

Medien und Kommunikation

Eine interdisziplinäre Einführung

Herausgegeben von
Hans Krahl und Michael Titzmann

Zweite, überarbeitete Auflage

Verlag Karl Stutz

Passau
2013

13 Interaktion und Kooperation von Texten und Bildern

13.1 Die Begriffe ‚Text‘ und ‚Bild‘

13.2 Strukturbedingte Leistungen und Grenzen von ‚Texten‘ und ‚Bildern‘

13.3 Kooperations- und Interaktionsformen von Text und Bild

13.4 Zwei Beispiele für Text-Bild-Interaktion und -Kooperation

Kaum eine semiotische Äußerung, also ein ‚Text‘ im weitesten Sinne, bedient sich nur eines Zeichensystems. So werden mündliche sprachliche Äußerungen notwendig von den Zeichen der Intonation, Gestik, Mimik usw. begleitet (und partiell interpretiert), so bedienen sich schriftliche Texte, ob handschriftlich oder gedruckt tradiert, unterschiedlicher Grapheme (z.B. Wahl der Schrifttypen/-größen im Druck) und einer bestimmten Verteilung der Zeichen auf dem Schreibmaterial (z.B. Abstände, Absätze, usw.); und alle diese non-verbalen Elemente sind mehr oder weniger bedeutungstragend. In ein und derselben Äußerung können im Prinzip Elemente beliebig vieler der kulturell jeweils verfügbaren Zeichensysteme kombiniert werden. Videoclips oder Filme z.B. können praktisch alle Zeichensysteme nutzen: Bild, Sprache, Ton, Musik, Gestik, Mimik, Proxemik, die mit dem Aussehen von anthropomorphen Figuren verbundenen kulturellen Zeichensysteme, vestimentäre, kulinarische Kodes, Piktogramme, die Kodes der Verkehrszeichen, der Einrichtung von Wohnräumen, der Wahl der Automarke usw. Einige Aspekte der *Interaktion und Kooperation von Zeichensystemen* sollen am Beispiel von Text-Bild-Relationen exemplarisch skizziert werden. ‚Text‘ wird im folgenden Beitrag nur im eingeschränkten (kulturell normalen) Sinne der sprachlichen Äußerung gebraucht.

Bilder sind Gegenstand unterschiedlichster Disziplinen geworden. Neben Semiotik, Kunstwissenschaft, Filmwissenschaft, Werbungsforschung usw. versucht sich in jüngster Zeit eine Bildwissenschaft zu etablieren, die einstweilen aus der Addition aller Disziplinen besteht, die mit Bildern bzw. den Strukturen ihrer neuronalen, psychischen, sozialen Verarbeitung zu tun haben (Sachs-Hombach/Rehkämper 1998, Sachs-Hombach 2005, Frank/Lange 2010). Bild-Text-Relationen insbesondere wurden Thema in Semiotik (z.B. Titzmann 1990, Decker 2005), in der kunstwissenschaftlichen Richtung der Ikonologie (Klibansky/Panofsky/Saxl 1990; Panofsky 1978, 1979, 1980, 1997; Wind 1981; Wittkower 1983), der Filmwissenschaft (z.B. Kanzog 1991 und 2007), der Werbeforschung (z.B. Karmasin 1993).

13.1 Die Begriffe ‚Text‘ und ‚Bild‘

1. ‚Text‘ soll *hier* nun jede Äußerung heißen, die (dominant) aus einer begrenzten Menge sprachlicher Zeichen einer oder mehrerer, ‚natürlicher‘ oder ‚künstlicher‘ Sprachen (z.B. die Zeichensysteme der formalen Logik, der Mathematik, der Informatik) besteht; sowohl in auditiver wie in visueller Wahrnehmung bilden die Zeichen von Texten eine *lineare Sukzession*; in mündlichen wie in schriftlichen Texten ist die Reihenfolge korrekter Wahrnehmung der Zeichen festgelegt. Ein Text ist also eine zeitlich und/oder räumlich *geordnete Menge von Zeichen*: eine *Folge* $T := \langle z_1, z_2, \dots, z_n \rangle$. ‚Text‘ soll eine solche Zeichenfolge unabhängig davon heißen,

- ob die Zeichenmenge als ‚vollständige‘ oder ‚unvollständige‘ (z.B. Fragment) wahrgenommen wird;
- ob die Zeichen zu syntaktisch wohlgeformten Einheiten (also z.B.: grammatisch korrekten Sätzen) kombiniert sind.

2. ‚Bild‘ soll *hier* jede Äußerung heißen, die (dominant) aus einer Menge visuell wahrnehmbarer, non-verbaler – *räumlich geordneter* – Elemente besteht,

- die durch einen selbst visuell wahrnehmbaren ‚Rahmen‘ (Gemälderahmen, Bildrand bei Foto, Film etc.) begrenzt wird, und
- die in dem Sinne zeichenhaft sind, dass sie
 - etwas im logisch-mathematischen Sinne ‚*abzubilden*‘, und/oder
 - etwas im semiotischen Sinne zu ‚*bedeuten*‘ scheinen, und
- die dem Betrachter *simultan* gegeben sind; und
- die *invariant* sind, das heißt weder sich noch ihren Ort in der räumlichen Anordnung ändern.

‚Bild‘ in diesem Sinne sind also z.B. gegenständliche oder abstrakte Gemälde, Landkarten, Darstellungen des Systems städtischer Verkehrslinien (U-/S-Bahn, Straßenbahn, Bus), Piktogramme, Verkehrsschilder, Schemata und Diagramme (z.B. in wissenschaftlichen Arbeiten wie in diesem Band), jede Teileinheit eines Comic Strip, Fotos (inklusive deren Reproduktion in Printmedien), die kleinsten (sukzessiven) Einheiten einer Dramen- oder Operninszenierung, eines Videoclips, eines Werbespots, einer Fernsehsendung, eines Spielfilms. Pragmatisch formuliert: ‚Bild‘ ist jede Menge non-verbaler Elemente, die sich durch ein einzelnes Foto dokumentieren ließe und als zeichenhaft/bedeutungstragend wahrgenommen wird; sobald sich eine Teileinheit eines Bildes transformiert, z.B. die Gestik oder Mimik einer Person, ihr Ort im Raum, ihre Gewandung usw., liegt im Sinne unseres Begriffsgebrauchs (mindestens) ein

zweites, logisch dem ersten sukzessives Bild vor; jedes *Einzelbild* stellt also eine *statische Situation*, sozusagen eine ‚Momentaufnahme‘ aus dem temporalen Kontinuum von ‚Realität‘, dar. Ein Comic Strip, ein Werbespot (zu Werbung: Kap. 9; und Karmasin 1993), ein Videoclip (vgl. Decker 2005), ein Spielfilm (Kap. 14; und Kanzog 1991, 2007), eine Dramen- oder Operninszenierung, ein Gemäldezyklus, der etwa die Stationen der ‚passio Christi‘ darstellt, besteht folglich aus einer Sukzession solcher Bilder. Comic Strips, Theateraufführungen, Filme usw. sind also unter anderem Bildfolgen. Die visuell wahrnehmbaren Gegebenheiten, mittels derer sich diese non-verbale Elemente konstituieren, die das Bild ausmachen, sind Linien und Formen, Helligkeits- und/oder Farbunterschiede (oder auch sonstige Materialien bei Collagen etc.).

3. Thema ist hier *nur*, was in Bildern und Texten *semiotisch gegeben* ist; wie das semiotisch Gegebene – *psychologisch* und *neurologisch* gesehen – rezipiert und verarbeitet wird, ist Gegenstand anderer Disziplinen: der sozialpsychologischen Rezeptions- und Wirkungsforschung (Kap. 9; Karmasin 1993), der Wahrnehmungspsychologie (vgl. hier Kap. 8) und der Neurobiologie (vgl. die einschlägigen Beiträge in Sachs-Hombach/Rehkämper 1998, Sachs-Hombach 2005).

13.2 Strukturbedingte Leistungen und Grenzen von ‚Texten‘ und ‚Bildern‘

1. Das sprachliche Zeichensystem, dessen sich Texte bedienen, und das nicht-sprachliche Zeichensystem, dessen sich Bilder bedienen, haben eine *fundamental verschiedene Struktur*.

1.1 Das jeweilige natürliche oder künstliche Sprachsystem besteht aus *diskreten Einheiten*, das heißt aus eindeutig abgrenzbaren Zeichen, z.B. den Lexemen (= ‚Wörtern‘) der natürlichen Sprachen. Diesen primären Signifikanten ist jeweils eine begrenzte Menge von möglichen Signifikaten zugeordnet (Kap. 1), die *lexikalisch* sind, das heißt in Wörterbüchern kodifiziert werden können. Die syntagmatische Kombination der Signifikanten (Kap. 2), der Kontext, das kulturelle Wissen (Kap. 4) legen mehr oder weniger eindeutig fest, welche der lexikalisch möglichen Signifikate diese Äußerung vermittelt. Aus diesen ‚einfachen, primären Einzelzeichen‘ können einerseits durch Kombination ‚komplexe primäre Zeichen‘ – anthropomorphe Figuren, Räume, Situationen, Ereignisse – gebildet werden, andererseits – z.B. durch Abweichung im Gebrauch der Zeichen – kontextuell dekodierbare ‚sekundäre Zeichen‘ (Kap. 1).

1.2 Ganz anders bei Bildern, seien es Schwarzweiß- oder Farbfotos, Werbung in Printmedien, Filmbilder, Gemälde im Museum: was wir hier semio-

tisch haben, sind zunächst *nicht-diskrete Einheiten*: durch Linien, Umrisse usw. begrenzte oder nicht begrenzte Helligkeits- und Farbunterschiede. Unmittelbar wahrnehmbare beliebige Elemente von Bildern oder Bildausschnitte sind also nicht (primäre) Zeichen, wie dies etwa Lexeme sind oder die Größen „e“, „=“, „m“, „.“, „c“, „²“ in „ $e = m \cdot c^2$ “. Bildelemente/-ausschnitte sind also zunächst nur *Zeichenäquivalente: potentielle Signifikanten für potentielle Signifikate*. Aus diesen Linien, Formen, Helligkeitsunterschieden, Farben werden erst in einem *Akt der Projektion* des Betrachters primäre Signifikanten, also diskrete zeichenhafte Einheiten, in die das Bild zerlegt wird. Eine bestimmte Kombination von Bildelementen identifizieren wir etwa als Abbildung einer Frau, einer Landschaft, eines Autos usw. Diese Projektion verläuft einerseits nach *wahrnehmungspsychologischen und neurologischen Gesetzmäßigkeiten* (etwa ‚Gestaltwahrnehmung‘ – Kap. 8), andererseits auf der Basis *kultureller – somit konventioneller – Kodierungen*. Dafür seien einige Beispiele genannt. Antike und Mittelalter kennen in Gemälde oder Mosaik keine dreidimensionale, räumliche Darstellung, die erst in der Renaissance erfunden wird, wohl aber sonstige konventionelle Kodierungen, bei denen z.B. der Größenunterschied zwischen anthropomorphen Figuren und/oder deren Positionierung in Bildmitte oder am Bildrand hierarchische Unterschiede ausdrückt: vgl. z.B. die Darstellung des christlichen Gottes als ‚Pantokrator‘ (‚Allbeherrscher‘) in spätantiken Mosaiken wie den Kirchenbauten in Ravenna.

Im einfachsten Falle weist eine Kombination von Bildelementen Merkmale auf, die den Merkmalen des Signifikats sprachlicher Zeichen entsprechen: Wir können diese Kombination dann als ‚Frau‘ oder als ‚Auto‘ oder ... identifizieren und benennen: So entstehen *primäre bildliche Signifikanten*, indem einer Farb-Form-Kombination ein sprachlich benennbares Signifikat zugeordnet wird, weshalb sie *danach* als Signifikant identifiziert werden kann. Auch Bilder können *sekundäre Signifikanten* aufweisen, also Zeichen, deren primäres Signifikat zum sekundären Signifikanten für sekundäre Signifikate wird. Solche sekundäre Zeichenhaftigkeit ist das Produkt *kultureller Kodierung*, die der Betrachter kennen muss. In einem Gemälde der Frühen Neuzeit mag etwa das primäre Signifikat ‚Anker‘ sekundär ‚Hoffnung, Vertrauenswürdigkeit, Verlässlichkeit‘ bedeuten, ein ‚Knabe mit Pfeil und Bogen‘ den antiken Gott Amor als Zeichen für erotische Anziehung, ein ‚Skelett mit Pfeil und Bogen‘ hingegen den Tod, wobei die Waffe in beiden Fällen signalisiert, dass das menschliche Subjekt als passives Objekt gedacht wird. Solche sekundäre Zeichenhaftigkeit von Bildelementen kann wiederum *am Anfang nur mit Hilfe sprachlicher Texte oder Kotexte* eingeführt werden: Dass ‚Anker‘ ‚Hoffnung‘ bedeutet, ist

eine arbiträre Setzung von Zeichenhaftigkeit, die das Produkt eines sozialen – sprachlich vermittelten – Konsenses sein muss. Dass etwa Attribute wie ‚Zep-ter‘ und ‚Krone‘ ‚Herrschaft‘ bedeuten, setzt kulturelles Wissen voraus, das letztlich auf Texten basiert.

Wo also im Falle eines Textes jeder sprachkompetente Benutzer sofort die primären Einzelzeichen erkennen kann, muss er im Falle von Bildern erst eruieren, welches potentielle Zeichenäquivalent denn auch tatsächlich primäres Zeichen ist. Wo also im Falle des Textes von vornherein die primären Signifikanten gegeben sind, denen Signifikate zugeordnet werden können, sind im Falle des Bildes zunächst – aufgrund wahrnehmungspsychologisch-neurobiologischer Strukturen und aufgrund des sprachlich artikulierbaren kulturellen Wissens über die Welt – die Signifikate gegeben, denen Bildelemente als Signifikanten zugeordnet werden. Auch die bildliche – z.B. filmische – Äußerung kann natürlich ein Objekt – einen Gegenstand, eine Person, was auch immer – als Signifikanten aufbauen, dessen Signifikat der Rezipient oder Interpret suchen muss: Solche Signifikanten sind aber logisch notwendig sekundäre.

2. Kommen wir zu etwas konkreteren Unterschieden zwischen Bildern und Texten.

2.1 Texte wie Bilder können sich in den von ihnen – mit ihren semiotischen Möglichkeiten – dargestellten Welten auch je anderer Zeichensysteme bedienen. Bilder können dabei alle kulturellen Codes *unmittelbar* darstellen, die auf visuellen Zeichen basieren, etwa eben Piktogramme, Verkehrsampeln, Einrichtungsstile von Räumen, Stile der Selbstpräsentation, im Outfit, Gestik, Mimik, Proxemik. Texte können visuell Wahrnehmbares, es sei kodiert oder nicht, nur *mittelbar* darstellen.

(1) Hans machte eine verzweifelte Geste; Grete schaute vielsagend und winkte ab; – nur ein Bild könnte zeigen, was die sprachliche Umschreibung als wahrnehmbare Mimik bzw. Gestik genau bezeichnet (= Referent: vgl. Kap. 1).

Texte können mittelbar auch über Töne, Geräusche, Musik oder über olfaktorische und gustatorische Wahrnehmungen reden, die Bilder nicht darstellen können. Sofern Texte und Bilder in ihren dargestellten Welten also andere, nicht-sprachliche bzw. nicht-bildliche Zeichen überhaupt darstellen können, können es Texte nur mittelbar, Bilder unmittelbar.

2.2 Das bedeutet, dass Texte wie Bilder jeweils *systembedingte Nullpositionen* haben, nicht nur *textspezifische* (Kap. 4). Ein sprachlicher Text kann etwa eine anthropomorphe Figur noch so genau in ihren visuell wahrnehmbaren Merkmalen zu beschreiben versuchen: Gegenüber ihrer bildlichen Darstellung etwa

einer Theateraufführung oder eines Films wird der Text, etwa ein Drama, ein Roman, immer notgedrungen Nullpositionen aufweisen. Nur ein Beispiel: In einem kleinen Roman von Friedrich Spielhagen, *Clara Vere* (1853), wird eine junge Frau „in der reizenden Fülle ihrer schlanken Schönheit“ (Spielhagen 1890, S. 73) wahrgenommen. „Schön“ bedeutet natürlich ‚ästhetisch erfreulich‘, ‚reizend‘, ‚erotisch anziehend‘: Aber wie muss die Dame aussehen, um diese Bedingungen zu erfüllen? „Fülle“ und „Schlankheit“ würden sich in unserem heutigen Sprachgebrauch und kulturellem Wissen eher als oppositionell ausschließen (wo Frauenzeitschriften magersüchtige Models als Ideal präsentieren), während sie in unserem Text offenbar kompatible Merkmale sind. Wie wir uns eine solche junge Schöne vorzustellen hätten, könnte also nur ein Bild entscheiden; hätten wir aber nur ein Bild der jungen Frau ohne Text, wüssten wir nicht, ob sie kulturell als ‚schön‘, ‚reizend‘, ‚schlank plus füllig‘ zu klassifizieren ist. Das Bild zum Text würde also eine *Referentialisierung* erlauben: Dem Signifikat könnte ein potentieller Referent zugeordnet werden. Der Text zum Bild würde hingegen umgekehrt eine *Semantisierung* erlauben: Dem Bildsignifikat als Referenten könnte eine semantische Klassifikation zugeordnet werden.

2.3. Ein Bild kann eine kulturell als ‚schön‘ und/oder ‚erotisch attraktive‘ Frau zeigen, es kann *nicht sagen*, aufgrund welcher Merkmale sie das ist. Ein Text kann *nicht zeigen*, wie kulturell eine ‚schöne‘ bzw. ‚erotische‘ Frau beschaffen sein muss; aber er kann *sagen*, sie sei es bzw. sie müsse die Merkmale x, y, z, ... aufweisen, um es zu sein. Das heißt mit anderen Worten: Ein Bild kann grundsätzlich nur *konkrete, singuläre Größen/Individuen* (Personen, Situationen, Objekte) darstellen, ein Text kann *abstrakte Klassen bilden und Regeln formulieren* („alle Frauen, die x sind, sind schön“): Er ist fähig zu *generalisierender Abstraktion*.

2.4 Damit verknüpft ist eine weitere semiotische Differenz von Texten und Bildern. Sprache kann ‚normal‘ gebraucht werden, um mittels ihrer eine *dargestellte Welt* zu konstruieren („objektsprachlicher“ Gebrauch), oder als ‚*Metasprache*‘, mittels derer objektsprachliche Äußerungen über eine reale oder fiktive Welt kommentiert oder das verwendete Zeichensystem, die Sprache, selbst thematisiert wird. Texte können also neben *objektsprachlichen Propositionen* auch *metasprachliche Propositionen* enthalten:

- (2) Als Hanna Hans fragte, ob er mit einer Scheidung einverstanden sei, nickte Hans, obwohl seine Seele „nein!“ schrie [a]. In tiefer emotionaler Verstörung scheiterte er wieder einmal kommunikativ [b]. Doch gestische Zeichen sind nun einmal eindeutig [c].

Satz [b] impliziert eine metasprachliche Proposition, insofern sie den Zeichengebrauch von Hans kommentiert:

- Hans wählt ein gestisches Zeichen, dessen Signifikat in Opposition zu seiner Intention steht,
- und diese Proposition kommentiert die aus [a] ableitbare Proposition,
- Hans drückt durch ein gestisches Zeichen seine Zustimmung zum Scheidungswunsch Hannas aus.

Aus [c] würde eine weitere metasprachliche Proposition folgen, die diesmal nicht den Sprachgebrauch von Hans, sondern das von ihm verwendete Zeichensystem selbst kommentiert:

- Gesten haben eindeutige Signifikate. [Vorsicht: der Satz stimmt selbst nicht.]

2.5 ‚Sprache‘ ist nun aber nicht nur ein Zeichensystem, das seine eigenen Äußerungen und seine eigene Struktur zum Thema machen kann: ‚Sprache‘ ist generell die *letzte Metasprache für alle Zeichensysteme bzw. zeichenhaften Äußerungen*. Denn

a) *jedes nicht-sprachliche Zeichensystem kann letztlich nur mit Hilfe von Sprache eingeführt werden*: Ein neues Piktogramm, ein neues Verkehrsschild, ein neues modisches Outfit, eine neue Semantik eines Waschmittels oder einer Automarke, neue logische oder mathematische Zeichen, können nur durch Texte zum Zeichensystem werden – die erstrebte Konvention setzt soziale, sprachliche Kommunikation voraus, wie jede(r) Werbetreibende weiß. Wenn scheinbar ein neues non-verbales Zeichen ohne die Hilfe sprachlicher Zeichen eingeführt wird, dann geschieht es mittels von Zeichen, die schon früher sprachlich eingeführt worden sind.

b) *Aussagen über die Struktur von Zeichensystemen oder über den Gebrauch von Zeichen in einer Äußerung* sind nur mit Hilfe der Sprache möglich (linguistische Aussagen über natürliche Sprachen, semiotische Aussagen über Zeichensysteme im Allgemeinen usw.). Über formale logische und mathematische Sprachen kann zwar in formalen – logischen oder mathematischen – Sprachen gesprochen werden: über diese aber wiederum in letzter Instanz nur in der natürlichen Sprache. Jede Definition eines wissenschaftlichen Begriffs z.B. mag aus wiederum wissenschaftlichen Begriffen bestehen; aber jeder davon basiert am Ende auf normalsprachlichen Formulierungen. Die natürliche Sprache ist die ultimative Metasprache aller Zeichensysteme.

c) *Jede Bedeutung einer semiotischen Äußerung*, welcher Zeichensysteme sie sich auch immer bedienen möge, wird in letzter Instanz durch eine *Menge sprachli-*

cher Äußerungen repräsentiert und ausgedrückt. Die *Bedeutung eines Bildes* z.B. ist notwendig einer *Menge aus dem Bild ableitbarer Propositionen* (Kap. 4) äquivalent und nur durch eine solche artikulierbar. Sowohl die Bedeutung von Texten als auch die von Bildern wird also wiederum durch einen (Meta-)Text abgebildet.

2.6 Hier sei nochmals an die Unterscheidung von *Bedeutung* und *Wirkung* erinnert. Wenn etwa Gefühle einer Figur der dargestellten Welt in einem Text sprachlich benannt oder umschrieben werden oder in einem Bild gestisch und mimisch ausgedrückt werden, sind diese (sprachlich benennbaren) Gefühle Teil der *Bedeutung der Äußerung*. Wenn hingegen Elemente der Äußerung, z.B. eben die dargestellten Gefühle von Figuren oder deren sonstige Lebensäußerungen, Gefühle beim Hörer/Leser des Textes oder Betrachter des Bildes auslösen, sind diese nicht Teil der *Bedeutung*, sondern der *Wirkung*, die individuell verschieden und subjektiv ausfallen mag: Ein und dieselbe Figur einer dargestellten Welt mag bei dem einen Rezipienten Sympathie, beim anderen Antipathie auslösen, wie der eine schön finden mag, was dem anderen hässlich erscheint. Je traditioneller und homogener eine Kultur/Gruppe ist, desto mehr wird es auch an emotionalen und ästhetischen Kodierungen geben (vgl. auch Kap. 4); je ausdifferenzierter und inhomogener die Gruppe/Kultur ist, desto individueller und subjektiver wird auch ausfallen, welche Wirkungen die *Bedeutung der Äußerung* auslöst (und muss somit von der Rezeptions- und Wirkungsforschung untersucht werden: Kap. 8 und 9). Mit anderen Worten: beim Rezipienten ausgelöste kognitive, affektive, evaluative Reaktionen, die im ersten Falle starker kultureller Normierung zur *Bedeutung* gehören mögen, gehören im Falle schwacher Normierung nur zur *Wirkung*. Zwischen Texten und Bildern vergleichbarer Qualität wird dabei ein Unterschied zu erwarten sein: Wenn es beispielsweise in einem Text hieße

(3) Hans prügelte wieder einmal Hanna; sie litt fürchterlich;

und wenn der selbe Sachverhalt etwa in einer filmischen Bilderfolge dargestellt wäre, dann ist damit zu rechnen, dass zum einen das Gefühl Hannas als Teil der *Bedeutung* im Bild intensiver vermittelt wird und zum anderen auch die affektive *Wirkung* auf den Rezipienten intensiver sein wird. Der Text kann Gefühle besser klassifizieren und thematisieren, das Bild kann sie besser darstellen; der Text kann uns sagen, wer welche Gefühle hat, im Falle des Bildes müssen wir sie selbst identifizieren; im Text erfahren wir bei schriftlicher Kommunikation oder bei mündlichem Bericht über Dritte diese Gefühle nur indirekt, im Bild bzw. in der Bildfolge erleben wir sie als ‚Zeuge‘ mit.

2.7 Die Welt eines Textes kann *temporalisiert* sein, das heißt, ihre Elemente können ein *zeitliches Nacheinander* aufweisen. Im Extremfall kann ein und derselbe Satz auf verschiedenste Zeiträume Bezug nehmen:

(4) Nachdem Hanna sich von Hans hatte scheiden lassen (t_2), der früher Grete verlassen hatte (t_1), verfiel er in jener Nacht (t_4) nach Monaten der Unentschlossenheit (t_3) doch Lydias Reizen, die er sich später in neuer Ehe anzueignen suchen wird (t_5).

In der dargestellten Welt haben wir faktisch drei unterscheidbare vergangene Zeitpunkte bzw. Zeiträume (t_1 , t_2 , t_3), einen – bezogen auf die dargestellte Welt – ‚gegenwärtigen‘ (t_4), einen zukünftigen (t_5) und den – später als alle diese liegenden – der Sprechsituation selbst (t_6). Ein Bild hingegen kennt nur die *Gleichzeitigkeit aller Elemente*, keine Vergangenheiten und Zukünfte.

2.8 Da *Temporalisierung* deren (notwendige, aber nicht hinreichende) Voraussetzung ist, kennt das Bild auch keine Narrativisierung der dargestellten Welt. Ein Text (Alltagserzählung, Gedicht, Roman, Drama, Geschichtsschreibung, ...) kann eine *narrative Struktur* (Kap. 5; Lotman 1972, Renner 1983, Titzmann 2003) aufweisen: Er erzählt dann eine Geschichte bzw. eine solche ist von ihm abstrahierbar. Kein Bild – wohl aber eine Bildfolge – kann eine narrative Struktur haben (Titzmann 1990, Wünsch 1999). Das Einzelbild kann eine momentane Situation aus einer narrativen Abfolge darstellen, nicht aber diese selbst. Jede solche Situation impliziert natürlich ein Vorher und ein Nachher: Aber falls dieses aus ihr erschlossen werden kann, dann nur mit Hilfe kulturellen Wissens (Kap. 4 und 5); anderenfalls bleiben Vorgeschichte und Folgen dieser Situation Nullposition. Nehmen wir an, ein Gemälde zeige einen fast nackten Mann, der einen Kranz mit Stacheln auf dem Haupte trägt und an Händen und Füßen an ein aufrecht stehendes Holzkreuz genagelt ist (vgl. das Gemälde von Grünewald in Kap. 5). Nur aufgrund unseres kW über christliche Mythologie können wir diese schockierende Situation als Teil einer bekannten Geschichte identifizieren und wissen, um wen es sich handelt und was vorher und nachher geschehen sein soll; ein Mitglied einer Kultur/Gruppe ohne dieses kW stünde vor einer absolut brutalen und rätselhaften Situationsdarstellung.

2.9 Noch eine letzte strukturbedingte Differenz von Bildern und Texten sei hier genannt. Bei sprachlichen Äußerungen können wir zwischen ‚*eigentlichem*‘ und ‚*uneigentlichem*‘ Sprachgebrauch (die Tropen der Rhetorik) unterscheiden. Ein *Tropus* (Metapher, Synekdoche, Metonymie, Allegorie, ...) basiert darauf, dass (mindestens) ein sprachliches Zeichen im Kotext oder Kontext ab-

weichend verwendet wird, so dass dessen lexikalisierbares ‚eigentliches‘ Signifikat nicht vom Text gemeint sein kann, aber aufgrund des Ko(n)textes ein durch diesen bedingtes, neues, ‚uneigentliches‘ Signifikat erschlossen werden kann. Wenn es z.B. von unserem Hans – nachdem er sich Lydias bemächtigt hat – dann hieße:

(5) Gegenüber Lydia war Hans ein Schoßhündchen,

ist offenkundig nicht gemeint, dieser habe sich zum Köter transformiert, sondern er zeige Verhaltensweisen, die dieser Art Viechern kulturell zugeschrieben wird. Das Bild kennt keinen ‚uneigentlichen‘ Zeichengebrauch (Titzmann 1990, Wunsch 1999; für – multimediale – Bildfolgen wie Film gilt anderes: Kanzog 2001). Was ein Bild darstellt, ist genau das, was es darstellt. Ein Element der dargestellten Welt kann aufgrund kultureller Kodierungen zum sekundären Zeichen (z.B. eben ‚Anker‘ für ‚Hoffnung‘) werden, aber es ist nicht uneigentlich gebraucht: Sein primäres Signifikat bleibt erhalten (etwa ein dargestellter Anker) und die sekundäre Bedeutung ist nur durch Kodekenntnis bzw. kW erschließbar. Ein Bild kann einerseits aus Elementen bestehen, die sich mit keinem in unserem kW vorgesehenen Objekten der ‚Realität‘ identifizieren lassen; das gilt für alle *nicht-mimetischen*, also nicht eine potentielle ‚Realität‘ nachahmenden, abbildenden Gemälde: z.B. sowohl für eine *fantastische* wie für eine *abstrakte Malerei*. Ein Bild kann andererseits Elemente kombinieren, die sich in unserem kW nicht gleichzeitig in ein und derselben Situation finden: z.B. die Darstellung eines nackten, umschlungenen Paares auf einem Bett, vor dem ein Anker läge – dergleichen hat man normalerweise (= kW) nicht im Schlafzimmer. Aber der Anker bleibt ein Anker und wird nicht – z.B. – zur Metapher oder Allegorie. Wir haben einfach ein Paar mit sonderbarer Einrichtung des Schlafzimmers vor uns, solange wir nicht wissen, dass kulturell das Signifikat ‚Anker‘ sekundärer Signifikant für ein sekundäres Signifikat (z.B. eben ‚Zuverlässigkeit‘, ‚Treue‘) ist. *Bildliche Tropen* gibt es also nur in Ko(n)texten: ermöglicht durch Bildfolgen und/oder durch die Mitwirkung von Elementen der Sprache und des kW.

13.3 Kooperations- und Interaktionsformen von Text und Bild

1. Texte und Bilder können in ein und derselben semiotischen Äußerung koexistieren (Kooperation/Interaktion ‚in praesentia‘) oder von einander unabhängige Äußerungen bzw. Teile solcher darstellen (was Kooperation/Interaktion – ‚in absentia‘ – nicht ausschließt); zur ersten Gruppe gehören natürlich

multimediale Äußerungen wie Theater-/Operninszenierungen, Film, Fernsehen, Videoclip, Werbung usw.

2. Wenn Text und Bild in praesentia koexistieren, ist zunächst ihre hierarchische Relation innerhalb der Äußerung relevant.

2.1 Das Bild kann *in den Text eingebettet* sein ($B \subset T$). Das ist z.B. der Fall bei ikonischen Illustrationen und Dokumentationen zu Texten in handschriftlich tradierten Büchern des Mittelalters wie in gedruckten der Neuzeit, ob ihr Gegenstand fiktiv oder non-fiktiv, literarisch oder theoretisch ist, ob es sich um Fotos in den Printmedien, um Abbildungen der Spuren von Elementarteilchen in Beschleunigern (etwa in einer physikalischen Publikation), um die unzähligen Formen von Abbildungen in Enzyklopädien oder deren Äquivalent im Internet oder auf CD-Rom handelt. Auch wenn das Bild, etwa in den Nachrichten in Printmedien oder Fernsehen oder bei den Fotos von Beobachtungen im Weltraum mittels eines astronomischen Riesenteleskops, als Ausgangspunkt fungieren mag, als Dokumentation eines politischen oder militärischen oder physikalischen Ereignisses, das vom Text erläutert und kommentiert, interpretiert wird: In diesen Fällen ist das eingebettete Bild entweder dem Text untergeordnet ($B < T$), wenn die Bildpropositionen nur der Illustration und Visualisierung von Textpropositionen dienen, oder das eingebettete Bild ist dem Text übergeordnet ($B > T$), wenn der Text als Kommentar oder Interpretation des Bildes fungiert. Bei totaler oder partieller Inkompatibilität der Text- und der Bildpropositionen gilt im Falle $B < T$, dass das Bild vom Text quasi ‚widerlegt‘ wird, im Falle $B > T$, dass der Text vom Bild ‚widerlegt‘ wird.

Doch nun zu einigen konkreten Bild-Text-Kooperationen. Als Beispiel sei eine der vielen Abbildungen aus dem großen Handbuch der Anatomie des Andreas Vesalius – *De humani corporis fabrica* (1543) – genommen:

V. Bildwissen

1. Sichtbarkeit und Sagbarkeit

Die vorangegangenen Kapitel haben gezeigt, dass die Bildwissenschaft in vielen ihrer genuinen Aufgabenbereiche noch am Anfang steht und es trotz einer großen Reihe von Anregungen etwa aus der Kunstgeschichte, der Soziologie und Ethnologie sowie den *Cultural Studies* weiteren Bedarf an Theoriebildung und Methodenentwicklung gibt. Doch noch immer konzentriert die bildwissenschaftliche Diskussion ihre Energien auf die Unterscheidung von Bild und (Sprach-)Zeichen: „Wo ein Bild war, konnte kein Zeichen sein und umgekehrt.“ (Belting 2005a, 35, ausführlich Belting 2005b) Die Annahme einer wechselseitigen Ausschließlichkeit ist der aktuelle Stand im Kampf, den namhafte Vertreter der Bildwissenschaft gegen die Vorherrschaft des Wortes führen (Boehm 1995; Boehm 2004; Bredekamp 2004; Debray 1996; Mitchell 2008b). Die Fronten sind verhärtet. Vorbehalte von bildwissenschaftlicher Seite gelten weniger Wort und Sprache als solchen, als vielmehr dem Anspruch der Semiotik auf das Bild als Zeichen und auf die Re- oder Dekonstruktion des Bildes als sprachförmiger Zwischenträger von Bedeutungen. Um dagegen Eigenwert und Eigensinn des Bildes zu betonen, wird von den Kritikern nicht nur dessen Sprach(ähn)lichkeit zu Recht bestritten, sondern eine Beschreibungskompetenz der Semiotik für alle wesentlichen Aspekte der Bilder gelehnt. Dazu beruft man sich unter anderem auf die Autorität Foucaults, der in seiner viel zitierten Auseinandersetzung mit Diego Velázquez' Gemälde *Las Meninas* (1656) auch darauf hinweist, dass „die Beziehung der Sprache zur Malerei [...] eine unendliche Beziehung [ist]; [...] Sprache und Malerei verhalten sich zueinander irreduzibel: vergeblich spricht man das aus, was man sieht: das was man sieht, liegt nie in dem, was man sagt [...]!“ (Foucault 1971, 38) Foucault gibt hier klar zu erkennen, dass im besonderen Denken und Wissen einer Epoche weder Sprache noch Bild überflüssig oder austauschbar sind, sondern Aufgaben erfüllen, die ausschließlich sie aufgrund ihrer jeweiligen Form auch erfüllen können. Aber: Ist damit gemeint, dass diese Aufgaben wissenschaftlich nicht beschreibbar sind, dass besonders die Formen, Verfahrensweisen und Funktionen von Bildwerken in der jeweiligen visuellen Kultur nicht analysiert und interpretiert werden können, weil die Bildwissenschaft selbst sich überwiegend sprachlich ausdrückt?

Anti-Semiotik

Der Gegensatz zwischen Bildern und Worten wird zumeist auf eine unüberbrückbare ontologische Differenz zurückgeführt. Die Frontlinie verläuft dabei zwischen der Linguistik und Semiotik, die das Konzept des Zeichens auf das Bild anwenden wollen, und einer Spielart der Bildwissenschaft, die sich phänomenologisch der Präsenz und Wirkung des Bildkörpers und nicht seiner Funktion als Stellvertreter von Abwesendem (Repräsentation) zuwendet. Die Kritik an der Semiotik richtet sich mehrheitlich gegen deren Annahme einer Codiertheit und in deren Folge einer Decodierbarkeit (Lesbarkeit)

„ikonische Differenz“