl & von Cramon 2001), aber bisher hat das Fehlen von einem entsprechend differenzierten Zeichenmodalitätsmodell Untersuchungen außerhalb des Sprachsystems verhindert. Das Konzept der Diskurssemantik als einer vermittelnden Ebene, mit der diverse Sinnesmodalitäten in Beziehung zueinander gesetzt werden können, bietet eine Grundlage dafür, jetzt über diesen Zustand hinauszukommen.

5 Zusammenfassung und zukünftige Ausrichtungen

Die Idee, über verschiedene Disziplinen hinweg zu triangulieren, um komplementäre Zugangsmöglichkeiten zu einem komplexen Analyseobjekt zu erhellen, wird immer praktikabler und ermöglicht zunehmend raffiniertere (auch experimentelle) Untersuchungsdesigns. In diesem Beitrag haben wir über mehrere Verfeinerungen dieser allgemeinen Idee berichtet, in denen weitere ‚Punkte‘ der Triangulation hinzugefügt wurden, die direkt aus unserem allgemeinen semiotischen Modell der Multimodalität abgeleitet sind. Auf diese Weise werden differenzierte und funktional motivierte semiotische Artikulationen eines zu analysierenden Gegenstands sowohl mit abstrakteren Interpretationsschemata als auch mit direkt materiell feststellbaren Erscheinungsformen in Beziehung gesetzt. Dieser Auffassung zufolge dienen die verschiedenen Arten von Diskursanalysen als vermittelnde Beschreibungen auf einer ‚Zwischenebene‘, die einerseits in Bezug auf spezifische Kombinationen technischer Merkmale in einem Medium operationalisierbar sind und andererseits dazu dienen, abstraktere Deutungsverläufe auf Seiten der Rezipient*innen vorherzusagen sowie zu erklären. Diese letzteren sollten dann wiederum mit messbaren Rezeptionsunterschieden korreliert werden können. Die Relation von Diskursbeschreibungen zu anderen – mehr oder weniger abstrakten Ebenen – bietet daher einen Rahmen, innerhalb dessen empirische Vorhersagen über die flexible Aufnahme von Zeichen durch eine*n Leser*in, Zuhörer*in, Benutzer*in usw. während der Rezeption konzeptualisiert werden können (Bateman & Wildfeuer 2014).

Solche Vorhersagen können jedoch nur so genau sein wie die Darstellung der angenommenen Diskursorganisation. Um diese Praxis aufzuzeigen, haben wir in diesem Beitrag über konkrete Beispiele aus unseren aktuellen Studien zum Film und zur visuellen Erzählung in Comics und Graphic Novels berichtet. Es wurde gezeigt, wie eine breite semiotische Konzeption der Multimodalität explizit auf die empirische Untersuchung ausgerichtet werden kann. In allen Fällen bleiben jedoch unsere Ergebnisse in diesem Stadium weitgehend explorativ. Der Hauptzweck dieses Beitrags besteht daher darin, Richtungen aufzuzeigen, in denen semiotische Analysen in Zukunft weitergeführt werden können, damit die immer wieder auftretende Schwäche der traditionellen semiotischen Forschung in Bezug auf ihre mangelhaften empirischen Grundlagen durch reflektierten Einsatz von linguistisch inspirierten Methoden behoben werden kann.

Die stratifizierte Auffassung von Bedeutung, die als Folge unseres semiotischen Modells angenommen wurde, deutet damit an, wie wir eine tiefe Kluft zwischen verschiedenen Ansätzen der Bedeutungsanalyse, die sich manchmal sogar feindselig gegenüberstehen, wirksam überbrücken können. Insbesondere erachten wir die anhaltende Trennung zwischen eher hermeneutischen und diskursiven Interpretationen einerseits und formal bzw. quantitativ orientierten Ansätzen der Analyse und Ästhetik andererseits als besonders kontraproduktiv. Die mangelnde Interaktion zwischen diesen Forschungsrichtungen macht es für die hermeneutische Interpretation schwierig, eine solide empirische Unterstützung zu erreichen, und verhindert gleichzeitig, dass empirische Analysen auf hermeneutisch und/oder soziokulturell relevante Fragen angewendet werden. Dazu benötigten wir Methoden zur Konstruktion von Hypothesen, die eigens dafür formuliert wurden, um zu untersuchen, wie Erkenntnisse zur Kognition und die abstrakten Ebenen der Stilistik, des Genres und der breiteren sozialen und kulturellen Kontexte formal zusammenhängen.

Die Grundlage unseres Ansatzes war daher die Annahme, dass ein stratifiziertes, mehrstufiges Verfahren für die semiotische Analyse ein sensibleres Instrument zur Herausarbeitung signifikanter formaler Unterschiede darstellen kann, die auch als konstitutive Unterschiede auf höheren, soziokulturellen Abstraktionsebenen betrachtet werden können. Methodisch funktioniert dieser Ansatz durch die Identifizierung ‚vertikaler‘ Schnitte durch die oben in Abbildung 1 (obere Grafik) dargestellten Schichten der medialen Konfigurationen und der Zeichenmodalitäten, um Beschreibungen auf verschiedenen Abstraktionsebenen zu liefern, die dennoch komplementär in Beziehung zueinander gesetzt werden können. Dadurch wird ein allgemeines Modell für die Untersuchung von semiotisch anspruchsvollen Gegenständen jeder Art zur Verfügung gestellt.

Eine der wichtigsten Erkenntnisse aus der hier vorgestellten Forschungsrichtung und den verfolgten empirischen Untersuchungen ist die erweiterte Anwendbarkeit analytischer Methoden aus der Sprachwissenschaft, so dass sie für jede multimodale Kommunikation relevant werden. Die Diskussion über solche Erweiterungen wird schon seit langem in Bezug auf die gesprochensprachliche Interaktion geführt, aber selten für darüber hinausgehende Fragestellungen – in solchen Diskussionen wird sogar ‚Multimodalität‘ häufig als Synonym für gesprochene Face-to-face-Interaktion gesehen. Erweiterungen in diesem Bereich befassen sich folglich lediglich mit Kombinationen von gesprochener Sprache, Gestik, Intonation oder Proxemik. Perniss (2018) bietet einen nützlichen Überblick über solche Erweiterungen, darunter auch Argumente aus neueren Ergebnissen der Neurowissenschaften, die Ansätze, die von Unterschieden zwischen ‚linguistischem‘ und ‚nicht-linguistischem‘ Kommunikationsverhalten ausgehen, grundsätzlich in Frage stellen. Diese Forschungslinie weist jedoch insofern eine anhaltende Schwäche auf, als sich die vorgeschlagenen Ansätze überwiegend auf Überlegungen zu Erweiterungen des Grammatik-Begriffs beschränken. Wie wir im vorliegenden Beitrag argumentiert haben, ist Multimodalität jedoch ein weit aus umfassenderer Phänomenbereich. Erstens erstreckt sich Multimodalität auf alle Formen der Kommunikation, einschließlich der visuellen, akustischen, bewegungsbasierten und haptischen und ist somit keineswegs auf die gesprochene Interaktion beschränkt. Zweitens kommt, wie wir bereits mehrfach betont haben, ein großer Teil der Kraft der multimodalen Kommunikation aus der Vielfalt und Flexibilität von Diskursprozessen, die jenseits der Grammatik ablaufen.

Auch in Bezug auf visuelle Kommunikation gibt es Vorschläge, die für eine engere Verbindung zwischen bildwissenschaftlichen Studien und sprachwissenschaftlichen Ansätzen plädieren (vgl. Spillner 1982; Schmitz 2003; van Leeuwen 2004; Fix 2008; Klug 2013). Laut van Leeuwen:

Speech acts should be renamed communicative acts and understood as multimodal micro-events in which all signs present combine to determine its communicative intent. (van Leeuwen 2004: 8)

Genres of speech and writing are in fact multimodal: speech genres combine language and action in an integrated whole, written genres combine language, image, and graphics in an integrated whole. Speech genres should therefore be renamed ‘performed’ genres and written genres ‘inscribed’ genres. Various combinations of performance and inscription are of course possible. (van Leeuwen 2004: 10)

Wir haben diese Auffassung von Kommunikation und ihrer Erforschung zum Ausgangspunkt genommen und darüber hinaus gezeigt, wie diese radikale Erweiterung dessen, was als Gegenstand einer linguistischen Untersuchung zu be-

trachten ist, dennoch präzise erfasst und empirisch untersucht werden kann. Für viele Forschungsfragen, die Themen der Grammatik, der Semantik, des Diskurses, des Genres und des Stils umfassen, betrachten wir die Ergebnisse, die wir für die multimodale Kommunikation erhalten, als von unmittelbarer Bedeutung für die adäquate Theoretisierung dieser Begriffe auch innerhalb der Sprachwissenschaft. Und damit sind wir wieder beim Anfang und dem dort angeführten Zitat von Saussure. In allen Fällen gilt jedoch: Je präziser der Zusammenhang zwischen empirischer Analyse und geeigneten sprachwissenschaftlichen Begrifflichkeiten hergestellt werden kann, desto aussagekräftiger werden die Ergebnisse sowohl für die sprachwissenschaftliche Diskussion und Theoriebildung als auch für multimodale Anwendungskontexte im weiteren Sinne sein.

Literatur

- Abusch, Dorit (2013): Applying Discourse Semantics and Pragmatics to Co-reference in Picture Sequences. In: Chemla, Emanuel/Homer, Vincent/Winterstein, Grégoire (Hg.): *Proceedings of Sinn und Bedeutung* 17. Paris, 9–25. URL: <http://hdl.handle.net/1813/30598>.
- Adami, Elisabetta (2017): Multimodality. In: García, Ofelia/Flores, Nelson/Spotti, Massimiliano (Hg.): *The Oxford Handbook of Language and Society*. Oxford, New York: Oxford University Press, 451–472.
- Aguado, Karin (2015): Triangulation: Möglichkeiten, Grenzen, Desiderate. In: Elsner, Daniela/Viebrock, Britta (Hg.): *Triangulation in der Fremdsprachenforschung*. Frankfurt a.M.: Lang, 203–220.
- Aldama, Frederick Luis (2015): The science of storytelling. Perspectives from cognitive science, neuroscience, and the humanities. In: *Projections: Journal for Movies and Mind* 9 (1), 80–95.
- Aldama, Frederick Luis (2018): *Comics Studies Here and Now*. London, New York: Routledge.
- Baker, Paul/Egbert, Jesse (Hg.) (2016): *Triangulating Methodological Approaches in Corpus-Linguistic Research*. Abdingdon, New York: Routledge.
- Bateman, John A. (2007): Towards a *grande paradigmatique* of film: Christian Metz reloaded. In: *Semiotica* 167 (1/4), 13–64.
- Bateman, John A. (2014): *Text and Image: A Critical Introduction to the Visual/Verbal Divide*. London, New York: Routledge.
- Bateman, John A. (2016): Methodological and theoretical issues for the empirical investigation of multimodality. In: Klug, Nina-Maria/Stöckl, Hartmut (Hg.): *Handbuch Sprache im multimodalen Kontext*. Berlin: De Gruyter Mouton (HSW, 7), 36–74.
- Bateman, John A. (2019): Afterword: Legitimizing Multimodality. In: Wildfeuer, Janina/Pflaeging, Jana/Bateman, John A./Seizov, Ognian/Tseng, Chiao-I (Hg.): *Multimodality: Disciplinary Thoughts and the Challenge of Diversity*. Berlin: De Gruyter, 297–321.
- Bateman, John A./Beckmann, Annika/Varela, Rocío (2018): From Empirical Studies to Visual Narrative Organization: Exploring Page Composition. In: Dunst, Alexander/Laubrock,

- Jochen/Wildfeuer, Janina (Hg.): Empirical Comics Research. Digital, Multimodal, and Cognitive Methods. London, New York: Routledge, 127–153.
- Bateman, John A./Velooso, Francisco O.D./Wildfeuer, Janina/HiuLaam Cheung, Felix/Songdan Guo, Nancy (2017): An open multilevel classification scheme for the visual layout of comics and graphic novels: motivation and design. In: *Journal of Digital Scholarship in the Humanities* 32 (3), 476–510. URL: <http://dx.doi.org/10.1093/jllc/fqw024>.
- Bateman, John A./Wildfeuer, Janina (2014): A multimodal discourse theory of visual narrative. In: *Journal of Pragmatics* 74, 180–218. URL: <http://dx.doi.org/10.1016/j.pragma.2014.10.001>.
- Bergs, Alexander/Zima, Elisabeth (Hg.) (2017): Towards a multimodal construction grammar. Berlin: De Gruyter Mouton (Themenheft: *Linguistic Vanguard. A Multimodal Journal for the Language Sciences*).
- Blaikie, Norman W.H. (1991): A critique of the use of triangulation in social research. In: *Quality & Quantity: International Journal of Methodology* 25 (2), 115–136.
- Bucher, Hans-Jürgen (2011): Multimodales Verstehen oder Rezeption als Interaktion. Theoretische und empirische Grundlagen einer systematischen Analyse der Multimodalität. In: Diekmannshenke, Hans-Joachim/Klemm, Michael/Stöckl, Hartmut (Hg.): Bildlinguistik. Theorien – Methoden – Fallbeispiele. Berlin: Erich Schmidt, 123–156.
- Changeux, Jean-Pierre (1994): Art and neuroscience. In: *Leonardo* 27 (3), 189–201.
- Cohn, Neil (2014): Building a better ‘comic theory’: Shortcomings of theoretical research on comics and how to overcome them. In: *Studies in Comics* 5 (1), 57–75.
- Cohn, Neil (2016): A multimodal parallel architecture: A cognitive framework for multimodal interactions. In: *Cognition* 146, 304–323.
- Da, Nan Z. (2019): The Computational Case against Computational Literary Studies. In: *Critical Inquiry* 45, 601–639.
- Denzin, Norman K. (Hrsg.) (1970): Sociological Methods: A Sourcebook. London: Taylor & Francis.
- Deppermann, Arnulf (2013): Multimodal interaction from a conversation analytic perspective. In: *Journal of Pragmatics* 46, 1, 1–7.
- Deppermann, Arnulf/Streeck, Jürgen (Hg.) (2018): Time in Embodied Interaction: Synchronicity and Sequentiality of Multimodal Resources. John Benjamins.
- Dunst, Alexander/Laubrock, Jochen/Wildfeuer, Janina (Hg.) (2018): Empirical Comics Research. Digital, Multimodal, and Cognitive Methods. London and New York: Routledge.
- Erzberger, Christian/Kelle, Udo (2003): Making inferences in mixed methods: the rules of integration. In: Tashakkori, Abbas/Teddlie, Charles (Hg.): Handbook of Mixed Methods in Social and Behavioral Research. Thousand Oaks: Sage, 457–488.
- Ferstl, Evelyn C./von Cramon, D. Yves (2001): The role of coherence and cohesion in text comprehension: an event-related fMRI study. In: *Cognitive Brain Research* 11 (3), 325–340.
- Fine, Jonathan (1985): What do surface markers mean? Towards a triangulation of social, cognitive and linguistic factors. In: Benson, James D./Greaves, Williams S. (Hg.): Systemic perspectives on discourse: selected applied papers from the 9th International Systemic Workshop. Band 2. Norwood: Ablex, 102–115.
- Fish, Stanley (1980): Is there a text in this class?: The authority of interpretative communities. Cambridge, MA: Harvard University Press.
- Fix, Ulla (2008): Nichtsprachliches als Textfaktor: Medialität, Materialität, Lokalität. In: *Zeitschrift für Germanistische Linguistik* 36 (3), 343–354.
- Flanagan, Owen (1992): Consciousness reconsidered. Cambridge: MIT Press.

- Forceville, Charles J./Urios-Aparisi, Eduardo (Hg.) (2009): *Multimodal Metaphor*. Berlin, New York: De Gruyter.
- Fricke, Ellen (2012): *Grammatik multimodal: Wie Wörter und Gesten zusammenwirken*. Berlin, New York: De Gruyter.
- Gardt, Andreas (2012): Textsemantik. Methoden der Bedeutungerschließung. In: Bär, Jochen A./Müller, Marcus (Hg.): *Geschichte der Sprache – Sprache der Geschichte. Probleme und Perspektiven der historischen Sprachwissenschaft des Deutschen*. Oskar Reichmann zum 75. Geburtstag. Berlin: De Gruyter, 61–82.
- Gries, Stefan Th./Hampe, Beate/Schönefeld, Doris (2005): Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. In: *Cognitive Linguistics* 16 (4), 635–676. URL: doi:https://doi.org/10.1515/cogl.2005.16.4.635.
- Halliday, Michael A. K. (1992): Language as system and language as instance: the corpus as a theoretical construct. In: Svartvik, Jan (Hg.): *Directions in corpus linguistics: proceedings of Nobel Symposium 82, Stockholm, 4–8 August 1991*, Berlin: De Gruyter, 61–78.
- Hammersley, Martyn (2008): Troubles with triangulation. In: Bergman, Manfred Max (Hg.): *Advances in mixed methods research*. London: Sage, 22–36.
- Hedger, Stephen C. Van/Nusbaum, Howard C./Heald, Shannon L. M./Huang, Alex/Kotabe, Hiroki P./Berman, Marc G. (2019): The Aesthetic Preference for Nature Sounds Depends on Sound Object Recognition. In: *Cognitive Science* 43 (5), e12734.
- Holly, Werner (2009): Der Wort-Bild-Reißverschluss. Über die performative Dynamik der audiovisuellen Transkriptivität. In: Feilke, Helmuth/Linke, Angelika (Hg.): *Oberfläche und Performanz*. Tübingen: Niemeyer, 93–110.
- Jackendoff, Ray (2007): A parallel architecture perspective on language processing. In: *Brain Research* 1146, 2–22.
- Janesick, Valerie J. (1998): The dance of qualitative research design: Metaphor, methodolatry, and meaning. In: Denzin, Norman K./Lincoln, Yvonne S. (Hg.): *Strategies of Qualitative Inquiry*. Los Angeles: Sage, 35–55.
- Jewitt, Carey/Henriksen, Berit (2016): Social Semiotic Multimodality. In: Klug, Nina-Maria/Stöckl, Hartmut (Hg.): *Handbuch Sprache im multimodalen Kontext*. Berlin: De Gruyter (HSW, 7), 145–164.
- Kappelhoff, Hermann/Müller, Cornelia (2011): Embodied meaning construction. Multimodal metaphor and expressive movement in speech, gesture, and in feature film. In: *Metaphor and the Social World* 1 (2), 121–153.
- Kesselheim, Wolfgang (2011): Sprachliche Oberflächen: Musterhinweise. In: Habscheid, Stephan (Hg.): *Textsorten, Handlungsmuster, Oberflächen. Linguistische Typologien der Kommunikation*. Berlin, New York: De Gruyter, 337–366.
- Klug, Nina-Maria (2013): Bilder als Texte: Methoden einer semiotischen Erweiterung angewandter Diskursanalyse. In: Roth, Kersten Sven/Spiegel, Carmen (Hg.): *Angewandte Diskurslinguistik*. Berlin: De Gruyter, 163–187.
- Klug, Nina-Maria (2016): Multimodale Text- und Diskurssemantik. In: Klug, Nina-Maria/Stöckl, Hartmut (Hg.): *Handbuch Sprache im multimodalen Kontext*. Berlin: De Gruyter (HSW, 7), 165–189.
- Krampen, Martin (1981): Ferdinand de Saussure und die Entwicklung der Semiologie. In: Krampen, Martin/Oehler, Klaus/Posner, Roland/Uexküll, Thure von (Hg.): *Die Welt als Zeichen: Klassiker der modernen Semiotik*. Berlin: Wolf Jobst Siedler Verlag, 99–142.

- Kress, Gunther (2010): *Multimodality: a social semiotic approach to contemporary communication*. London: Routledge.
- Kress, Gunther/Leeuwen, Theo van (2001): *Multimodal discourse: the modes and media of contemporary communication*. London: Arnold.
- Kress, Gunther/Leeuwen Theo van (2006 [1996]): *Reading Images: the grammar of visual design*. London, New York: Routledge.
- Ledin, Per/Machin, David (2017): Multi-modal critical discourse analysis. In: Flowerdew, John/Richardson, John E. (Hg.): *The Routledge Handbook of Critical Discourse Studies*. London: Routledge, 60–76.
- Leeuwen, Theo van (2004): Ten reasons why linguists should pay attention to visual communication. In: LeVine, Philip/Scollon, Ron (Hg.): *Discourse and technology: multimodal discourse analysis*. Washington, D.C.: Georgetown University Press, 7–19.
- Leeuwen, Theo van (2005): *Introducing social semiotics*. London: Routledge.
- Leeuwen, Theo van (2020): The new visuality of writing. In: Stöckl, Hartmut/Caple, Helen/Pflaeging, Jana (Hg.): *Shifts towards Image-centricity in Contemporary Multimodal Practices*. London, New York: Routledge, 64–85.
- Loschky, Lester C./Larson, Adam M./Magliano, Joseph P./Smith, Tim J. (2015): What Would Jaws Do? The Tyranny of Film and the Relationship between Gaze and Higher-Level Narrative Film Comprehension. In: *PLoS One* 10 (11), e142474.
- Magliano, Joseph P./Dijkstra, Katinka/Zwaan, Rolf A. (1996): Generating predictive inferences while viewing a movie. In: *Discourse Processes* 22, 199–224.
- Martinec, Radan/Salway, Anthony (2005): A system for image-text relations in new (and old) media. In: *Visual Communication* 4 (3), 337–371.
- Mathison, Sandra (1988): Why triangulate? In: *Education Researcher* 17 (2), 13–17.
- Meier, Stefan (2014): *Visuelle Stile. Zur Sozialsemiotik visueller Medienkultur und konvergenter Design-Praxis*. Bielefeld: Transcript.
- Mikkonen, Kai (2017): *The Narratology of Comic Art*. London and New York: Routledge.
- Mondada, Lorenza (2016): Challenges of multimodality: Language and the body in social interaction. In: *Journal of Sociolinguistics* 20 (2), 2–32.
- Norris, Sigrid (2019): *Systematically Working with Multimodal Data: Research Methods in Multimodal Discourse Analysis*. Hoboken, Chichester: Wiley.
- O'Toole, Michael (2011 [1994]): *The language of displayed art*. Abingdon, Oxon: Routledge.
- Packard, Stephan/Rauscher, Andreas/Sina, Veronique/Thon, Jan-Noel/Wilde, Lukas/Wildfeuer, Janina (2019): *Comicanalyse. Eine Einführung*. Berlin: J.B. Metzler.
- Patel-Grosz, Pritty/Grosz, Patrick Georg/Kelkar, Tejaswinee/Jensenius, Alexander Refsum (2019): Coreference and disjoint reference in the semantics of narrative dance. In: Sauerland, Uli/Solt, Stephanie (Hg.): *Proceedings of Sinn und Bedeutung* 22, Band 2. Berlin: ZAS, 199–216.
- Perniss, Pamela (2018): Why We Should Study Multimodal Language. In: *Frontiers in Psychology* 9, 1109. URL: <https://doi.org/10.3389/fpsyg.2018.01109>.
- Posner, Roland (1991): *Kultur als Zeichensystem. Zur semiotischen Explikation kulturwissenschaftlicher Grundbegriffe*. In: Assmann, Aleida/Harth, Dietrich (Hg.): *Kultur als Lebenswelt und Monument*. Frankfurt: Fischer, 37–74.
- Royce, Terry D. (2016): Intersemiotic Complementarity in Print Advertisements. In: Klug, Nina-Maria/Stöckl, Hartmut (Hg.): *Handbuch Sprache im multimodalen Kontext*. Berlin: De Gruyter (HSW, 7), 348–371.

- Saussure, Ferdinand de (2001 [1916]): Grundfragen der allgemeinen Sprachwissenschaft. Hrsg. von Charles Bally und Albert Sechehaye unter Mitwirkung von Albert Riedlinger. Übersetzt von Herman Lommel. Berlin: De Gruyter.
- Schlenker, Phillippe (2017): Outline of Music Semantics. In: *Music Perception* 35, 3–37.
- Schmitz, Ulrich (2003): Blind für Bilder. Warum sogar Sprachwissenschaftler auch Bilder betrachten müssen. In: *LAUD. Series A: General & Theoretical Papers* No. 581. URL: <http://www.linse.uni-due.de/linguistische-publikationen/blind-fuer-bilder-warum-sogar-sprachwissenschaftler-auch-bilder-betrachten-muessen.html>
- Schmitz, Ulrich (2004): Sprache in modernen Medien. Einführung in Tatsachen und Theorien, Themen und Thesen. Berlin: Erich Schmidt Verlag.
- Schmitz, Ulrich (2016): Multimodale Texttypologie. In: Klug, Nina-Maria/Stöckl, Hartmut (Hg.): *Handbuch Sprache im multimodalen Kontext*. Berlin: De Gruyter (HSW, 7), 327–347.
- Siefkes, Martin/Arielli, Emanuele (2018): The Aesthetics and Multimodality of Style. *Experimental Research on the Edge of Theory*. Berlin etc.: Lang.
- Smith, Murray (2012): Triangulating aesthetic experience. In: Shimamura, Arthur P./Palmer, Stephen E. (Hg.): *Aesthetic science: Connecting minds, brains, and experience*. London, New York: Oxford University Press, 80–106.
- Snow, Charles Percy (2001 [1959]): *The two cultures*. London: Cambridge University Press.
- Spillner, Bernd (1982): Stilanalyse semiotisch komplexer Texte. Zum Verhältnis von sprachlicher und bildlicher Information in Werbeanzeigen. In: *Kodikas/Code. Ars Semeiotica* 4/5 (1), 91–106.
- Spitzmüller, Jürgen/Warke, Ingo H. (2011): *Diskurslinguistik: eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse*. Berlin: De Gruyter.
- Stöckl, Hartmut (2004a): 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 (2004b): In between modes: language and image in printed media. In: Ventola, Eija/Charles, Cassily/Kaltenbacher, Martin (Hg.): *Perspectives on Multimodality*. Amsterdam: John Benjamins, 9–30.
- Stöckl, Hartmut (2016): Multimodalität – Semiotische und textlinguistische Grundlagen. In: Klug, Nina-Maria/Stöckl, Hartmut (Hg.): *Handbuch Sprache im multimodalen Kontext*. Berlin: De Gruyter (HSW, 7), 3–35.
- Tseng, Chiao-I (2013): *Cohesion in Film: Tracking Film Elements*. Basingstoke: Palgrave Macmillan.
- Tseng, Chiao-I/Bateman, John A. (2018): Cohesion in Comics and Graphic Novels: An Empirical Comparative Approach to Transmedia Adaptation in *City of Glass*. In: *Adaptation* 11 (2), 122–143. URL: <https://doi.org/10.1093/adaptation/apx027>.
- Tseng, Chiao-I/Kautonnen, Janne/Bateman, John A. (2017): Textual analysis of event patterns and neural functions of narrative constructions in film. *Paper presented at the 2017 conference of the society for the cognitive study of the moving image (SCSMI 2017)*, Aalto University, Helsinki.
- Tseng, Chiao-I/Laubrock, Jochen/Bateman, John A. (2021): The impact of multimodal cohesion on attention and interpretation in film. In: *Discourse, Context & Media* 44 (December): 100544. aop.
- Tseng, Chiao-I/Laubrock, Jochen/Pflaeging, Jana (2018): Character Developments in Comics and Graphic Novels: A Systematic Analytical Scheme. In: Dunst, Alexander/Laubrock,

- Jochen/Wildfeuer, Janina (Hg.): Empirical Comics Research. Digital, Multimodal, and Cognitive Methods. London, New York: Routledge, 154–175.
- Wildfeuer, Janina (2014): Film Discourse Interpretation. Towards a New Paradigm for Multimodal Film Analysis. London, New York: Routledge.
- Wildfeuer, Janina (2017): Diskurssemiotik = Diskurssemantik + multimodaler Text. In: Hess-Lüttich, Ernest/Warnke, Ingo H./Reisigl, Martin/Kämper, Heidrun (Hg.): Diskurs – semiotisch. Aspekte multimodaler Diskursforschung. Berlin, New York: De Gruyter, 189–209.
- Wildfeuer, Janina/Bateman, John A./Hiippala, Tuomo (2020): Multimodalität: Grundlagen, Forschung und Analyse. Eine problemorientierte Einführung. Berlin: De Gruyter.
- Winkler, Hartmut (2008): Zeichenmaschinen: oder warum die semiotische Dimension für eine Definition der Medien unerlässlich ist. In: Münker, Stefan/Roesler, Alexander (Hg.): Was ist ein Medium? Frankfurt a.M.: Suhrkamp, 211–222.
- Zacks, Jeffrey M./Speer, Nicole K./Reynolds, Jeremy R. (2009): Segmentation in Reading and Film Comprehension. In: *Journal of Experimental Psychology* 138, 307–327.

