

Aus: Ellen Fricke (2008). Grundlagen einer multimodalen Grammatik des Deutschen. Syntaktische Strukturen und Funktionen. Habilitation, Europa-Universität Viadrina Frankfurt (Oder), Manuskript.

2 Multimedialität und Multimodalität: Gesten und Lautsprache

2.1 Problemdarstellung

In diesem Kapitel geht es darum, den Begriff der sprachlichen Multimodalität in den Kontext eines leistungsfähigen, semiotisch und kommunikationstheoretisch fundierten Medienbegriffs zu stellen. Jede Artikulation setzt notwendig ein Medium voraus. Medien sind jedoch gegenüber dem Mediatisierten nicht neutral, sondern im Mediatisierten selbst zeigen sich „Spuren“ des zugrunde liegenden Mediums (Krämer 1998: 78ff., Stetter 2005: 79). Daher, so Christian Stetter (2005: 266) in seinem Buch „System und Performanz“, gehört die Klärung des Medienbegriffs zu den „unverzichtbaren Grundlagen einer erkenntnistheoretisch reflektierten Linguistik“.¹ Was ist unter Multimedialität und Multimodalität in der Sprache zu verstehen? Wie lassen sich diese Begriffe im Hinblick auf die von uns in diesem Buch verfolgten Untersuchungsziele sinnvoll definieren? Was unterscheidet Multimodalität von Multimedialität? Inwieweit ist die Lautsprache als multimodal zu betrachten?

In einem ersten Schritt werden wir zunächst gängige Medienbegriffe betrachten und systematisieren, die dann als Grundlage einer Definition sprachlicher Multimodalität dienen. Der Grundgedanke ist folgender: Der Begriff der sprachlichen Multimodalität ist nicht primär an das Kriterium des Vorliegens verschiedener Sinnesmodalitäten gebunden, sondern sprachliche Multimodalität liegt erst dann vor, wenn verschiedene Modalitäten dieselben sprachlichen Funktionen erfüllen kön-

1 Stetter (2005: 265f.): „Artikulation setzt immer ein mediales Substrat voraus. So legitim der Gedanke Saussures auch gewesen ist, die Sprachwissenschaft aus ihrer Abhängigkeit von der Philologie zu lösen, und so wesentlich dafür die Wendung hin zur oralen Sprache war: mit der Evolution der chinesischen Schrift und des Alphabets – um nur die beiden Extrema zu nehmen – tritt die Schrift als mehr oder weniger autonomes Medium mit eigenen Charakteristika neben die orale Sprache. Und neben diesen beiden Sprachmedien behauptet sich die Gebärdensprache als von beiden unabhängiges drittes. Medien sind aber gegenüber dem Mediatisierten nicht neutral. In dessen Gestalt bewahrt sich stets die Spur des Mediums.“

nen, wie z.B. die lautsprachliche und gestische Realisierung der syntaktischen Funktion des Attributs in einer Nominalgruppe (siehe Kapitel 5), oder in dieselben sprachlichen Strukturen eintreten, etwa die Substitution einer lautsprachlichen Konstituente durch eine gestische Konstituente in derselben syntaktischen Position.²

Das heißt Multimodalität setzt im Gegensatz zur Multimedialität nicht nur die Simultanität mindestens zweier Medien als notwendige Bedingung voraus, sondern zusätzlich als hinreichende Bedingung deren strukturelle und/oder funktionale Integration. Diese Integration liegt, das ist eine zentrale These dieses Buches, nicht nur auf der Ebene der Performanz, sondern auch auf der Ebene des Sprachsystems. Mit anderen Worten: Wir behaupten die systemlinguistische Relevanz redebegleitender Gesten. Damit stellt sich die Frage nach dem Verhältnis von System und Rede bzw. von System und Kompetenz. Christian Stetter (2005) schlägt eine Linguistik vor, die das System nicht hinter, sondern in der Performanz sucht und den Begriff der Medialität integriert. Seine Thesen zur Medialitätsproblematik und zur Typ-Token-Problematik werden in einem Unterkapitel referiert und diskutiert. Stetters These der Unhintergebarkeit der Medialität der Sprache wird in einem sich direkt anschließenden Kapitel ergänzt um die weiterführende These der Unhintergebarkeit der Multimodalität der Sprache.

2.2 Ist Sprache multimedial oder multimodal?

Eine begriffliche Klärung

„Grundlagen einer multimodalen Grammatik des Deutschen: Syntaktische Strukturen und Funktionen“ – so lautet der Titel dieses Buches. Nicht nur der Begriff der Grammatik, den wir verwenden, sondern auch der Begriff der sprachlichen Multimodalität bedarf der Klärung. Er ist unmittelbar an den Begriff des Mediums gebunden. Ist Sprache ein Medium? Ist Gestik ein Medium? Ist Fernsehen ein Medium? Ist die Bildzeitung ein Medium? Ist ein Museum ein Medium? Der Gebrauch des Terminus „Medium“ erweist sich als heterogen und inflationär. Ebenso unklar ist die Abgrenzung von Multimedialität und Multimodalität. Handelt es sich um zwei verschiedene Termini für denselben Begriff? Und, sollte es

² Siehe die Beispiele (1) und (2) im ersten Kapitel.

sich um zwei verschiedene Begriffe handeln, sind Sprache und Gestik dann zwei Medien oder ein Medium mit unterschiedlichen Modalitäten?

Konsultiert man neuere Publikationen zum Thema sprachliche Multimodalität (z.B. Stivers/Sidnell 2005, Schmitt 2007), dann bleibt die erhoffte Aufklärung aus. Zwar wird der Grundgedanke deutlich, nämlich dass die menschliche Rede in der Kommunikation von Angesicht zu Angesicht nicht nur hörbar, sondern auch sichtbar ist, insofern Gesten, Mimik, Blickverhalten und Körperhaltung zur Botschaft, die kommuniziert werden soll, in irgendeiner Weise beitragen. Wie dies genau geschieht, wird jedoch nicht expliziert. Allein mit dem Kriterium, dass beim Rezipienten mindestens zwei verschiedene Sinnesmodalitäten an der Rezeption einer Botschaft beteiligt sein müssen, ist jedoch ein „multimodales“ Medium wie die Rede, nicht von einer „multimedialen“ Diashow zu unterscheiden. So heißt es in der Einleitung des Themenheftes „Multimodal Interaction“ der Zeitschrift *Semiotica*:

“Face-to-face interaction is, by definition, multimodal interaction in which participants encounter a steady stream of meaningful facial expressions, gestures, body postures, head movements, words, grammatical constructions, and prosodic contours. In this chapter we follow Enfield (2005) in distinguishing between the vocal/aural and visual-spatial modalities. The vocal/aural modality encompasses spoken language including prosody. The visuospatial modality including gesture, gaze, and body postures. As Enfield notes, these differ only in terms of modality but also with respect to which semiotic ground plays a dominant role in their organization. Vocal-aural signs are prototypically symbolic whereas indexicality and iconicity are more important in the visuospatial modality.” (Stivers/Sidnell 2005: 2)

An diesem Zitat wird zweierlei deutlich, dass erstens der in diesem Themenheft verwendete Begriff der Multimodalität medien- und kommunikationstheoretisch nicht verankert ist und sich zweitens beschränkt auf Phänomene der Performanz. Darüber hinaus ist die Zuordnung der Oralität zur Symbolizität einerseits und der Visualität zur Ikonizität und Indexikalität andererseits ebenso trivial wie falsch. Zumindest für die Indexikalität lässt sich diese Behauptung nicht aufrechterhalten, wie die Deixistheorie seit Bühler (1934/1982) zeigt.³

3 Bühler (1934/1982) nimmt für lautsprachliche Zeichen sowohl ein Symbolfeld als auch ein Zeigfeld an.

Um in einem ersten Schritt die gängigen Medienbegriffe zu systematisieren, bieten die Unterscheidungen, die Roland Posner (1986) in seinem grundlegenden Aufsatz „Zur Systematik der Beschreibung verbaler und nonverbaler Kommunikation“ vornimmt, einen geeigneten Ausgangspunkt. Nach Posner (1986: 293) „ist ein Medium jeweils ein System von Mitteln für die Produktion, Distribution und Rezeption von Zeichen, das den in ihm erzeugten Zeichenprozessen bestimmte gleichbleibende Beschränkungen auferlegt“. Über die jeweiligen medienspezifischen Beschränkungen wird das Mediatisierte durch das Medium geprägt. In Abhängigkeit vom zugrunde gelegten Medientyp unterscheidet Posner einen biologischen, einen physikalischen, einen technologischen, einen soziologischen, einen kulturbezogenen und einen kodebezogenen Medienbegriff:

1. biologischer Medienbegriff (sinnliches Medium): Der biologische Medienbegriff „charakterisiert die Zeichensysteme nach den Körperorganen, die an der Produktion, Distribution, und Rezeption von Zeichen beteiligt sind“ (Posner 1986: 293). Unter den biologischen Medienbegriff nach Posner fällt also das Kriterium der beim Rezipieren eines Zeichens vorliegenden Sinnesmodalität. Danach lassen sich visuelle Medien (Augen), auditive Medien (Ohren), olfaktorische Medien (Nase), gustatorische Medien (Geschmacksknospen im Mund) und taktile Medien (Tastsinn der Haut) unterscheiden (vgl. Posner 1986: 294).
2. physikalischer Medienbegriff (physisches Medium): Der physikalische Medienbegriff „charakterisiert die Zeichensysteme, nach den chemischen Elementen und deren physikalischen Zuständen, soweit sie erforderlich sind, um eine physische Verbindung zwischen dem Produktionsorgan des Senders und dem Rezeptionsorgan des Empfängers herzustellen“ (Posner 1986: 294). Das grundlegende Kriterium ist hier die physische Kontaktmaterie wie z.B. die Luft beim Übertragen der Schallwellen oder ein elektromagnetisches Feld beim Übertragen optischer Wellen (ebd.).
3. Technologischer Medienbegriff (technisches Medium): Der technologische Medienbegriff „charakterisiert die Zeichensysteme nach den technischen Mitteln, die bei der Erzeugung von Zeichenprozessen zur Modifikation der Kontaktmaterie eingesetzt werden, die die physische Verbindung zwischen dem Produktionsorgan des Senders und dem Rezeptionsorgan des Empfängers herstellt“ (Posner 1986: 294). Es geht hier um die verwendeten technischen Apparaturen und ihre Produkte, z.B. Printmedien mit ihren Druckmaschinen und

Druckerzeugnissen oder das Medium CD mit einem CD-Player als Abspielgerät und der CD als Tonträger.

4. Soziologischer Medienbegriff (soziales Medium): Der soziologische Medienbegriff „charakterisiert die Zeichensysteme nach den sozialen Institutionen, die die biologischen, physischen und technischen Mittel mit dem Ziel der Erzeugung von Zeichenprozessen organisieren“ (Posner 1986: 295). Beispiele für solche Institutionen sind z.B. Fernsehanstalten, Museen, Theater, Bibliotheken oder Konzertagenturen (ebd.).
5. Kulturbezogener Medienbegriff (Textsorte): Der kulturbezogene Medienbegriff „charakterisiert die Zeichensysteme nach dem Zweck der Botschaften, die durch sie übermittelt werden“ (Posner 1986: 296). Es geht dabei um Gattungen, Genres oder Textsorten der übermittelten Botschaft. Unabhängig von den jeweils vorliegenden biologischen, physischen, technischen oder sozialen Medien verleihe der Kommunikationszweck, so Posner, den Botschaften gleichartige Strukturen. So wird sowohl im Radio, Fernsehen und in Printmedien zwischen Nachrichten, Kommentaren, Kritiken und Reportagen unterschieden (ebd.).
6. Kodebezogener Medienbegriff (Kodierungsmedium): Der kodebezogene Medienbegriff „charakterisiert die Zeichensysteme nach den Regeln, mithilfe derer ihre Benutzer bei der Zeichenproduktion den Botschaften Zeichenträger und bei der Rezeption den Zeichenträgern Botschaften zuordnen“ (Posner 1986: 296). Beispiele für kodebezogenen Medien wären die Unterteilung innerhalb eines Verlags in eine deutsche, englische oder französische Sektion oder in der Musik die Gegenüberstellung von tonaler und atonaler Musik (ebd.).

Nimmt man Posners Systematisierung zum Ausgangspunkt, um sprachliche Multimedialität von sprachlicher Multimodalität abzugrenzen, dann ist zunächst sowohl für die Multimedialität als auch die Multimodalität das Vorliegen mindestens zweier Medien eine jeweils notwendige, wenn auch nicht hinreichende Bedingung. Damit sind Multimodalität von Monomodalität und Multimedialität von Monomedialität unterschieden. Was ist nun aber das Abgrenzungskriterium zwischen Multimedialität und Multimodalität? Eine Möglichkeit wäre sicherlich, beide Termini als synonym zu betrachten. Damit jedoch würden ein durch Gesten und Mimik unterstützter Vortrag und ein computerunterstützter Vortrag, etwa eine PowerPoint-Präsentation, unter denselben Begriff fallen. Dies wäre jedoch kontraintuitiv, eine Differenzierungsmöglichkeit wäre „verschenkt“. Posner schlägt vor,

den Begriff der Multimedialität an die Beteiligung „apparativer Medien“ zu binden (Posner 1986: 300):

„Nimmt man den heutigen Sprachgebrauch zum Maßstab, so ist nonverbale Kommunikation somit nur dann als multimedial zu bezeichnen, wenn sie von apparativen Medien ermöglicht wird, die mindestens zwei Sinnesmodalitäten einbeziehen.“ (Posner 1986: 300)

Multimedial wären demnach das Fernsehen, das Kino oder eine PowerPoint-Präsentation mit integrierten Tonaufnahmen, nicht aber Theater- und Opernaufführungen (Posner 1986: 399). Allerdings konstatiert Posner im Weiteren:

„Zu fragen bleibt, ob dieser Sprachgebrauch sich konsolidieren wird oder ob mit der Zeit die Faszination der elektronischen Technik so stark nachlassen wird, daß Oper und Theater, Bälle und Sportveranstaltungen auch ohne ihren Einsatz als multimediale Massenkommunikation angesehen und analysiert werden. Das würde nicht nur die Definition des Terminus ‚multimedial‘ vereinfachen, sondern auch die Theoriebildung im Bereich der multimedialen Kommunikation erleichtern. Vieles spricht für eine solche Entwicklung.“ (Posner 1986: 300)

Wie lassen sich die bisher referierten Systematisierungen und Unterscheidungen für einen Begriff sprachlicher Multimodalität nutzbar machen? Gehen wir von einem kodebezogenen Medienbegriff aus, dann ist die Sprache allgemein und damit auch jede Einzelsprache wie das Deutsche ein Medium. Nach Merten (1999: 134) ist Sprache das „erste kommunikative Medium“ überhaupt, aus deren Eigenschaften sich die grundlegenden Kriterien für alle weiteren Medien (im Sinne eines technologischen Medienbegriffs, E.F.) gewinnen lassen, nämlich „Quantelung“ in kleinste syntaktische Einheiten (hohes Auflösungsvermögen), „Nichtverbrauch“, „Relationalität“ (Verweis auf anderes als sich selbst), „Distanzwahrnehmung“, „Fungibilität der prozessierten Inhalte“, „Kopplung“ von psychischen Systemen und „Multiplikatorfunktion“, insofern Sprache von mehreren Empfängern zugleich rezipiert werden kann (Merten 1999: 134ff.). Damit sich jedoch Sprache als erstes kodebezogenes Medium entwickeln kann, muss als Bedingung die hochauflösende Leistung des physikalischen Mediums vorhanden sein (Merten 1999: 141), wie z.B. der akustische Kanal bei den Lautsprachen oder der optische Kanal bei Gebärdensprachen. Physikalische Medien setzen Dinge zueinander in Beziehung und erzeugen dadurch unsere Wahrnehmung (ebd.), es sind physikalische Medien für die Wahrnehmung (Merten 1999: 144). Betrachten wir die folgende Grafik:

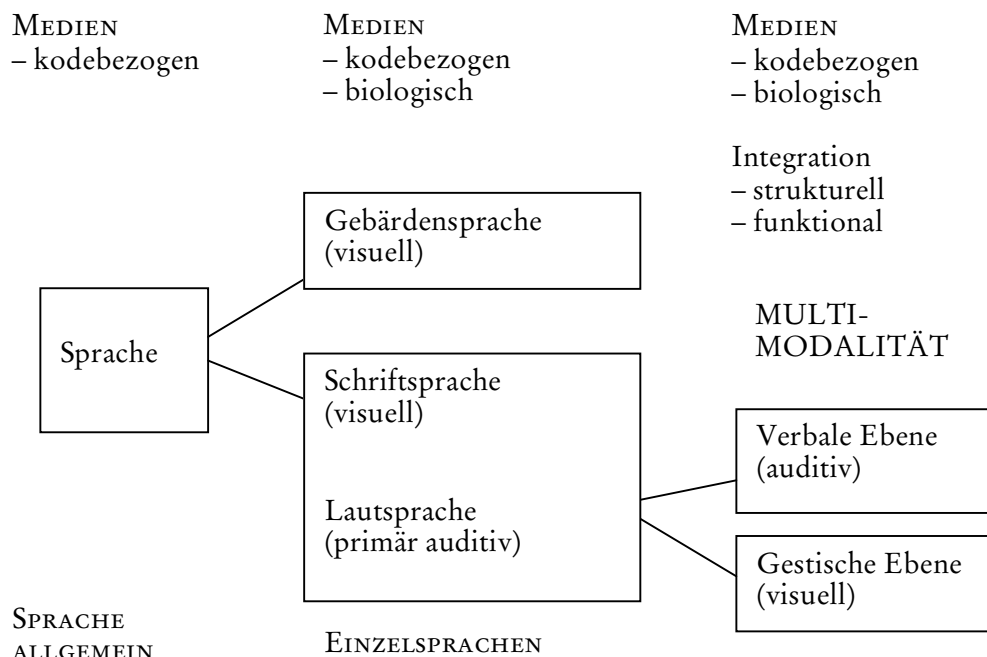


Abbildung 22: Sprachliche Medien und sprachliche Multimodalität

Verschränkt man den kodebezogenen Medienbegriff in einem zweiten Schritt mit dem biologischen Medienbegriff, dann lassen sich verschiedene Medien des Sprachlichen unterscheiden: Gebärdensprache, Schriftsprache und Lautsprache. Eine Gebärdensprache wie die Deutsche Gebärdensprache (DGS) ist, da nur eine Sinnesmodalität affiziert ist, nämlich die visuelle Sinnesmodalität, monomedial und monomodal. Das Deutsche als Schriftsprache ist, sieht man von Hypertextanwendungen im Internet ab, primär visuell und damit ebenfalls primär monomedial und monomodal. In Schrift integriertes Ton- und Bildmaterial setzt die Schriftsprache als Kode voraus, nicht aber umgekehrt. Schriftliche Kommunikation funktioniert unabhängig von anderen potentiell integrierbaren Kodes. Gleiches scheint für die Lautsprache zu gelten: Unabhängig von der visuellen Modalität können sich Menschen, wenn sie z.B. miteinander telefonieren, beschränkt auf die auditive Modalität verständigen. In solchen Kommunikationssituationen ist die Lautsprache ebenfalls monomedial und monomodal. Sichtbare redegleitende Gesten setzen die hörbare, lautsprachliche Rede voraus, nicht aber umgekehrt. Ist die Lautsprache damit analog zur Schriftsprache ebenfalls als primär monomedial und monomodal einzustufen? Es gibt starke Argumente dafür, dies nicht zu tun. Vergleicht man die Kommunikation via Telefon mit der Kommunikation von Angesicht zu Angesicht, dann zeigt sich, dass es sich bei der Kommunikation von

Angesicht zu Angesicht um die ontogenetisch und phylogenetisch primäre Kommunikationsform handelt. Das Telefon ist eine relativ späte technologische Neuerung, der Umgang mit dem Telefon wird von Kindern später erlernt als die unmittelbare Face-to-Face-Kommunikation. Die Kommunikation von Angesicht zu Angesicht nun aber im Sinne eines biologischen Medienbegriffs primär audiovisuell und nach Lyons (1977 und 1980) die sog. „kanonische Äußerungssituation“, von der sich alle anderen Kommunikationssituationen mit ihren kontextuellen Einschränkungen ableiten lassen.⁴

Auch wenn der Ausdruck „Lautsprache“ suggeriert, dass nur die auditive Sinnesmodalität affiziert ist, ist in der Kommunikation von Angesicht zu Angesicht die visuelle Sinnesmodalität via Gesten und anderer Körperbewegungen beteiligt. Es liegt also die Beteiligung zweier Medien (biologischer Medienbegriff) vor, die die notwendige Bedingung sowohl für Multimodalität als auch Multimedialität darstellt. Ist die Lautsprache nun als multimedial oder multimodal zu klassifizieren? Wie lassen sich beide Begriffe im Hinblick auf unsere Untersuchungsziele sinnvoll voneinander abgrenzen? Betrachtet man Charakterisierungen redebegleitender Gesten (z.B. nach McNeill 1992, 2005), dann handelt es sich um Körperbewegungen, die zu beobachten sind, wenn jemand spricht. Sie sind mit der begleiteten Rede eng verbunden im Hinblick auf die zeitliche Organisation und teilen mit ihr gemeinsame semantische und pragmatische Funktionen.

Hier deutet sich ein brauchbares Unterscheidungskriterium für die Abgrenzung von Multimodalität und Multimedialität an, nämlich dasjenige der strukturellen und funktionalen Integration in ein- und denselben bzw. in einen dominanten Kode. Sind zwei sprachliche Medien in denselben Kode strukturell und/oder funktional integriert, dann liegt Multimodalität vor. Sind zwei oder mehr Medien

4 Es wäre zu prüfen, ob die Schrift, die nach Stetter evolutionsbedingt autonom neben die Lautsprache tritt, nicht im Pikturalen der redebegleitenden Gesten ihren zunächst an die Lautsprache gebundenen Ursprung hat. Damit wären die redebegleitenden Gesten die primäre Quelle des Stetterschen „Graphismus“. Das zunächst audiovisuelle Sprechen spaltet sich auf in die autonome, primär auditive Lautsprache und eine autonome, primär visuelle Schrift. Dieser Zusammenhang kommt bei Stetter nicht in den Blick, da er die Lautsprache ausschließlich unter ihrem auditiven Aspekt betrachtet und Gesten für nicht typisierbar hält. Darüber hinaus sind auch noch Zusammenhänge wie die folgenden zu berücksichtigen: Ebenso wie die Schrift zurückwirkt auf die Analyse der Lautung, wirken bildliche Codes zurück auf die Analyse redebegleitender Gesten.

weder strukturell noch funktional in denselben integriert, dann handelt es sich um Multimedialität. Diese Abgrenzung liegt den folgenden Definitionen sprachlicher Multimedialität sowie sprachlicher Multimodalität im engeren und weiteren Sinn zugrunde.

2.3 Definitionen

2.3.1 Multimodalität im engeren Sinn

Der Begriff der sprachlichen Multimodalität beruht sowohl auf einem biologischen als auch auf einem kodebezogenen Medienbegriff. Sprachliche Multimodalität liegt dann vor, wenn sprachlichen Typen oder Token zum einen zwei verschiedenen Sinnesmodalitäten und zwei verschiedenen Kodierungsmedien angehören, und zum anderen eine strukturelle und/oder funktionale Integration in ein- und denselben Kode vorliegt. Ein Beispiel für eine funktionale Integration wäre die Attributfunktion redebegleitender Gesten in verbalen Nominalgruppen mit verbalem Nukleus, z.B. in der Äußerung *diese roten Kacheln* (+ Geste, die ein Dreieck abbildet). In diesem Fall wäre die redebegleitende ikonische Geste (als Bestandteil eines subordinierten Kodes) in den dominanten lautsprachlichen Kode funktional integriert. Ein Beispiel für strukturelle Integration sind redeersetzende Gesten, wenn z.B. in der Äußerung *das war o.k.* das Prädikativum *o.k.* durch eine emblematische Geste nahezu gleicher Bedeutung an derselben syntaktischen Position substituiert wird (siehe die Beispiele (1) und (2) in Kapitel 1.1).

2.3.2 Multimodalität im weiteren Sinn

Sprachliche Multimodalität im weiteren Sinn beruht ausschließlich auf einem kodebezogenen Medienbegriff. So lassen sich nicht nur in der Lautsprache, sondern auch in der Schriftsprache strukturelle und funktionale Verschränkungen beobachten (vgl. Stöckl 2004), und zwar zwischen Schrift und Bild. Da nun aber beide, anders als Geste und Lautsprache, derselben Sinnesmodalität angehören, liegen nur im Sinne eines kodebezogenen Medienbegriffs zwei verschiedene Medien vor. Um sprachliche Multimodalität im weiteren Sinn handelt es sich dann, wenn mindestens zwei verschiedene kodebezogene Medien in ein- und denselben Kode strukturell und/oder funktional integriert sind. Ein Beispiel für eine funktionale

Integration wäre die schriftliche Äußerung des Verbaldeiktikons *hier*, welches z.B. in der Schlagzeile eines Boulevardblatts „Hier hauste sie, die Mörderbande!“ einen Raumbereich denotiert (vgl. Fricke 2007), begleitet von einem graphischen Pfeil. Der denotierte Raumbereich könnte z.B. durch ein auf derselben Seite abgebildetes Foto eines Hochhauses instanziiert sein. In Kontiguität mit dem Deiktikon *hier* in der Schlagzeile und dem Foto eines Hochhauses zeigt der graphische Pfeil auf ein bestimmtes Fenster des Hochhauses. In diesem Fall leisten das Verbaldeiktikon *hier* und der graphische Pfeil gemeinsam die Lokalisierung der in der Schlagzeile erwähnten „Mörderbande“. Ein Beispiel für eine strukturelle Integration wäre z.B. die Einbindung von Links und Icons in Hypertexten des Internet (Janzen 2004: 283). Anders als bei Gestik und Rede in der Kommunikation von Angesicht zu Angesicht, in der jedes Sprechen auch immer potentiell Gestikulieren beinhaltet, also prinzipiell jede verbale Äußerung von einem ununterbrochenen Gestenfluss begleitet sein kann, ist die Interaktion von Schrift und Bild punktuell auf bestimmte „Orte“ im Text begrenzt, an denen Bilder eingefügt werden.

2.3.3 Multimedialität

Sprachliche Multimedialität ist dadurch charakterisiert, dass an einer mündlichen, schriftlichen oder gebärdensprachlichen Äußerung mindestens zwei verschiedene Medien beteiligt sind, die weder strukturell noch funktional in einen Kode integriert sind. Stellen wir uns eine Konzertsituation in der zeitgenössischen Musik vor. Eine Performerin schreibt auf einem mit Kontaktmikrophonen verstärkten Overhead-Projektor einen Text, den sie begleitend liest, bzw. sie spricht einen Text, den sie begleitend schreibt. Die Schreibgeräusche werden über Lautsprecher wiedergegeben. Wir haben es in diesem Beispiel mit verschiedenen Medien zu tun: Das Medium der Schrift als Schriftsprache beim Schreiben und sehendem Verstehen und das Medium der Lautsprache beim Sprechen/Lesen und hörendem Verstehen (kodebezogener Medienbegriff), des Weiteren die verschiedenen beteiligten Sinnesmodalitäten (die sicht- und hörbare Schrift, das hörbare Sprechen/Lesen), die verschiedenen beteiligten physischen Medien (Luft, optische Wellen), die technischen Medien/Apparaturen im Sinne eines technischen Medienbegriffs (Overhead-Projektor, Kontaktmikrophon, Verstärker, Lautsprecher), das Medium der Konzertveranstaltung (soziologischer Medienbegriff) sowie das Medium der Performance (kulturbezogener Medienbegriff).

Wenn wir diese Medien auf den jeweils zugehörigen Medienbegriff bezogen betrachten, dann zeigt sich nirgendwo eine funktionale oder strukturelle Integration, wie wir sie für die sprachliche Multimodalität im engeren und weiteren Sinn beobachten können. Vielmehr stehen die Medien, wenn auf der jeweiligen Bezugsebene mehrere vorliegen, parallel zueinander. Zwar dienen sie einem bestimmten ästhetischen und kommunikativen Zweck, dieser wird jedoch nicht realisiert über die strukturelle und funktionale Integration in denselben lautsprachlichen, schriftsprachlichen oder gebärdensprachlichen Kode. In unserem Beispiel laufen Schreiben und begleitendes Lesen bzw. Sprechen und begleitendes Schreiben zeitlich und aufeinander bezogen parallel. Keines der beiden Medien jedoch greift in den jeweils anderen Kode funktional oder strukturell ein. Es handelt sich also um einen multimedialen und nicht um einen multimodalen Sprech-Schreib-Leseakt.

2.3.4 Übersicht

In der folgenden Tabelle sind die einzelnen Begriffe und die sie konstituierenden Merkmale als Übersicht zusammengefasst:

BEGRIFFLICHE MERKMALE	BEGRIFFE		
	Multimodalität im engeren Sinn	Multimodalität im weiteren Sinn	Multimedialität
Mehr als ein Medium	+	+	+
Kodebezogener Medienbegriff	+	+	+
Biologischer Medienbegriff	+	–	+
Technologischer Medienbegriff	–	–	+
Funktionale/strukturelle Integration in denselben Primärkode	+	+	–

Tabelle 1: Die Abgrenzung von Multimedialität und Multimodalität

Allen zu unterscheidenden Begriffen gemeinsam ist, dass sowohl Multimodalität als auch Multimedialität nur dann vorliegt, wenn mehr als ein Medium involviert ist. Dies ist das Abgrenzungskriterium gegenüber Monomodalität bzw. Monome-

dialität. Multimedialität ist von Multimodalität dadurch unterschieden, dass keine strukturelle oder funktionale Integration in ein- und denselben Primärkode vorliegt. Innerhalb der Multimodalität unterscheiden wir zwischen einer Multimodalität im engeren und weiteren Sinn. Multimodalität im weiteren Sinn liegt dann vor, wenn die beteiligten Medien verschiedenen Kodes angehören, um Multimodalität im engeren Sinn handelt es sich dann, wenn die Medien nicht nur unterschiedlichen Kodes angehören, sondern darüber hinaus auch noch verschiedenen Sinnesmodalitäten im Sinne eines biologischen Medienbegriffs.

2.4 System und Performanz:

Stetters These der Unhintergebarkeit der Medialität der Sprache

Wenn wir nun in sprachlichen Äußerungen Phänomene der Multimodalität und Multimedialität beobachten können und sich gerade angesichts der Möglichkeiten, die das technische Medium Internet bereitstellt, neue Performanzstrukturen wie Chat-Kommunikation, E-Mail-Kommunikation oder Hypertext entwickeln, dann stellt sich die Frage nach dem Status des Medialen in der linguistischen Theoriebildung.

In seinem Buch „System und Performanz“ versucht Christian Stetter eine Neubestimmung der Relation von System und Performanz auf der Grundlage der Symboltheorie Nelson Goodmans. Goodman mache, so Stetter (2005: 9), „die Funktion des Darstellens an effektiv vollzogenen Referenzakten fest“, das heißt an der Parole in der Terminologie Saussures oder der Performanz in der Terminologie Chomskys. Folge man Goodman in seiner Konzeption, so Stetter (ebd.), dann werde „die Konstruktion einer Kompetenz unabhängig von der Berücksichtigung der ihr entsprechenden Performanz unhaltbar“. Mit anderen Worten: „Kompetenz ist dann je nur als Korrelat einer Performanz denkbar, und das heißt in erster Instanz: als Kompetenz, die sich medial realisiert (Hervorhebung E.F.)“ (ebd.). Dieser Gedanke scheint zunächst nicht neu und schließt durchaus an gängige Medienkonzeptionen an, wie etwa diejenige von Luhmann, dessen Konzeption in dem folgenden Zitat referiert wird:

„Das mediale Substrat ist insofern immer als *solches* ein deformiertes und formbares Substrat, die Form immer *als solche* eine mediale Form. Die Prägung einer Form etwa die Mitteilung einer sprachlichen Form, ist deshalb nie einfach nur eine Mitteilung als

solche, sondern immer zugleich und vor allem die Mitteilung einer Mitteilbarkeit: die Mitteilung der medialen Seite der Form. Das Medium, in dem die Differenz von medialem Substrat und Form prozessiert wird, teilt sich mithin ‚in sich selbst mit‘.“ (Khurana 2004: 108)

Linguistische und semiotische „Brisanz“ gewinnt Stetters Medialitätsbegriff erst im Zusammenhang mit einer Konzeption der Zeichen, die er zum einen als sukzessive voneinander erzeugte Kopien ohne Original begreift (s.u.) und zum anderen an den Goodmanschen Begriff der Exemplifikation bindet:

„Zeichen sind die Ausdrücke, die das Material linguistischer Untersuchungen bilden, nicht insofern, als sie *repräsentativ* verwendet werden. Zeichen sind sie vielmehr, insofern sie in ihrem aktuellen Gebrauch bestimmte Eigenschaften *exemplifizieren*: Ähnlichkeiten und Unähnlichkeiten der Gestalt, Typen von Kontextreferenzen, Möglichkeiten von Referenzen auf Erfüllungsgebiete.“ (Stetter 2005: 11)

Goodman (1968/1997) charakterisiert den Begriff der Exemplifikation am Beispiel der Stoffprobe eines Schneiders. Eine rote Stoffprobe verfügt selbst über die Eigenschaft, auf die es Bezug nimmt, nämlich die Eigenschaft, rot zu sein. Symbole, die sich selbst denotieren, exemplifizieren sich auch selbst. Nehmen wir das Wort „dreisilbig“, es denotiert und exemplifiziert dreisilbige Wörter.⁵ Den Unterschied zwischen Denotation und Exemplifikation erläutert Goodman an folgendem Beispiel:

„In der Normalsprache ist die Bezugnahme von ‚Mann‘ auf Churchill und von ‚Wort‘ auf ‚Mann‘ eindeutig Denotation; wenn dagegen Churchill ‚Mann‘ symbolisiert und ‚Mann‘ ‚Wort‘ symbolisiert, ist die Bezugnahme eindeutig Exemplifikation.“ (Goodman 1968/1997: 64)

Stetter benutzt den Begriff der Exemplifikation, um die Dichotomie von System und Performanz durch die Beziehung zwischen Typ und Token neu zu bestimmen. Die Fixierung auf das System habe, so Stetter (2005: 34f.), aus der Opposition von Typ und Token das letztere „weggekürzt“. Stetter begreift den Typ nicht als etwas, was *hinter* den einzelnen Token steht, sondern als „Menge sukzessiver voneinander erzeugter Kopien, für deren keine es ein Original gibt“ (Stetter 2005:

5 Goodman's eigenes Beispiel „kurz“, das kurze Wörter exemplifiziert und denotiert, ist weniger eindeutig. (Goodman 1968/1997: 65).

15).⁶ Wie können wir nun aber von verschiedenen Äußerungsvorkommnissen als demselben Wort oder derselben Phrase sprechen? Welche Berechtigung haben wir dafür? Stetter knüpft auch hier an Goodman an, der in seiner Theorie der Notation die Frage stellt, was die Identität von verschiedenen Aufführungen eines Werks ausmacht:

„Zurecht hat Goodman darauf aufmerksam gemacht, daß als hinreichende wie notwendige Bedingung dafür, A' als Replica von A'' anzusehen, keineswegs gefordert ist, A' müsse gleich mit A'' sein. Es genügt, wenn A' A'' so weit ähnlich ist, daß A' eben aufgrund dieser Ähnlichkeit keinem anderen Typ zugerechnet wird als A'' und umgekehrt. Jedes Token ist qua Token die mehr oder weniger ähnliche Kopie eines anderen Token, und dies hat zur Pointe, daß es *keines Originals* bedarf. Ein beliebiges Token tut es an seiner Stelle. Das verändert den kategorialen Status von Typ und Token grundlegend. Der Typ wird zu einer Menge von mehr oder weniger verbindlichen Unterscheidungskriterien, von denen aber für keines kategorisch gilt, daß es für einen bestimmten Zeitraum mit sich selbst identisch bleiben müßte, einfach deswegen, weil dies der Natur empirischer Gegebenheiten widersprechen würde.“ (Stetter 2005: 199)

Das sprachliche Zeichen – hier knüpft Stetter an Saussure an – bildet sich „im intersubjektiven Gebrauch“ (Stetter 2005: 204), es ist weder ein physischer noch ein psychischer Zustand oder Prozess, sondern ein Typ, „dessen Extension eine Klasse von intersubjektiv verwendeten ähnlichen Tokens ist“ (Stetter 2005: 205). Damit nimmt Stetter im Gegensatz zu Chomsky eine nominalistische und keine universalienrealistische Position ein. Eine Grammatik hat zum Gegenstand „das formal Typisierbare“ (Stetter 2005: 225) und beschreibt nach Stetter notwendige, nicht aber hinreichende Bedingungen für die Formulierung von Äußerungen. Stetter wendet sich explizit gegen das „Verdoppelungsproblem“ der Sprache, das sowohl auf Chomsky als auch auf Saussure zurückgeht (Stetter 2005: 211). Saussure charakterisiert die *Langue* als eine soziale Institution, die von den Mitgliedern der Sprachgemeinschaft psychisch internalisiert wird (vgl. Saussure

6 Mit diesem Verhältnis von System und Gebrauch knüpft Stetter, ohne diese explizit zu nennen, an die Junggrammatiker des 19. Jahrhunderts an, die einen Usus als Durchschnitt aller Äußerungsvorkommnisse annehmen. Vergleiche etwa Hermann Paul (1880/1968: 29): „Aus der Vergleichung der einzelnen Sprachorganismen lässt sich ein gewisser Durchschnitt gewinnen, wodurch das eigentlich Normale in der Sprache, der Sprachusus bestimmt wird. Dieser Durchschnitt kann natürlich um so sicherer festgestellt werden, je mehr Individuen und je vollständiger jedes einzelne beobachtet werden kann.“

1967: 23), Chomsky setzt in der generativen Linguistik Grammatik mit dem internem System gleich (Chomsky 1965). Dies führt zu zwei kategorial verschiedenen Extensionen:

„Die generative Linguistik verwendet Symbole im Zusammenhang von Ausdrücken, mit denen die Konstituentenstruktur von effektiv gegebenen Sätzen beschrieben wird, ‚S‘, ‚NP‘, ‚VP‘ usw. Gleichzeitig, *uno actu*, werden aber diese Symbole durchgängig als Namen für Kategorien u.ä. einer ‚internen‘ Grammatik verwendet. Damit gibt man diesen Symbolen, die ja nur durch die Endsymbole der betreffenden Phrasenstrukturgrammatik extensional definiert sind – insofern eben jedem Endsymbol durch eine Lexikonregel ein Menge von Lexemen zugeordnet wird –, eine *zweite* Extension: Das Symbol ‚S‘ soll nun nicht mehr nur den Satz *The man hit the ball* bezeichnen und die ihm zugeordnete Konstituentenstruktur [...], sondern gleichzeitig etwas ganz anderes, ein Verfahren, das das eine wie das andere erzeugt.“ (Stetter 2005: 211)⁷

Man muss also, so Stetter, zwischen der Repräsentation der Anwendung einer Kompetenz – also der „Dokumentation einer *Performanz*“ – und der Repräsentation der Kompetenz selbst unterscheiden (Stetter 2005: 212).

„Die Psycho- bzw. Neurophrase wird entsprechend etwas anderes sein als Phrasen unserer Sprache usw. Die Frage, was dieses Anderssein ausmacht, wird gestrichen, wenn man die Extension linguistischer Symbole in der beschriebenen Weise verdoppelt. Linguistische Symbole müssen, anders gesprochen, durch psychologische oder neurobiologische Modelle interpretiert werden [...].“ (Stetter 2005: 212)

Neuropsychische Systeme als die das Sprechen erzeugende Instanz sind nach Stetter nicht Gegenstand der Linguistik, sondern der Gegenstand der Neurologie und Psychologie und müssen mit den diesen Disziplinen jeweils eigenen Theorien und Methoden beschrieben werden (Stetter 2005: 216). Die Fähigkeit zur Ausbildung von Verfahren, die Stetter „Neuro- bzw. Psycho-Phonem, -Wort, -Phrase“ (Stetter 2005: 212) usw. nennt, ist – das bestreitet Stetter nicht – angeboren. Er bestreitet aber, dass diese Fähigkeit notwendig auf einer angeborenen Universalgrammatik beruht. Er verweist auf die Theorie neuronaler Netzwerke, die den

7 Stetter vergleicht die Gleichsetzung von Grammatik und internem System als Sprachkompetenz mit der Kompetenz eines Musikers: „Jedermann hielte es wohl für absurd, wenn man die Kompetenz, die Yehudi Menuhin befähigte, Beethovens Violinkonzert zu spielen, durch eine Transkription, die literale Kommentierung der Partitur dieses Konzerts, oder durch eine auf

Nachweis gebracht hat, dass das menschliche Gehirn zu analogischem und induktivem Lernen fähig ist. Dies wiederum korrespondiert mit dem von ihm vertretenen Grundgedanken, „daß sprachliche Typen analogisch organisierte Mengen ähnlicher Inskriptionen sind“ (Stetter 2005: 212). Und so kommt Stetter zu dem Schluss, „daß es hinter dem Sprechen nichts gibt und geben kann, das *als Sprache* anzusprechen wäre“ (Stetter 2005: 213):

„Sprache ist per se ein öffentliches, im beschriebenen Sinne distributiv allgemeines Phänomen. Nur wenn die Linguistik dies als ihre Konstitutionsprämisse akzeptiert, bleibt ihr als Sprach-Wissenschaft ein eigenständiger Ort neben Psychologie und Neurologie auf der einen, Logik, Soziologie etc. auf der anderen Seite. [...] Dies macht es erforderlich, daß die Linguistik das von ihr so lange vernachlässigte Gebiet der phänomenalen Manifestation sprachlicher Einheiten jedweder Komplexität – die vermeintliche Oberfläche – wieder als ihren genuinen Objektbereich begreift und eine diesem angemessene Theorie wie Empirie entwickelt.“ (Stetter 2005: 213f.)

Was heißt nun aber „die langue in der Performanz suchen“ (Stetter 2005: 217)? Es stellt sich die Frage nach dem kategorialen Apparat der von Stetter angemahnten Empirie. Einerseits, so Stetter, sei eine Organisation von einem einheitlichen Gesichtspunkt erforderlich, andererseits müsse dieser Apparat so anpassungsfähig sein, dass damit das Spezifische eines jeden Sprachsystems erfasst werden könne. Stetter knüpft im Weiteren an die doppelte Artikulation und an die Konstituenz als diejenigen Gesichtspunkte der Linguistik des 20. Jahrhunderts an, die sich nach seiner Auffassung unmittelbar zu psychologischen und neurobiologischen Spracherzeugungsprozessen in Beziehung setzen lassen (Stetter 2005: 219). Dauerhaftigkeit und Flüchtigkeit sprachlicher Tatsachen oder die Organisation der linearen Ordnung sprachlicher Einheiten in Gebärdensprachen, in der Schriftsprache und in der Lautsprache sind medial bedingt. Diese die sprachlichen Tatsachen konstituierenden Eigenschaften sind nach Stetter in einem eigenen Teilgebiet der linguistischen Theoriebildung kategorial zu fassen, der Begriff der Sprachkompetenz, so seine These, lasse sich medienunabhängig nicht formulieren. So diagnostiziert er der gegenwärtigen Linguistik eine fehlende Phänomenologie sowie eine unzureichende semiotische und logische Fundierung, um Prozesse medialer Konstituierung angemessen zu erfassen:

einer CD festgehaltene Aufführung des Konzerts durch Menuhin ‚repräsentieren‘ zu wollen“ (Stetter 2005: 212).

„Bis heute fehlt der Linguistik weitestgehend eine Phänomenologie, geschweige denn eine Axiomatik solcher symbolkonstituierender Modi. Dazu fehlen der Disziplin sowohl der Blick auf die mediale Konstitution ihres Gegenstands wie die semiologischen und vor allem die logischen Mittel.“ (Stetter 2005: 225)

Medien sind für Stetter „*symbolisierende Performanzen*, genauer gesagt das, was an der performance reiner Vollzug ist“ (Stetter 2005: 74):

„So hat man also in erster Näherung unter einem Medium das Sichvollziehen einer Operation über oder in einem materiellen Substrat zu verstehen, über einem Apparat oder auch einem Konglomerat von Dingen, sodaß in diesem Vollzug etwas Wahrnehmbares von bestimmter Gestalt erzeugt wird: eine Sonate, ein Tanz, ein geschriebenes Wort, ein Bild, ein Stierkopf aus Fahrradlenker und Sattel. Dieser Vollzug ist Performanz und ist es doch nur, soweit man ihn gleichsam als reine Mechanik betrachtet, nicht als intentionalen Prozeß, ein Vollzug wie das bewußtlose Schwimmen des Fisches im Wasser [...].“ (Stetter 2005: 70)

Das Medium bildet nach Stetter eine zeitliche Einheit mit dem Mediatisierten und beschränkt sich auf die notwendigen Bedingungen zu dessen Realisierung. Damit fällt Stetters Medienbegriff primär unter einen kodebezogenen Medienbegriff nach Posner. Er unterscheidet drei sprachliche Medien, die orale Sprache, die Schriftsprache und die Gebärdensprache und weist ihnen eine je eigene mediale Verfasstheit zu:

„Die sprachliche Einheit existiert im Oralen im eindimensionalen, in der Schrift im zweidimensionalen, in der Gebärdensprache schließlich nicht im dreidimensionalen, sondern in einem vierdimensionalen Raum, [Stetter meint hier den dreidimensionalen Raum und die Zeit, E.F.] und sie je adäquat zu beschreiben ist demgemäß die schwierige, weil eine eigene Phänomenologie erfordernde Grundaufgabe der Linguistik.“ (Stetter 2005: 212)

Mit dieser Charakterisierung sitzt jedoch Stetter derselben Schriftzentriertheit der Grammatik auf, die er in „Schrift und Sprache“ (1997) so scharfsinnig herausarbeitet.⁸ Nach Stetter (2005: 10) ist Grammatik „nichts anderes als die Systematisierung

8 Stetter (1997: 9): „Die Schrift ist wie die Sprache ein Produkt der Evolution, nicht ‚willkürliche‘ menschliche Erfindung. [...] Sie hat klargemacht, daß von dem Augenblick an, wo sich der Graphismus dem Phonismus vollständig untergeordnet hatte, wo mit anderen Worten die Schrift durch ihre Unterwerfung unter die Sprache den gleichen Grad an Differenzierungsmächtigkeit erlangt hatte, sie in ein neues Verhältnis zu dieser tritt. Zwar bleibt die Sprache

einer Sprachauffassung, die mit der allmählichen Ausbildung der orthographischen Register ausgeprägt wurde“. Verhaltensweisen wie Gesten werden nach diesem Grammatikkonzept gleichsam herausgekürzt, weil bezogen auf Alphabetschriften nur die auditive Modalität der Lautsprache von der Schriftsprache durchformt und „analysiert“ wird. Dieser „Herauskürzung“ unterliegt auch Stetter, wenn er die orale Sprache als eindimensionales Medium charakterisiert.

Zwar konstatiert er die eigenständige visuelle, raum-zeitliche Verfasstheit der Gebärdensprache, die damit ihren eigenen medial bedingten Regeln der Kombinatorik unterliegt, jedoch entgeht ihm der multimodale Aspekt der oralen Sprache, deren mediales typisierbares Substrat er auf das rein Auditive beschränkt. Dies liegt neben der schon genannten Schriftzentriertheit an der These der syntaktischen Dichte von Bildern und Gesten, die Stetter von Goodman übernimmt:

„Derselbe Satz ‚Es lebe der König‘ wird nicht einmal von derselben SchauspielerIn zweimal in genau derselben Weise gespielt. Das Artikulationsschema, das in diesem Fall aus dem Verbund von sprachlicher Artikulation, Mimik und Gestik gebildet wird, ist schon syntaktisch dicht, denn Mimik und Gestik haben keine Syntax.“ (Stetter 2005: 35)

Zwar kritisiert Stetter Chomsky und dessen Auffassung der Performanz als defektem Appendix der Kompetenz, denn erst in der Performanz werde „ja die Verknüpfung der syntaktischen mit der gestischen und mimischen Artikulation geleistet, die jede orale Artikulation *als solche* aufweist [...]“, er negiert jedoch eine potentielle systemlinguistische Relevanz von Gesten und deren Integrierbarkeit in lautsprachliche Struktur- und Funktionstypen und betrachtet sie ausschließlich „der Fülle der Artikulation als Erscheinung“ zugehörig (vgl. Stetter 2005: 34f.). Was ist nun unter syntaktischer Dichte nach Goodman zu verstehen, die Stetter als Kriterium anführt?

„Ein Schema ist syntaktisch dicht, wenn es unendlich viele Charaktere bereitstellt, die so geordnet sind, daß es zwischen jeweils zweien immer ein drittes gibt.“ (Goodman 1968/1997: 133)

Und weiter:

immer das ‚ältere‘ Medium, aber sie wird nun von dem jüngeren mehr und mehr durchformt, je länger ihre Kohabitation dauert.“

„Ein Symbolschema ist analog, wenn es syntaktisch dicht ist; [...]“ (Goodman 1968/1997: 154)

Nach diesem Kriterium können, so Stetter, redebegleitende Gesten keine Einheiten aufweisen, die sich zu höheren Einheiten gemäß eines Regelsystems kombinieren. Das Argument der syntaktischen Dichte wird insbesondere in der Bildwissenschaft gegen die Möglichkeit einer Bildsyntax oder Bildgrammatik angeführt (siehe Kapitel 1). Nun fallen nicht alle Gesten unter einen abstrakten Bildbegriff. Darüber hinaus zeigen die Gebärdensprachen der Gehörlosen, deren Gebärden durchaus bildhafte Komponenten besitzen, dass bildhafte Gesten als sprachliche Äußerungen nicht notwendig syntaktisch dicht sein müssen.

Im vierten Kapitel werde ich exemplarisch zeigen, dass redebegleitende Gesten derart strukturiert sind, dass ihnen rekursive Konstituentenstrukturen zugewiesen werden können. Damit widerspreche ich erstens der Annahme, dass Gesten lediglich ein Phänomen der Performanz als Erscheinung sind und damit einer nicht-typisierbaren „Fülle“ zugehörig, und zweitens der Annahme, dass Gesten syntaktisch dicht im Sinne Goodmans sind. Unter dieser Voraussetzung lässt sich dann im Weiteren zeigen, dass Gesten erstens partiell typisierbar und zweitens in den lautsprachlichen Kode potentiell strukturell und funktional integrierbar sind. Der Nachweis, den Stetter für die Unhintergebarkeit der medialen Verfasstheit der Sprache führt, lässt sich ohne weiteres auf die Annahme der Unhintergebarkeit der Multimodalität der Lautsprache übertragen, die Gegenstand des nächsten Unterkapitels ist. So liegt ein wesentliches Verdienst Stetters für die Ziele des vorliegenden Buches darin, dass seine Konzeption es erlaubt, die Gestenforschung in eine Theorie der Sprache einzubinden, die der medialen Konstituierung von Sprache ein eigenes Teilgebiet einräumt.

2.5 System und Performanz:

Argumente für die Unhintergebarkeit der Multimodalität der Sprache

Nicht nur die Medialität, sondern weiterführend auch die Multimodalität der Sprache ist unhintergebar. Innerhalb der Sprachwissenschaft sind dafür methodologische und empirische Argumente angeführt worden, die, obwohl sie im Weiteren kaum rezipiert wurden und daher in Vergessenheit gerieten, nichts an Aktualität eingebüßt haben. Ich werde exemplarisch vier Autoren herausgreifen, die zu

unterschiedlichen Zeiten und in unterschiedlichen linguistischen Traditionen stehend die These einer zumindest partiellen sprachlichen Multimodalität am pointiertesten vertreten haben. Es handelt sich erstens um Wilhelm Wundt, den Begründer der experimentellen Psychologie, der den Gesten, unter die bei ihm das Spektrum von den redebegleitenden Gesten bis zu den Gebärdensprachen der Gehörlosen fällt, ein eigenes Kapitel in seinem Hauptwerk „Völkerpsychologie“ (1900) widmet, zweitens um den Sprachpsychologen Karl Bühler, der in seiner „Zweifelderlehre“ (1934) in Anlehnung an den Junggrammatiker Karl Brugmann eine Theorie des sprachlichen Zeigens entwickelt, die der Zeigegeste eine prominente Stellung einräumt, drittens um den formalen Linguisten und Semiotiker Louis Hjelmslev, der in seiner „Prolegomena zu einer Sprachtheorie“ (1943) die Sprache als reine Struktur beschreibt und die These vertritt, dass dieselbe Struktur sich in verschiedenen Substanzen wie Lauten und Gesten manifestieren kann, viertens um den Linguisten und Verhaltensforscher Kenneth Pike, der in den sechziger Jahren des letzten Jahrhunderts eine einheitliche Theorie und Methodologie fordert, um die Interaktion von verbalem und nonverbalem Verhalten von einem einheitlichen Standpunkt beschreiben zu können. Von den vier genannten Autoren ist Pike derjenige, der seinen Ansatz im Hinblick auf eine multimodale Sprachbeschreibung am stärksten ausgearbeitet hat. Wir beginnen daher zunächst mit der Darstellung seines Ansatzes und fahren zeitlich rückläufig fort bis zu Wundt, der in seinem Kapitel zur Gebärdensprache vergessene Parallelen zur heutigen Gestenforschung aufweist und insbesondere in seinen syntaktischen Analysen einen Ausgangspunkt für die weiteren Kapitel dieses Buches darstellt.

2.5.1 Pike

In seinem Buch „Language in relation to a unified theory of the structure of human behavior“ (1967) fasst Kenneth Pike sprachliches Verhalten als einen Teilbereich allgemeinen kulturellen Verhaltens auf (Hübler 2001: 219f.). Pikes programmatische Position lässt sich nach Hübler folgendermaßen umreißen:

„Da in der Wirklichkeit verbale und nichtverbale Aktivitäten vermischt und wechselseitig vermittelt auftreten, bedarf es auch einer Theorie, in der verbale und nichtverbale Aktivitäten berücksichtigt und als wechselseitig vermittelbar behandelt werden. Diese *unified theory*, die schon im Buchtitel versprochen wird, macht auch eine einheitlichen Methodologie erforderlich, welche das gewährleistet, was angestrebt ist,

nämlich den sprachlichen und den nicht-sprachlichen Anteil gleichermaßen zu berücksichtigen. An die Vorgabe, daß Sprachverhalten nur ein Spezialfall allgemeinen menschlichen Verhaltens ist, knüpft sich die Erwartung, daß zwischen sprachlichem und nicht-sprachlichem Verhalten Gleichheiten existieren, die es ihrerseits ermöglichen, einen Beschreibungsapparat zu entwickeln, der beides in einem erfaßt und als strukturelles Ganzes behandelt.“ (Hübler 2001: 220)

Woraus ergibt sich für Pike die Notwendigkeit, verbales und nonverbales Verhalten einheitlich zu beschreiben? In seinem Artikel “Towards a theory of the structure of human behavior“ (1956/1972) gibt Pike das Beispiel eines „gesture song“, an dem er die methodologische Unzulänglichkeit der herkömmlichen Linguistik demonstriert, die nur die lautsprachliche Ebene als ihrem Gegenstandsreich zugehörig betrachtet. Es handelt sich um ein Lied, bei dem die Sänger die Wörter mit lexikalischer Bedeutung durch eine Geste ersetzen, ohne dass die zeitliche Struktur des Liedes sich verändert. Es entstehen immer mehr Pausen, die Melodie wird von den Sängern aber weiter mitgedacht, bis zum Schluss lediglich die Funktionswörter als vereinzelte Melodiefragmente übrig bleiben.

“Consider, for example, a party game in which people sing the stanza of a song; on the repetition of that same stanza, one word is deleted and a gesture is put in its place. For example, from the phrase *Under the spreading chestnut tree* the word *spreading* may be deleted, and a gesture — the arms outspread — may be substituted for it. During the third singing of that stanza, a further word is replaced by another gesture: for the word *tree* a gesture with the arms upright may be substituted. This kind of replacement can continue, until practically all the words in the stanza are replaced by gestures.” (Pike 1956/1972: 113)

Welche Schlüsse zieht nun Pike aus diesem Beispiel? Zunächst, so Pike, handele es sich um ein einziges Spiel und nicht um ein aus zwei Spielen zusammengesetztes. Eine adäquate Analyse erfordere daher einen ganzheitlichen Zugang und keine Fraktionierung.

“The integration of the verbal and nonverbal activity is here so complex, however, that an attempt to describe it by the mere summation of an atomistic linguistic approach and an atomistic non-linguistic anthropological approach would utterly fail to reveal the structure of that unit.” (Pike 1956/1972: 114)

Versetzen wir uns in die Situation eines Linguisten, der die grammatischen Strukturen dieses Liedes untersuchen will. Den Text der ersten Strophe kann er in Sätze,

Satzteile und weitere kleinere Einheiten wie Morpheme oder Phoneme gliedern. Aber schon ab der zweiten Strophe machen die Daten für ihn keinen Sinn, weil er kein Handwerkzeug dafür hat, andere Körperbewegungen als diejenigen des Mundes zu analysieren. Daher hat er auch keinen Ansatzpunkt, diejenigen Ersetzungsvorgänge zu beschreiben, bei denen Gesten lexikalische Wörter substituieren und die die Essenz des Spiels ausmachen. In der letzten Strophe schließlich findet er nur noch Funktionswörter wie Artikel und Konjunktionen, zwischen denen sich unterschiedlich lange Pausen befinden. Mit anderen Worten: Er findet keine linguistische Struktur.

Nichtlinguisten wie Anthropologen oder Soziologen stehen vor anderen Problemen. Sie beschreiben vielleicht detailliert die soziale Situation und das gestische Verhalten. Aber auch sie erfassen die Äußerung nicht als Einheit, da ihnen elaborierte linguistische Beschreibungskategorien fehlen. Musiker hingegen können die Melodie und die Pausen notieren und, wenn man an Notationsverfahren der zeitgenössischen Musik denkt, auch die Gesten, nur sind ihre Notationskategorien nicht mit linguistischen Beschreibungskategorien korreliert. Pike fordert daher eine einheitliche Theorie der Struktur menschlichen Verhaltens, die es erlaubt, derartige Integrationen von verbalem und nonverbalem Verhalten zu erfassen.

“It is precisely this substitutability of nonverbal materials for verbal materials — in the same class with them, occurring in the same spot, as structurally integrated in the identical fashion — which allows this party game to proceed as a functioning unit. Only a unified theory of the structure of human behavior, such as we have been attempting to develop, can possibly describe such a unitary instance as this and do it justice.” (Pike 1956/1972: 114)

Pike schlägt ein Analyseverfahren nach „spot-class combinations“ als integrierten Einheiten vor (Pike 1956/1972: 110ff.), das in späteren Veröffentlichungen in sein Konzept des „Tagmems“ mündet⁹. Die Grundidee ist jedoch auch schon in seinen frühen Artikeln entwickelt. Es geht darum, eine Äußerungsposition („spot“ später „slot“), an der eine Substitution erfolgen kann, mit einer Klasse („class“), die diese Äußerungsposition „füllen“ kann („filler“), zu korrelieren.

“By ‘spot’ we mean the place at which substitution can occur, just as above we used the phrases *Shut the door* and *Shut the drawer* to show the substitution of the one word

9 Zu unterschiedlichen Fassungen des Tagmemkonzepts im Laufe seiner Veröffentlichungen siehe Pikes Übersicht in „Text and Tagmeme“ (Pike und Pike 1983: 104ff.).

for the other. By this theory, all behavior is considered to contain significant spots at which behavior occurrences may be found. At each spot a series of alternatives is possible, so that if no such series of alternatives is to be found in any particular place within a continuum of behavior, it is assumed that no emic spot is covered by that particular area.” (Pike 1956/1972: 114)

Den großen Vorteil seines Ansatzes sieht Pike darin, dass das Analyseverfahren auf jeder beliebigen Ebene menschlichen Verhaltens beginnen und nach oben hin zu größeren Einheiten oder nach unten hin zu kleineren Einheiten fortgesetzt werden kann. Minimale Einheiten wie etwa Phoneme werden nicht als Anfangspunkt vorausgesetzt.

“The very great advantage of the spot-class start is that, with it, one is able to bypass the necessity for a *minimum* unit in the early stages of the analyses. At any point within behavior, whether the behavior event under consideration be large or small, the analyst can start looking for substitutable items which elicit, in conjunction with the matrix in which they are embedded, differential responses from members of the community.” (Pike 1956/1972: 111)

Hübler (2001: 223) fasst ein Beispiel Pikes in einer tabellarischen Übersicht zusammen, in dem in einer Sequenz der Begrüßung und Verabschiedung die offenen Positionen durch gestische und verbale Verhaltenseinheiten gefüllt sind.

HANDLUNGS- SEQUENZEN	SLOTS			
	Begrüßung (John)	Nachfrage (John)	Erwiderung (Bill)	Verabschiedung (John)
(A)	<i>Howdy</i>	<i>Come along with me?</i>	<i>No</i>	<i>Bye</i>
(B)	<i>Howdy</i>	<i>Come along with me?</i>	Headshake	<i>Bye</i>
(C)	waving a greeting	crooking a finger	<i>No</i>	Waving goodbye
(D)	waving a greeting	crooking a finger	Headshake	Waving goodbye

Tabelle 2: Slots und ihre möglichen Füllungen nach Pike (vgl. Hübler 2001: 223)

In dieser Tabelle gibt es eine Abfolge von vier pragmatischen Funktionen: Die Interaktionssequenz beginnt mit einer Begrüßung von Person y (Bill) durch Person x (John), die an ihre Begrüßung eine Nachfrage anschließt, Person y (Bill) beantwortet die Nachfrage abschlägig, Person x (John) verabschiedet sich. Diese pragmatischen Funktionen und die damit verbundenen Slots können sowohl verbal als auch gestisch instanziiert werden. Zeilen (A) bis (D) stellen unterschiedliche Filler-Varianten dar. In Zeile (A) ist die gesamte Verhaltenssequenz verbal, in Zeile (D) gestisch realisiert. Zeilen (B) und (C) stellen Mischformen dar. So kann die Begrüßung durch John sowohl verbal durch die Äußerung *Howdy* als auch gestisch durch ein Winken instanziiert sein, die Aufforderung mitzukommen lässt sich nicht nur durch die explizite Nachfrage *Come along with me?*, sondern auch durch ein „Herbeiwinken“ mit gekrümmtem Zeigefinger realisieren, ein verbales *Nein* lässt sich durch ein Kopfschütteln ersetzen, die verbale Verabschiedung *Bye* durch ein Abschiedswinken.

An seinen Beispielen wird deutlich, dass Pikes Fokus vor allem auf konventionalisiertem nonverbalem Verhalten liegt und dessen Instanziierung von primär pragmatischen Funktionen. Dennoch gilt Pikes methodologisches Argument auch für grammatische Funktionen und Kontexte:

„The activity of man constitutes a structural whole, in such a way that it can not be subdivided into neat ‘parts’ or ‘levels’ or ‘compartments’ with language in a behavioral compartment insulated in character, content, and organization from other behavior. Verbal and nonverbal activity is a unified whole, and theory and methodology should be organized or created to treat it as such.“ (Pike 1967: 26)

Eine wichtige Voraussetzung für einen Ansatz, wie Pike ihn entwickelt, ist die Neutralität der wissenschaftlichen Beschreibungssprache im Hinblick auf die Zeichenmaterie. Dieser Aspekt wird explizit von Hjelmslev herausgestellt, dessen Ansatz Gegenstand des folgenden Abschnitts ist.

2.5.2 Hjelmslev

Louis Hjelmslev knüpft in seiner Glossematik an zwei Aussagen Saussures an, erstens an die Aussage, die Sprache sei Form und keine Substanz, zweitens an Saussures Auffassung von Sprache als einem System von Differenzen. Gegenstand der Linguistik ist nach Hjelmslev die Sprache als reine Form, als „eine Struktur *sui*

generis“ (Hjelmslev 1943/1974: 9), die Untersuchung der Substanz wird anderen nicht-linguistischen Disziplinen zugewiesen.

„Die Linguistik muß also ihre Hauptaufgabe darin sehen, eine Ausdruckslehre und eine Inhaltslehre auf einer internfunktionalen Basis zustande zu bringen, und zwar eine Ausdruckslehre ohne lautliche oder phänomenologische Voraussetzungen und eine Inhaltslehre ohne ontologische oder phänomenologische Voraussetzungen.“ Hjelmslev (1943/1974: 78)

Hjelmslevs Ausgangspunkt ist der unanalysierte Text in seinem Verlauf. Jedem Verlauf entspricht dabei ein System, „mit dem der Verlauf sich analysieren und beschreiben läßt mit Hilfe einer begrenzten Anzahl von Prämissen“ (Hjelmslev 1943/1974: 13). In seiner Analyse geht Hjelmslev deduktiv vor: Für die sprachwissenschaftliche Analyse wird statt eines induktiven Vorgehens von der Komponente zur Klasse ein deduktives Verfahren von der Klasse zur Komponente gefordert:

„Wenn dem Sprachwissenschaftler irgend etwas gegeben ist [...], dann ist es der noch unanalysierte *Text* in seiner ungeteilten und absoluten Ganzheit. Die einzig mögliche Art des Vorgehens, ein System zu suchen, das dem Verlauf dieses Textes zugeordnet werden kann, wird eine Analyse sein, in der der Text als eine in Komponenten aufgeteilte Klasse aufgefaßt wird, dann diese Komponenten als wiederum in Komponenten aufgeteilte Klassen und weiter so, bis die Analyse erschöpfend durchgeführt ist.“ (Hjelmslev 1943/1974: 17)

Wie verhalten sich nun in Hjelmslevs Ansatz Theorie und Empirie zueinander? Hjelmslev postuliert die „Arbitrarität“ seiner Theorie, insofern sie unabhängig von jeder Erfahrung ist, und ihre „Angemessenheit“, insofern sie Voraussetzungen einführt, die „die Bedingungen zur Anwendung auf bestimmte Erfahrungsdaten erfüllen“ (Hjelmslev 1943/1974: 18f.). Ein induktives Verfahren, wie sie etwa eine Sprachbeschreibung anwendet, die von einzelnen Lauten als kleinsten Einheiten ausgeht, widerspricht Hjelmslevs Empirieprinzip, das eine Beschreibung fordert, die widerspruchsfrei, erschöpfend und so einfach wie möglich ist (Hjelmslev 1943/1974: 15).

„Die induktiv gewonnenen Klassenbegriffe wie ‚Genitiv‘, ‚Perfekt‘, ‚Konjunktiv‘, ‚Passiv‘ usw. liefern schlagende Beispiele. Keiner von ihnen läßt sich in seinen bisherigen Verwendungen allgemein definieren: Genitiv, Perfekt, Konjunktiv und Passiv sind in der einen Sprache, wie z.B. Latein, etwas ganz anderes als in einer anderen, wie z.B. Griechisch. Entsprechendes gilt ohne jede Ausnahme für die übrigen Begriffe der

klassischen Sprachwissenschaft. Auf diesem Gebiet führt deshalb Induktion nicht von der Fluktuation zur Konstanz, sondern lediglich zum Zufall. Die Induktion steht letztlich in Konflikt mit dem von uns aufgestellten Empirieprinzip: sie kann keine widerspruchsfreie und einfache Beschreibung gewährleisten.“ (Hjelmslev 1943/1974: 16)

Welchen Raum lässt Hjelmslevs Ansatz einer multimodalen Sprachbeschreibung? Dadurch dass er vom unanalysierten Text als Gesamtheit ausgeht, und zwar von Äußerungen, die nicht notwendig lautlich sein müssen, sondern verschiedenen Medien und Modalitäten angehören können, wären z.B. redegleitende Gesten Bestandteil dieser noch unanalysierten Gesamtheit und potentieller Gegenstand des sich weiter deduktiv vollziehenden Analyseverfahrens. Hjelmslev äußert sich dazu sehr explizit:

„Die Vorherrschaft der klassischen Phonetik über einen langen Zeitraum hat im übrigen bewirkt, daß sogar der Begriff ‚natürliche‘ Umgangssprache in der Auffassung der Linguisten eingeschränkt worden ist in einer Weise, die nachweislich unempirisch ist, d.h. unangemessen, weil nicht-erschöpfend. Man nahm an, daß die Ausdruckssubstanz der Umgangssprache ausschließlich aus ‚Lauten‘ bestehen würde. Besonders [E. und K.] Zwirner haben in letzter Zeit darauf aufmerksam gemacht, daß man dabei übersehen hat, daß die Rede von Mimik und Gestik begleitet wird und daß gewisse Redeteile durch sie ersetzt werden können und daß in Wirklichkeit wie [E. und K.] Zwirner sagen, nicht nur die sogenannten Sprechorgane (Hals, Mund, Nase), sondern die ganze quergestreifte Muskulatur beim Gebrauch ‚natürlicher‘ Sprachen mitwirkt.“ (Hjelmslev 1943/1974: 101)

Hjelmslevs Annahme, dass sich ein und dieselbe sprachliche Form in unterschiedlichen Substanzen manifestieren kann, lässt sich am besten am Beispiel seines Zeichenmodells verdeutlichen.

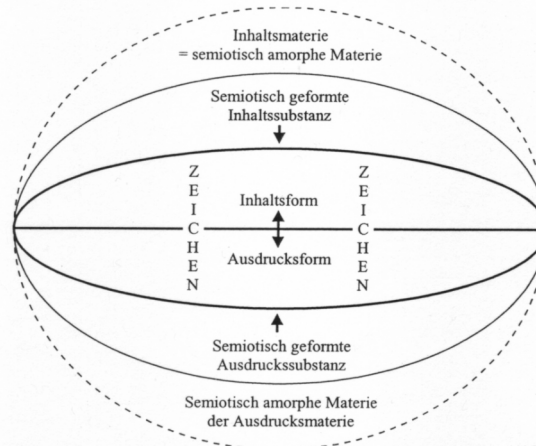


Abbildung 23: Die Zeichenfunktion nach Hjelmslev (vgl. Nöth 2000: 81)

Hjelmslev unterscheidet zwei Ebenen des Zeichens, die Inhaltsebene und die Ausdrucksebene. Beide Ebenen setzen sich wechselseitig voraus, sie sind in Hjelmslevs Terminologie „solidarisch“ und gehen als Funktoren eine Zeichenfunktion ein (Hjelmslev 1943/1974: 52). Anders als Saussure unterscheidet Hjelmslev nicht nur zwischen Substanz und Form, sondern verfeinert diese Unterscheidung jeweils für die Ausdrucksebene und Inhaltsebene. Die Inhaltsebene ist differenziert nach Inhaltsform, Inhaltssubstanz und Inhaltsmaterie, die Ausdrucksebene nach Ausdrucksform, Ausdruckssubstanz und Ausdrucksmaterie. Beide Ebenen sind nicht notwendig isomorph. Die Materie beider Ebenen wird aufgefasst als ungeformte, amorphe Ausdrucks- und Inhaltsmasse, die Substanz ist jeweils geformte Substanz, die Form ist jeweils substanz- bzw. materielose Form.

Jede Sprache gibt nun der Materie eine andere Form, wie auch Sandkörner durch verschiedene Spielzeugformen jeweils anders geformt werden (Hjelmslev 1943/1974: 56). Stellt man sich potentielle Artikulationsorte als ungeformte Lautmaterie vor, dann legen verschiedene Sprachen unterschiedliche Grenzen in das Kontinuum der Artikulationsorte. So ist das Deutsche etwa bei den Plosiven in drei Bereiche eingeteilt, einen vorderen /p/-Bereich, einen mittleren /t/-Bereich und einen hinteren /k/-Bereich, während das Eskimoische und Lettische zusätzlich zwei /k/-Bereiche unterscheiden (Hjelmslev 1943/1974: 58).

Für die Inhaltsebene gibt Hjelmslev das Beispiel der Benennung bestimmter Farben im Deutschen und Kymrischen. Die Inhaltsmaterie ist hier ein Teil des Farbspektrums, das von den Einzelsprachen jeweils unterschiedlich geformt wird, in dem sie voneinander abweichende Grenzen in das Farbspektrum hineinlegen.

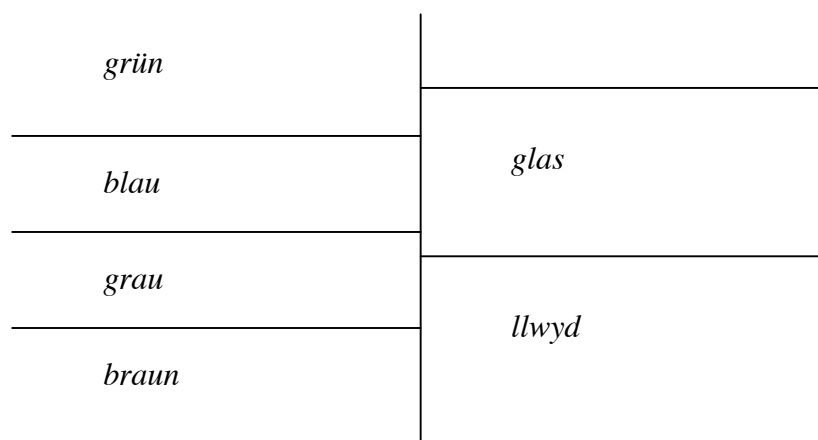


Abbildung 24: Die Formung des Farbspektrums durch verschiedene Sprachen
(Hjelmslev 1943/1974: 57)

Nun verhält es sich nicht nur so, dass dieselbe amorphe Materie durch verschiedene sprachliche Formen zu jeweils verschiedenen Inhalts- und Ausdruckssubstanzen geformt werden kann, sondern dieselbe Form kann sich auch in unterschiedlichen Substanzen manifestieren. Aus Hjelmslevs Annahme, dass die Beschreibung der sprachlichen Substanz die Beschreibung der sprachlichen Form voraussetzt (Hjelmslev 1943/1974: 76) folgt:

„daß ‚die Substanz‘ an sich nicht sprachdefinierend sein kann. Man muß sich, als ein und derselben Form zugeordnet, mehrere Substanzen vorstellen können, die unter dem Gesichtspunkt der Substanzhierarchie wesensverschieden sind; das arbiträre Verhältnis zwischen Sprachform und Sinn macht das zu einer logischen Notwendigkeit.“ (Hjelmslev 1943/1974: 101)

Hjelmslev gibt verschiedene Beispiele für Manifestationen ein- und derselben Form. So kann nach Hjelmslev dieselbe Form in Schrift manifestiert sein, wie dies bei phonetischer oder phonematischen Notationen der Fall ist oder bei an der lautlichen Ebene orientierten Orthographien wie dem Finnischen, Flaggenkodes können zur Manifestation einer natürlichen Sprache verwendet werden, Handalphabete der Gehörlosen können an Stelle der schriftsprachlichen Alphabete treten (Hjelmslev 1943/1974: 102). Die Tatsache, dass eine Sprachform sich in verschiedenen Substanzen manifestieren kann, lässt Hjelmslev zu einem erweiterten Aufgabenbereich der Sprachtheorie gelangen:

„Die Aufgabe des Sprachtheoretikers ist demnach nicht bloß die, das faktisch vorliegende Ausdruckssystem zu beschreiben, sondern zu berechnen, welche Ausdruckssysteme überhaupt als Ausdruck für ein gegebenes Inhaltssystem möglich sind, *und vice versa*. Aber es ist eine Tatsache, die sich leicht experimentell nachweisen lässt, daß ein beliebiges sprachliches Ausdruckssystem in sehr verschiedenen Ausdruckssubstanzen manifestiert werden kann.“ (Hjelmslev 1943/1974: 101)

Damit gehört Hjelmslev zu den ersten Sprachwissenschaftlern, die die Medialität der Sprache systematisch reflektieren. Ihre Untersuchung gehört in Hjelmslevs Konzeption in den Gegenstandsbereich der von der Glossematik abhängigen Substanzwissenschaften. Neben der Phonetik und Phonologie wäre im Sinne von Hjelmslev auch einer potentiellen Kinetik und Kinematik ein Ort zugewiesen.

2.5.3 Bühler

Während bei Pike und Hjelmslev die Beschreibungskategorien auf abstrakten, zeichenmaterieneutralen Relationen wie z.B. Determination beruhen und daher auf unterschiedliche Medien und Modalitäten angewandt werden können, gewinnt Bühler in seiner „Sprachtheorie“ (1934) die Kategorie der Deixis aus der Beobachtung von konkreten Zeigegesten in Relation zu einer Gruppe sprachlicher Ausdrücke wie *ich*, *hier*, *jetzt*, *dieser*, *dort*, die nur bezogen auf die Situation, in der sie geäußert werden, interpretiert werden können.

Im ersten Abschnitt seines Kapitels „Das Zeigfeld der Sprache und die Zeigwörter“ fragt Bühler nach der Gemeinsamkeit von Zeigwörtern wie *hier* und *dort* mit Wegweisern und Zeigegesten. Seine zentrale Behauptung ist, dass „die Bedeutungserfüllung der Zeigwörter an sinnliche Zeighilfen gebunden, auf sie und ihre Äquivalente angewiesen bleibt“ (Bühler 1934/1982: 80).

„Phänomenologisch aber gilt der Satz, daß der Zeigefinger, das natürliche Werkzeug der demonstratio ad oculos zwar ersetzt wird durch andere Zeighilfen; ersetzt schon in der Rede von präsenten Dingen. Doch kann die Hilfe, die er und seine Äquivalente leisten, niemals schlechterdings wegfallen und entbehrt werden; auch nicht in der Anaphora, dem merkwürdigsten und spezifisch sprachlichen Modus des Zeigens. Diese Einsicht ist der Angelpunkt unserer Lehre vom Zeigfeld der Sprache.“ (Bühler 1934/1982: 80f.)

So wie nun die Zeigegesten von einem Ausgangspunkt, der beim Zeigenden liegt, auf einen Zielpunkt verweisen, der beim Zeigobjekt liegt, so verweisen auch Deiktika von einem Ausgangspunkt, der Origo, die beim Sprecher liegt, auf einen Zielpunkt, an dem sich das Zeigobjekt befindet: „Von der Origo des anschaulichen Hier aus werden sprachlich alle anderen Positionen gezeigt, von der Origo Jetzt aus alle anderen Zeitpunkte“ (Bühler 1934/1982: 107). Bühlers Kategorie des Zeigfeldes ist origorelativ und stellt damit ein zur Zeigegeste analog strukturiertes Zeigegestenäquivalent dar. Darüber hinaus betont Bühler, dass in bestimmten Kontexten Zeigegesten für gewisse Ausdrücke eine notwendige Bedingung ihres Gebrauchs darstellen. Die neuere Deixisliteratur spricht dann von „Zeigegestennotwendigkeit“ (Harweg 1990, Sennholz 1985).

Stellen wir uns zur Illustration folgende Situation vor: Zwei Personen sitzen in einem Café an einem Tisch und unterhalten sich. Nach einiger Zeit fragt Person A: *Wo sind denn hier die Toiletten?* Person B antwortet: *Dort!*. Was passiert nun, wenn

die Äußerung von *dort* von keiner Zeigegeste oder einer anderen hinweisenden Bewegung begleitet wird? Wäre Person A in der Lage, die Toilettenräume sofort zu finden? Die Antwort wäre nein, mehr noch, die Äußerung von *dort* in dieser Situation ohne begleitende hinweisende Körperbewegung würde als unangemessen empfunden werden. Das Vorliegen einer Zeigegeste ist in dieser Situation eine notwendige Bedingung für den angemessenen Gebrauch von *dort*: Innerhalb eines von *dort* bezeichneten Raumbereichs referiert sie auf einen Raumpunkt der mit der gesuchten Toilette als vom Sprecher intendierten Referenzobjekt assoziiert ist. Mit anderen Worten: Zeigegestennotwendigkeit ist in verbalen Deiktika oder Zeigwörtern lexikalisch enkodiert.

Wie betrachtet Bühler die Interaktion von Zeigegeste und verbalem Deiktikon? Die Analyse der verbalen Äußerung *dér Hut*, die von einer hinweisenden Zeigegeste begleitet wird, lässt ihn zu dem folgenden Schluss gelangen:

„Die hinweisende *Geste*, welche in der lebendigen Wahrnehmungssituation beobachtet wird, ist unentbehrlich, kann höchstens durch Äquivalente vertreten werden. Wozu dann aber das *dér*, allgemein das demonstrierende Wort der *to-Deixis? Es leistet dem Anschein nach nichts Neues, sondern wiederholt nur, was die Geste auch schon bietet. Doch gerade dies dürfte eine Täuschung sein. Man könnte sagen, das demonstrierende Lautzeichen kopuliere die Fingergeste mit dem Namen *Hut* und mache das Ganze erst zu einem ordentlichen Gefüge. Eine derartige Vermittlerrolle vermöge es deshalb zu spielen, weil es auf der einen Seite dem Material nach mit dem Namen zu den Lautzeichen und auf der anderen Seite der Funktion nach mit der Geste zu den hinweisenden Zeichen gehört.“ (Bühler 1934/1982: 88f.)

Den Zeigwörtern stellt Bühler die Nennwörter gegenüber. Ein Nennwort ist gekennzeichnet durch seine „Wasbestimmtheit“ (Bühler 1934/1982: 103). Es hat die Eigenschaft, „daß es im Munde jedes und aller als Symbol für denselben Gegenstand verwendet wird, sofern er die und die nicht grundsätzlich mit dem Gebrauchsfall wechselnden Eigenschaften hat“ (Bühler 1934/1982: 103). Dies gelte aber nicht, so Bühler, für ein Zeigwort: „Denn ich kann jeder sagen und jeder, der es sagt, weist auf einen anderen Gegenstand hin als jeder andere“ (Bühler 1934/1982: 103).

Zeigwörter gehören dem Zeigfeld, Nennwörter dem Symbolfeld der Sprache an. Zeigfeld und Symbolfeld sind zwei verschiedene Arten von „Umfeldern“ (Bühler 1934/1982: 154), in denen das sprachliche Zeichen seine „Bedeutungserfüllung“ (Bühler 1934/1982: 80) erfährt. Das Zeigfeld ist instanziiert durch die je-

weilige Situation, in der das Zeichen geäußert wird. Das Symbolfeld besteht aus dem durch andere sprachliche Zeichen gebildeten Kontext, „dem syntaktischen und lexikalischen Moment der Sprache“ (Bühler 1934/1982: 151) oder „synsemantischen Umfeld“ (Bühler 1934/1982: 81).

Interessant ist im Weiteren für die Diskussion sprachlicher Multimodalität, dass Bühler ursprünglich in seiner Felderlehre nicht nur ein Symbolfeld und ein Zeigfeld, sondern darüber hinaus auch noch ein „Malfeld“ (Bühler 1924/1982: 195ff.) vorsah. Damit gleicht seine Zeichenkonzeption in diesem Stadium derjenigen von Peirce, die zwischen symbolischen, indexikalischen und ikonischen Zeichen unterscheidet. Was bringt Bühler dazu, schließlich doch von einem Malfeld der Sprache abzusehen, das gleichberechtigt neben dem Zeigfeld und dem Symbolfeld steht?

Bühler unterscheidet zunächst zwischen sprachlicher und nicht-sprachlicher Lautmalerei und bestimmt die sprachliche Lautmalerei als den eigentlichen Gegenstand seiner weiteren Ausführungen:

„Kein Zweifel: wer die Sprache beiseite schiebt, kann lautmalen nach Herzenslust; die Frage ist einzig und allein, ob und wie man es innerhalb der Sprache zu tun vermag.“ (Bühler 1934/1982: 196)

Während Bühler der Zeigegeste im Zeigfeld der Sprache eine prominente Stellung einräumt, zieht er für das Malfeld Gesten überhaupt nicht in Betracht, sondern beschränkt sich ausschließlich auf die lautliche Substanz. Innerhalb der Strukturen der Lautsprache sieht Bühler das Prinzip der Ikonizität verschiedenen Beschränkungen unterworfen, und zwar dem „Syntaxriegel“, dem „Phonemriegel“ und Beschränkungen des Lexikons. Der Riegel, den die Syntax der Lautmalerei vorschreibt, ist vor allem durch die unterschiedliche Freiheit der Wortstellung in den Einzelsprachen bestimmt:

„Wo immer die Sprache als Darstellungsmittel benutzt wird, kann man nur ‚trotzdem‘ malen und soweit es *erstens* die Syntax der Sprache zulässt. Es gibt am Tore zur lautmalenden Sprache einen *Syntax-Riegel*, der leichter oder schwerer zu umgehen ist. Wenn in einer Sprache z. B. die Wortfolge im Satze weitgehend syntaktisch relevant ist wie im modernen Französischen oder Englischen, dann sind dadurch dem Aufbau einheitlicher Lautgemälde von vornherein engere Grenzen gezogen als etwa im Griechischen, Lateinischen oder auch noch im Deutschen.“ (Bühler 1934/1982: 203)

Das Lexikon kann zwar beispielsweise durch onomatopoetische Wortschöpfungen erweitert werden, dennoch basiert die Mehrzahl der Wortbildungen auf nicht-ikonischen Regularitäten:

„Sind die Ansprüche der Satzbildung befriedigt, so bleiben Wortwahl und Wortbildung als kleinere Freiheitsgebiete, worin sich Maltendenzen auswirken können. [...] Allein auch dabei gibt es noch einmal neue Beschränkungen. Der einzelne Sprecher kann das Lexikon nicht beliebig malend erweitern; es wird immer die Frage sein, ob er im Lexikon vorfindet, was er braucht, oder ob das Ganze der Sprache diese und jene Neubildung, die ihm gelegen wäre, gestattet oder nicht.“ (Bühler 1934/1982: 203)

Der Phonemriegel schließlich besteht in dem zugrunde liegenden Phonemsystem und den Regularitäten der phonotaktischen Kombination:

„Die Lautmale am Wort, die Phoneme, müssen mit genügender Schärfe und in der richtigen Reihenfolge realisiert werden; jedes Phonem aber läßt der Realisierung einen Spielraum offen, und in diesen Spielräumen mag die Lautmaterie malend gemodelt werden. Vollkommen frei ist diese Behandlung der Lautmaterie nur in jenen Dimensionen, die in der gegebenen Sprache von vornherein phonologisch unbesetzt, irrelevant sind.“ (Bühler 1934/1982: 203)

Zwar konstatiert Bühler, dass mit der Lautnachahmung ein in der Sprache wirksames Prinzip vorliege (Bühler 1934/1982: 209), jedoch führe dieses Prinzip eher zu marginalen Effekten, so dass die Annahme eines dem Symbol- und Zeigfeld nebengeordneten Malfeldes nicht zu rechtfertigen sei:

„Die Sprache wäre nicht, was sie ist, wenn es ein kohärentes, leistungsfähiges Malfeld in ihr gäbe. Die Sprache ist aber tolerant genug, an bestimmten Grenzen, wo ihre eigengesetzlichen Mittel erschöpft sind, das andersartige Malprinzip zuzulassen.“ (Bühler 1934/1982: 196)

Die Lautmalerei kann sprachlich nur innerhalb derjenigen Grenzen wirksam werden, die durch Zeig- und Symbolfeld gezogen sind:

„Das Malbedürfnis sprießt in den weiten Fugen, die der andersartige, malfremde Aufbau der Sprache frei und unbekümmert offen lässt; es hat sogar, um im Bilde zu bleiben, ein Fleckchen Erde für sich selbst und last not least fungiert es in eigenartiger Weise im Gebiete des sprachlichen A u s d r u c k s, was noch einmal eigens anerkannt werden soll.“ (Bühler 1934/1982: 198)

Trotz des eher marginalen Status des lautmalerischen Prinzips in der Sprache regt Bühler synchrone sprachvergleichende Untersuchungen verschiedener „Maltechniken“ an:

„Systemvergleiche und aus ihnen die Erkenntnis von charakteristischen *Maltechniken* dort und hier sind vorerst das einzige greifbare Forschungsziel, das ich mir als Ersatz der fehlenden Sprachgeschichte vorstellen kann. Von unserer Muttersprache her ergibt sich z. B. sofort die Frage, ob überall ebenso wie bei uns gemischt vokalisch-konsonantisch gemalt wird. Die Vokale unserer Geräuschnamen wie ‚brummen, summen, surren‘ gegen ‚klirren, schwirren, bimmeln, schrill‘ sind bestimmt nicht irrelevant; wie ist es in anderen Sprachen und Sprachfamilien? Gibt es solche, die vorwiegend konsonantisch, und andere, die vorwiegend vokalisch malen, oder wie ist es sonst?“ (Bühler 1934/1982: 215)

Beschränkte man anders als Bühler ein potentiell Malfeld nicht auf die Lautsubstanz, sondern bezöge redegleitende Gesten mit ein, ergäbe sich ein ganz anderes Bild: Aufgrund der andersartigen Substanz entfällt der „Phonemriegel“, darüber hinaus sind Gesten den Regularitäten der lautsprachlichen Morphologie nicht unterworfen, sondern erfahren z.B. als emblematische Gesten eine eigenständige, von lautsprachlichen Wortbildungsprinzipien unabhängige Lexikalisierung. Lediglich der „Syntaxriegel“ bleibt bestehen. Mit anderen Worten: Das Bühlersche Malfeld und damit das Prinzip Ikonizität der Sprache manifestiert sich weniger in der lautlichen als vielmehr in der gestischen Substanz. Verfolgt man diesen Gedanken konsequent weiter, dann ist der Status des Malfeldes möglicherweise zu revidieren. Es ergäbe sich eine Art Arbeitsteilung: Das Symbolfeld manifestiert sich primär lautsprachlich, das Zeigfeld sowohl lautsprachlich als auch gestisch, das Malfeld primär gestisch. Eine Integration von Geste und Lautsprache ergibt sich insofern, als dieselben sprachlichen Prinzipien und Regularitäten, z.B. Deixis und Ikonizität, sich in beiden Modalitäten manifestieren.

2.5.4 Wundt

Wilhelm Wundt ist Begründer der experimentellen Psychologie und des ersten deutschen Instituts zur Gebärdenforschung in Leipzig. Er gehört zu den ersten Forschern überhaupt, die den Gebärdensprachen der Gehörlosen den Status einer eigengesetzlichen natürlichen Sprache zugestehen, die der Lautsprache grundsätzlich gleichwertig ist.

Der erste Band seines Hauptwerks „Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte“ (1900) hat die Sprache zum Gegenstand. Das zweite Kapitel „Die Gebärdensprache“ untersucht Gesten unter linguistischem und semiotischem Aspekt. Wundt vertritt darin die Auffassung, dass Gesten, wie sie beispielsweise in den Gebärdensprachen der Gehörlosen vorliegen, Eigenschaften des angeborenen menschlichen Sprachvermögens reflektieren. Gebärdensprachen, darunter fallen bei Wundt auch emblematische Zeichensysteme und redebegleitende Gesten, haben sich, so Wundt, „aus allgemeingültigen Bedingungen der psychophysischen Organisation des Menschen nach bestimmten, überall wiederkehrenden psychologischen Gesetzen entwickelt [...]“ (Wundt 1900/1911: 253) und bilden ein einfaches und klares Modell des menschlichen Spracherwerbs sowie sprachlicher Konzeptualisierungsprozesse, wie Bedeutungsübertragung und -wandel im Allgemeinen. Er vertritt als einer der Ersten die These, dass es sich bei den Gebärden der Gehörlosen nicht nur um Zeichensysteme wortähnlicher Gesten handelt, sondern um eine eigenständige, von der Lautsprache unabhängige Sprache mit einer entwickelten Syntax:

„Man muß daher, statt aus der indifferenten Beschaffenheit der einzelnen Gebärden auf das Fehlen des Satzes zu schließen, vielmehr aus dem Dasein bestimmter syntaktischer Gesetze folgern, daß auch die Gebärdensprache nicht bloß aus einzelnen Zeichen, sondern aus Sätzen besteht. Ja der Satz spielt in ihr sogar eine größere Rolle, insofern er es ist, der erst dem einzelnen Zeichen seine grammatische Bedeutung verleiht.“ (Wundt 1900/1911: 218)

Wundt (1900/1911: 164) definiert Gesten „als ein System von Ausdrucksbewegungen, dem der Trieb der Mitteilung und Verständigung seine besonderen Eigenschaften verliehen hat“. Aus den jeweils vorliegenden Kommunikationsbedingungen heraus entwickelt er eine Einteilung der Gesten, die „Kendons Kontinuums der Gebärden“ (vgl. McNeill 1992) vorwegnimmt: Er unterscheidet die eine eigenständige Syntax aufweisenden Gebärdensprachen der Gehörlosen als Manifestation einer angeborenen Sprachfähigkeit und der Notwendigkeit, Gebärden als einziges sprachliches Kommunikationsmittel zu gebrauchen, die sprachersetzenden konventionalisierten Gesten der nordamerikanischen Indianer, die als zusätzlich verfügbares, übereinzelsprachliches Kommunikationsmittel zur Lautsprache hinzutreten und die Kommunikation zwischen den verschiedensprachigen Indianerstämmen ermöglichen, die neapolitanischen Gesten als konventionalisierte For-

men des ursprünglich zugrunde liegenden „Gestus“ (Wundt 1900/1911: 155, diejenigen Gesten, die das Sprechen begleiten, sowie der südeuropäischen Tendenz zu einer lebhafteren Pantomime, die nicht nur das Sprechen begleitet, sondern eingesetzt wird, wenn etwas verbal nur schwer kommuniziert werden kann (Wundt 1900/1911: 155). Des Weiteren führt Wundt rituelle Gesten an, die fester Bestandteil einer rituellen Handlung sind, wie z.B. die Geste des Segnens oder des Schwörens (Wundt 1900/1911: 158f.).

Wundt unternimmt eine weitere Klassifikation der Gebärden im Wesentlichen unter zwei Gesichtspunkten: erstens unter einem semiotischen Gesichtspunkt nach der Relation zum durch die Gebärde bezeichneten Objekt und zweitens unter psychologischem Gesichtspunkt nach ihrem psychologischen Ursprung. Die semiotisch angelegte Klassifikation der Gebärden umfasst, wie auch alle späteren Klassifikationssysteme, die hinweisenden Zeigegesten und ikonischen Gesten als große Gestenklassen. Ikonische oder „nachbildende Gesten“ in der Terminologie Wundts sind, so Wundt, darstellende Gesten, die sich aus nachahmenden Bewegungen entwickelt haben. Dabei unterscheidet Wundt zwei Formen der Nachahmung, eine „zeichnende“ und eine „plastische“: „Entweder nämlich werden die Umrißlinien des vorgestellten Gegenstands mit dem bewegten Zeigefinger in die Luft gezeichnet; oder die Gestalt des Gegenstandes wird durch die Hände in einer bleibenden Form nachgebildet“ (Wundt 1900/1911: 170).¹⁰

In seiner psychologischen Klassifikation zielt Wundt auf eine „Etymologie“, „eine Klassifikation der Gebärden, die von genetischen Gesichtspunkten aus unternommen wird [...]“ (Wundt 1900/1911: 162). Er versteht darunter die Rekonstruktion nicht des historischen, sondern des psychologischen Ursprungs einer Geste:

„Muß sich die Etymologie der Lautsprache mit der Ermittlung der Anfangsbildungen begnügen, die sie als geschichtlich gegebene und nicht weiter abzuleitende, eben deshalb aber auch in der Regel als unerklärbare anzusehen hat, so ist das „Etymon“ einer Gebärde dann nachgewiesen, wenn ihre psychologische Bedeutung und ihr Zusammenhang mit den allgemeinen Prinzipien der Ausdrucksbewegungen erkannt ist“ (Wundt 1900/1911: 163).

Wie hundert Jahre später McNeill interessieren Wundt die Gesten als Fenster zur Kognition: „[...] die Art, wie sich in der Zusammenfügung der Gebärden die

¹⁰ Diese beiden Formen der Nachahmung finden sich später in Müllers Darstellungsweisen (Müller 1998). Siehe dazu Kapitel 1.4.3.

Gedankenfolge spiegelt“ (Wundt 1900/1911: 157). Allerdings begründet er diese Spiegelung mit der spezifischen Medialität der Gesten und stellt ihre syntaktischen Prinzipien der Strukturierung in den Zusammenhang mit verschiedenen syntaktischen Strukturtypen der Lautsprachen. Mit anderen Worten: Wundt unterscheidet explizit zwischen der Abfolge der Gedanken auf einer kognitiven Ebene und einer Abfolge lautsprachlicher bzw. gestischer Zeichen auf der sprachlichen „Oberfläche“. Die These, dass es sich bei Gestik und lautsprachlicher Rede auf der kognitiven Ebene um ein integriertes System handle, ist bei Wundt schon implizit angelegt. Wie steht es nun mit der syntaktischen „Oberfläche“?

Wundt weist am Beispiel der Gebärdensprachen der Gehörlosen nach, dass der visuell-gestische Modus prinzipiell sprachfähig ist. Er wendet sich gegen die These Delbrücks, „eine Syntax der Gebärdensprache sei, wo sie überhaupt existiere, von der Lautsprache aus eingedrungen“ (Wundt 1900/1904: 210). Dem Einwand, dass Gesten über keine grammatischen Kategorien und keine Flexionsmerkmale verfügen, begegnet Wundt mit dem Argument, dass die Gebärdensprache damit nicht der Fähigkeit zur Satzbildung beraubt sei. Zur fehlenden bzw. geringen Ausprägung von Eigenschaften, die Flexionsmerkmalen in der Lautsprache äquivalent sind, merkt Wundt (1900/1904: 209) an:

„Noch weniger gilt dies von jenem relativen Mangel grammatischer Unterscheidung, wonach die einzelne Gebärde als solche nicht erkennen läßt, welche Stellung sie in der ganzen Mitteilung einnimmt. Denn eben hier läßt sich die logische Kategorie, der das einzelne Zeichen zuzurechnen ist, aus dem Zusammenhang unzweideutig erkennen. Dabei stellt es sich aber heraus, daß gerade das, was der Gebärdensprache angeblich fehlen soll, die Verbindung der einzelnen Vorstellungen zu einem Satzganzen, für sie das Hilfsmittel ist, durch das der grammatische Wert der einzelnen Gebärden bestimmt wird.“ (Wundt 1900/1904: 209)

Für Wundt ergibt sich aus diesen Überlegungen,

„daß von einer Syntax der Gebärdensprache mit vollem Recht geredet werden kann, insofern eben syntaktische Stellung der Wörter und Satz zusammengehörige Wechselbegriffe sind. Wo ein Satz existiert, da muß es auch bestimmte Gesetze zur Wortfügung geben, und umgekehrt: wo diese nachzuweisen sind, da ist auch der Satz vorhanden. Man muß daher, statt aus der indifferenten Beschaffenheit der einzelnen Gebärden auf das Fehlen des Satzes zu schließen, vielmehr aus dem Dasein bestimmter syntaktischer Gesetze schließen, daß auch die Gebärdensprache nicht bloß aus einzelnen Zeichen, sondern aus Sätzen besteht.“ Wundt (1900/1904: 209f.)

Welches sind nun die syntaktischen Gesetze der Gebärden, die Wundt herausarbeitet? Mit den meisten Lautsprachen stimmt die Gebärdensprache darin überein, daß in der Regel das Subjekt des Satzes zuerst kommt. Anders als im Deutschen, aber wie im Französischen, ist die Position des Attributs nachgestellt, das Objekt hingegen steht wie im Lateinischen vor dem Verb (Wundt 1900/1904: 211). Die Gebärdensprache (deutscher Gehörlose, E.F.), so Wundt (ebd.), stimmt zwar in ihrer syntaktischen Abfolge nicht mit dem lautsprachlichen Deutschen, wohl aber mit anderen Sprachen überein. Wenn nun deutsche Gehörlose in deutscher Lautsprache unterrichtet werden, dann kommt es vor, dass sie zunächst wie beim Erlernen einer Fremdsprache, wie Wundt anhand von Unterrichtsbeispielen darlegt, die Satzgliedfolge der Gebärdensprache auf diejenige der deutschen Lautsprache übertragen. In diesen Phänomenen sieht Wundt einen Beleg für die Sprachlichkeit der Gebärdensprache.

„Alle diese Erscheinungen zeigen, daß sich in dieser Beziehung die Gebärdensprache nicht anders verhält wie jede andere Sprache. Die eingeübte Sprachform ist nicht bloß ein äußeres Gewand des Gedankens, sondern sie beeinflusst diesen selbst, so daß sie sich zunächst jede neu angeeignete Sprachform untertan macht.“ (Wundt 1900/1904: 213f.)

Da nun Wundt in der Zeichensprache der Indianer, die er in einem Kapitel untersucht, eine ähnliche Abfolge findet, kommt er zu dem Schluss, dass die Gebärdensyntax primär auf gemeinsame psychologische Ursachen zurückzuführen sei. Er nimmt drei zugrunde liegende Prinzipien an: die „logische“, die „zeitliche“ und die „räumliche“ Abhängigkeit (Wundt 1900/1904: 217).

„Dabei ist zunächst das Prinzip der zeitlichen Anschaulichkeit für die Aufeinanderfolge der größeren Zusammenhänge, das der räumlichen für die engeren Verbindungen innerhalb eines einzelnen Satzes vorzugsweise bestimmend. Die Gebärdensprache berichtet Ereignisse genau in der Folge, in der sie erlebt wurden. Sie beschreibt Gegenstände genau in der Ordnung, in der sich ihre Teile der Beobachtung aufdrängen. [...] Zu dieser Folge wird sie aber schon deshalb gedrängt, weil die einzelnen Gebärden in ihren wichtigsten Formen selbst Nachbildungen aufeinander folgender Handlungen sind. So überträgt das Prinzip der zeitlichen Anschaulichkeit nur eine Eigenschaft der einzelnen Gebärden auf deren Zusammenhang.“ (Wundt 1900/1904: 20)

In anderer Weise ist das Prinzip der zeitlichen Anschaulichkeit für die Wortstellung verantwortlich. Eigenschaften treten im Allgemeinen gleichzeitig mit einem

Gegenstand auf und gehören zu den Merkmalen, an denen der Gegenstand erkannt wird. Objekte hingegen treten zugleich mit der Handlung auf, von der sie affiziert werden. Wundt zerlegt nun diese Beziehungen in eine „konstantere selbständige“ und in eine „variablere abhängige“ Vorstellung. Er gibt das Beispiel der Wortfolge *ein großes Haus*. In dieser Verbindung ist das Haus die unabhängige, die Größe hingegen die abhängige Vorstellung, da letztere an einen Gegenstand, in diesem Beispiel das Haus, gebunden ist. In der Wortfolge *der Baumeister baut das Haus* kann „die Handlung des Bauens nicht vorgestellt werden, ohne den Gegenstand, der gebaut wird“ (Wundt 1900/1904: 221). Was ergibt sich nach Wundt aus diesen Überlegungen für die Abfolge der Satzglieder und Satzgliedteile in der Gebärdensprache?

„Auf diese Weise sind die beiden Regeln der Stellung des Adjektivs hinter dem Substantiv und des Verbums hinter dem Objekt einerseits einfache Folgen der realen Koexistenz des Gegenstandes und seiner Eigenschaften, der Handlung und ihres Objektes. Andererseits entspringen sie der relativen Langsamkeit der Gebärdenfolge [...]. Dies verhält sich in der Lautsprache, in der ein Substantiv und sein Attribut, ein Verbum und sein Objekt im Fluß der Rede vollständig zu einer Worteinheit verbunden sein können, wesentlich anders. Verbindungen wie *mons ingens* und *ingens mons*, *puerum laudat* und *laudat puerum* sind beide für unser Denken simultanen Verbindungen äquivalent. Bei der Gebärdensprache, wo sich jeder Begriff selbständiger vom andern abhebt, würde eine Gebärde, die erst durch eine folgende ihre Stellung im Satz erhielte, leicht eine unerträgliche Hemmung im Flusse der Vorstellungen erzeugen.“ (Wundt 1900/1904: 221)

Wundt führt die syntaktischen Eigenschaften der Gebärdensprache auf zwei allgemeine Bedingungen zurück: Zum einen auf das ikonische Prinzip, das die Abfolge der Zeichen die Abhängigkeitsverhältnisse und Abfolgen in der Anschauung abbildet, zum anderen auf die mediale Bedingung, dass die gebärdensprachlichen Zeichen relativ langsam aufeinander folgen. Dies bedingt, „daß ein gegebenes Symbol, soweit es nicht an sich selbst deutlich ist, durch vorangehende, nicht erst durch nachfolgende Symbole seine Bedeutung erhält“ (Wundt 1900/1904: 221).

Neuere Untersuchungen zu Gebärdensprachen wie z.B. Boyes Braem (1992) zeigen, dass deren syntaktische und andere grammatische Eigenschaften komplexer sind, als Wundt es seinerzeit noch annimmt. Dennoch sind seine Argumente für eine multimodale Grammatik relevant: Wundt zeigt erstens bezogen auf die sprachliche Substanz, dass Gesten grundsätzlich sprachfähig sind. Damit können

aus Substanzgründen Gesten nicht mehr vom Gegenstand der Sprachwissenschaft ausgeschlossen werden. Da Wundts Konzept der Gebärdensprache das spätere Kendonsche Kontinuum der Gebärden antizipiert, ist nun grundsätzlich auch für Lautsprachen die Möglichkeit eröffnet, dass sie sich sowohl auf der verbalen auch auf der gestischen Ebene manifestieren können. Zweitens klassifiziert er Gesten nach ihrem psychologischen Ursprung und nach ihrem Verhältnis zum Bezeichneten. Damit trennt er die kognitive Ebene von der sprachlichen Ebene. Unsere Unterscheidung zwischen objektbezogenen und interpretantenbezogenen Gesten ist an Wundt anschließbar und im Kern in seinen Unterscheidungen schon angelegt. Drittens ist Wundt meines Wissens der erste Wissenschaftler überhaupt, der innerhalb der Sprache mediale Bedingungen syntaktischer Strukturen und Funktionen reflektiert. Damit ist er der Wegbereiter für eine Linguistik, die die Sprache in Abhängigkeit von ihrer Medialität untersucht.

2.5.5 Fazit

Die Arbeiten von Wundt, Bühler, Hjelmslev und Pike unterstützen aus jeweils unterschiedlicher Perspektive, die in diesem Kapitel vertretene These, dass die Annahme einer grundsätzlich multimodalen Natur der Sprache unhintergebar ist. Was gilt es nun an methodologischen und empirischen Argumenten für eine multimodale Grammatik festzuhalten?

1. Wie die Gebärdensprachen der Gehörlosen zeigen, sind Gesten grundsätzlich sprach- und grammatikfähig. Im Hinblick auf ihre Medialität kann sich Sprache nicht nur in Laut und Schrift, sondern auch in der visuell-räumlichen Substanz der Gebärden manifestieren.
2. Wie die Zeigegestennotwendigkeit in der Deixis der Lautsprache aufzeigt, ist das Vorliegen redegleitender Zeigegesten, eine Bedingung bestimmte verbale Deiktika zu gebrauchen. Zeigegesten sind somit systemlinguistisch relevant und nicht nur ein Phänomen des Sprachgebrauchs. Damit kann sich Sprache nicht nur in der Gestik als Medium wie in der Gebärdensprache der Gehörlosen, sondern auch multimodal in Lautsprache und begleitender Gestik der Hörenden manifestieren. Die Multimodalität der Sprache ist in der linguistischen Beschreibung also unhintergebar.
3. Aufgrund der nachgewiesenen Sprachfähigkeit in Gebärdensprachen und der systemlinguistischen Relevanz in Lautsprachen gehören Gesten nicht nur in

den Gegenstandsbereich der Linguistik, sondern überdies in den Kernbereich der Grammatikschreibung.

4. Daraus folgt, dass linguistische Beschreibungskategorien in den Lautsprachen nicht beschränkt auf die Substanz der Schrift und des Lautes definiert werden können.
5. Da Gesten potentiell in sprachliche Strukturen und Funktionen eintreten können, muss man, wie Pike und Hjelmslev es fordern, von der unanalysierten Äußerung als Ganzheit ausgehen und sie einer zeichenmaterieneutralen Analyse unterziehen, die sie als strukturelle und funktionale Einheit erfasst.

2.6 Zusammenfassung

Gegenstand dieses Kapitels ist zunächst die begriffliche Abgrenzung zwischen sprachlicher Multimedialität und sprachlicher Multimodalität. Multimedialität ist von Multimodalität dadurch unterschieden, dass bei der Multimedialität keine Integration in ein- und denselben Primärkode vorliegt. Anknüpfend an den Medienbegriff von Posner (1986) differenzieren wir innerhalb der Multimodalität zwischen Multimodalität im weiteren und engeren Sinn. Multimodalität im weiteren Sinn liegt dann vor, wenn die beteiligten Medien verschiedenen Kodes angehören (kodebezogener Medienbegriff nach Posner, z.B. Text und Bild). Wenn die beteiligten Medien nicht nur verschiedenen Kodes angehören, sondern darüber hinaus auch noch unterschiedlichen Sinnesmodalitäten (biologischer Medienbegriff nach Posner, z.B. Rede und Gestik) angehören, dann handelt es sich um Multimodalität im engeren Sinn.

Im Weiteren reflektieren wir den Status des Multimedialen und Multimodalen in der linguistischen Theoriebildung. Anknüpfend an die Symboltheorie Nelson Goodmans postuliert Christian Stetter die Unhintergebarkeit der Medialität der Sprache. Der Begriff der Sprachkompetenz, so seine These, lasse sich medienunabhängig nicht formulieren. Dabei unterscheidet er für die Sprache zwischen den drei Medien Lautsprache, Schriftsprache und Gebärdensprache.

Demgegenüber behaupten wir nicht nur die Unhintergebarkeit der Medialität, sondern auch der Multimodalität der Sprache und referieren die Ideen und Positionen von vier Klassikern der linguistischen Theoriebildung, nämlich Wilhelm Wundt, Karl Bühler, Louis Hjelmslev und Kenneth Pike, deren erste Ansätze zu

einer multimodalen Sprachbeschreibung, wenn überhaupt, dann nur eingeschränkt rezipiert wurden. Sie zeigen am Beispiel der Gebärdensprachen der Gehörlosen, dass das Medium der Gestik grundsätzlich sprachfähig ist (Wundt). Sprache kann sich also, in unterschiedlichen Medien manifestieren (Hjelmslev). Am Beispiel der Zeigegesten wird nachgewiesen, dass diese für gewisse lautsprachliche Ausdrücke eine notwendige Bedingung darstellen, diese zu gebrauchen (Bühler). Damit liegt eine systemlinguistische Integration von Gesten in die Lautsprache vor und nach unserer Definition Multimodalität im engeren Sinn. Daraus folgt, dass sprachliche Äußerungen als Ganzheit aufzufassen und einer zeichenmaterieneutralen Analyse zu unterziehen sind, deren Kategorien nicht von der Lautsubstanz her definiert sind (Pike, Hjelmslev).