Ellen Fricke

Grundlagen einer multimodalen Grammatik des Deutschen: Syntaktische Strukturen und Funktionen

Habilitationsschrift

eingereicht an der Fakultät für Kulturwissenschaften der Europa-Universität Viadrina Frankfurt (Oder)

im März 2008

Gefördert von der Volkswagenstiftung im Rahmen des interdisziplinären Forschungsprojekts “Towards a Grammar of Gesture: Evolution, Brain, and Linguistic Structures”.

Zeichnungen: Karin Becker.

„Wenn ich zu irgendetwas mehr Anlage als die allermeisten besitze, so ist es zu einem Verbinden sonst gewöhnlich als getrennt angesehener Dinge, einem Zusammennehmen mehrerer Seiten [...] und dem Entdecken der Einheit in einer Mannigfaltigkeit von Erscheinungen."

Wilhelm von Humboldt (1767–1835)
in einem Brief an F.A. Wolf (Lehmann 1985: 225)

"Verbal and nonverbal activity is a unified whole, and theory and methodology should be organized or created to treat it as such."

Kenneth L. Pike (1912–2000)
(1967: 26)

Inhaltsverzeichnis

1 Einführung ... 1
1.1 Gegenstand, Fragestellung und Ziel ... 1
1.2 Zum Forschungskontext ... 10
1.3 Datenerhebung, Transkription und Analyse ... 18
1.4 Redebegleitende Gesten: eine kurze Übersicht aus semiotischer Perspektive ... 21
1.4.1 Gestische und lautsprachliche Zeichen ... 21
1.4.2 Gestentypen: Ikon, Index, Symbol ... 26
1.4.3 Gestische Darstellungsweisen nach Müller: die Hand agiert, modelliert, zeichnet, verkörpert ... 30
1.5 Das Problem der gestischen Bezugnahme: mittelbare und unmittelbare Referenz bei ikonischen Gesten ... 32
1.6 Zum Aufbau der Arbeit ... 36

2 Multimedialität und Multimodalität: Gesten und Lautsprache ... 39
2.1 Problemdarstellung ... 39
2.2 Ist Sprache multimedial oder multimodal? Eine begriffliche Klärung ... 40
2.3 Definitionen ... 46
2.3.1 Multimodalität im engeren Sinn ... 47
2.3.2 Multimodalität im weiteren Sinn ... 47
2.3.3 Multimedialität ... 48
2.3.4 Übersicht ... 49
2.4 System und Performanz: Stetters These der Unhintergehbarkeit der Medialität der Sprache ... 50
2.5 System und Performanz: Argumente für die Unhintergehbarkeit der Multimodalität der Sprache ... 57
2.5.1 Pike ... 58
2.5.2 Hjelmslev ... 62

2.5.3 Bühler ... 67
2.5.4 Wundt ... 71
2.5.5 Fazit ... 77
2.6 Zusammenfassung ... 78

3 Syntaktische Komponenten: Semantisierung und Typisierbarkeit ... 80
3.1 Problemdarstellung ... 80
3.2 Stetters Begriff der Typisierung ... 82
3.3 Phonaestheme und die Semantisierung submorphematischer Einheiten in der Lautsprache ... 92
3.4 Kinaestheme und die Semantisierung gestischer Formmerkmale ... 108
3.5 Kendons Kontinuum der Gebärden: eine Revision ... 124
3.6 Zusammenfassung ... 130

4 Syntaktische Strukturen: Konstituenz und Rekursivität ... 132
4.1 Problemdarstellung ... 132
4.2 Syntax und Medialität: die mehrdimensionale Syntax der Gesten ... 135
4.3 Lineare Strukturen: Gestenphasen nach Kendon ... 148
4.4 Lineare Strukturen: Konstituentenstruktur und Markierungsstruktur ... 152
4.5 Lineare Strukturen: Gesten und Rekursion ... 163
4.5.1 Selbsteinbettung: Gesture Units ... 166
4.5.2 Koordination: Strokes ... 179
4.6 Kontextfreie Phrasenstrukturgrammatik der Gestenphasen ... 188
4.7 Syntaktische Relationen: Kopf, Kern und Modifikation ... 196
4.8 Zusammenfassung ... 200

5 Syntaktische Funktionen: Multimodale Attribuierung in Nominalgruppen ... 202
5.1 Problemdarstellung ... 202
5.2 Sind Gesten attributfähig? Eine Beispielanalyse ... 205
5.3 Linguistische Attributkonzepte ... 215
5.3.1 Attribuierung und Prädikation ... 216
5.3.2 Attribuierung und Determination ... 222

5.4 Die strukturelle Integration von Gesten in Nominalgruppen durch *so*, *solch* und *son* ... 228
5.5 Extensionale und intensionale Determination: objekt- und interpretantenbezogene Gesten ... 244
5.6 Zusammenfassung ... 269

6 Schluss ... 272
6.1 Zusammenfassung ... 272
6.2 Implikationen für eine multimodale Grammatik ... 275
6.3 Forschungsausblick ... 278

7 Anhang ... 280
7.1 Abkürzungen und Notationskonventionen ... 280
7.2 Der Parcours am Potsdamer Platz ... 284
7.3 Abbildungsverzeichnis ... 289
7.4 Tabellenverzeichnis ... 294

Literatur ... 297

1 Einführung

1.1 Gegenstand, Fragestellung und Ziel

Das Ziel dieses Buches besteht darin, erste theoretische, methodische und empirische Grundlagen für ein längerfristiges Projekt zu legen, nämlich für eine multimodale Grammatik – zunächst des Deutschen –, die redebegleitende Gesten partiell integriert. Unter Grammatik verstehen wir hier eine einzelsprachliche deskriptive Korpusgrammatik, „die auf eine umfassende Beschreibung der beobachtbaren Regularitäten einer Sprache bzw. eines repräsentativen Ausschnitts abzielt" (Bußmann 1990: 288).

Betrachtet man neuere wissenschaftliche Grammatiken des Gegenwartsdeutschen (Duden 1984 und 2006, Eisenberg 1998 und 1999, Engel 1996 und 2004, Heidolph et al. 1981, Helbig/Buscha 1998, Hentschel/Weydt 1994 und 2003, Zifonun et al. 1997), dann stellt man Unterschiede im Umfang der behandelten Gegenstände fest: Es gibt einen Kernbereich, zu dem die Syntax und Flexionsmorphologie gehören, und eine Peripherie mit Wortbildungslehre, Lautlehre, Orthographie, Diskurs- und Textgrammatik, die nicht von allen Grammatiken behandelt werden. Will man für eine zumindest partiell multimodal verfasste Grammatik argumentieren, dann ist es strategisch günstig, an dem unzweifelhaften Kernbereich jeder Grammatik anzusetzen: der Satzlehre. Weist man in diesem Bereich multimodale Strukturen und Funktionen nach, dann ist jedes Argument, dass diese nicht oder nur peripher Gegenstand der Grammatik seien, obsolet. Wir orientieren uns dabei an den folgenden Leitfragen:

1. Sind Gesten, die das Sprechen begleiten, unabhängig von der Lautsprache – hier des Deutschen – typisierbar und semantisierbar?
2. Lassen sich diesen Gesten unabhängig von der Lautsprache syntaktische Konstituentenstrukturen zuweisen?
3. Inwieweit können Gesten z.B. über syntaktische Funktionen in die lautsprachliche Syntax integriert werden?

Der Nachweis der Typisierbarkeit von Gesten ist die Voraussetzung für die Annahme syntaktischer Konstituenten, die in syntaktische Konstituentenstrukturen

eingehen. Der Nachweis ihrer Semantisierbarkeit ist die Voraussetzung für die Zuweisung der syntaktischen Relation der Modifikation bei multimodalen Attribuierungen in lautsprachlichen Nominalgruppen. Darin können Gesten die syntaktische Funktion eines Attributs übernehmen, das die Extension des durch das Kernsubstantiv Bezeichneten einschränkt. Sowohl Typisierung als auch Semantisierung in Gestik und Lautsprache lassen sich nach denselben strukturellen Prinzipien beschreiben, die auf dem Begriff der relativen Motiviertheit nach Saussure bzw. auf dem Konzept des diagrammatischen Ikonismus nach Peirce aufbauen. Mit anderen Worten: Dieselben strukturellen Prinzipien wirken in Gestik und Lautsprache und leisten eine strukturelle Integration auf der Ebene des Sprachsystems.

Unter Struktur verstehe ich Relationen zwischen Einheiten oder Mengen von Einheiten, die nach bestimmten gemeinsamen Eigenschaften klassifiziert sind:

> „Eine größere Einheit, etwa eine Wortform hat Struktur oder ist strukturiert, wenn sie aus solchen kleineren Einheiten nach kombinatorischen Regularitäten aufgebaut ist.“ (Eisenberg 1998: 2)

Im vierten Kapitel zeige ich, dass Gestenphasen unabhängig von der Lautsprache rekursive Konstituentenstrukturen zugewiesen werden können, die über die Eigenschaft der Selbsteinbettung verfügen und die sich prinzipiell zu Strukturen beliebiger Komplexität kombinieren. Damit ist die syntaktische Eigenschaft der Rekursivität nicht spezifisch für die Lautsprache, sondern ist auch für redebegleitende Gesten anzunehmen.

Wenn man nun dieselben Strukturen und Funktionen in Gestik und Lautsprache findet, dann lässt eine Umkehrung der Betrachtungsweise den Schluss zu, dass dieselben sprachlichen Strukturen und Funktionen sich zugleich in Gestik und Lautsprache manifestieren. Die Aufgabe einer multimodalen Grammatik ist dabei zu untersuchen, wo – das heißt z.B. auf welchen sprachlichen Beschreibungsebenen – die Integrationsbereiche von Gestik und Lautsprache liegen und wie diese beschaffen sind.

Gestische Konstituenten kombinieren sich jedoch nicht nur auf der gestischen Ebene zu höheren Einheiten, sondern können auch in lautsprachliche Konstituentenstrukturen integriert werden und dort syntaktische Funktionen übernehmen. Funktionen wie Subjekt, Prädikat, Objekt und Attribut sind nach Eisenberg (1999: 38) „relationale Begriffe“:

„Sie kennzeichnen eine Konstituente nicht für sich selbst und unabhängig von der Umgebung, sondern sie kennzeichnen, welche Funktion die Konstituente innerhalb einer größeren Einheit hat. Sie wird damit in Beziehung zu anderen Konstituenten gesetzt und diese Beziehungen oder Relationen werden als Subjekt-Beziehung, Objekt-Beziehung usw. bezeichnet." (Eisenberg 1999: 38)

Betrachten wir beispielsweise die Äußerung

(1) *Das war okay.*

Diese Äußerung kann von einer Geste der folgenden Form begleitet sein:

Abbildung 1: Die o.k.-Geste mit aufgerichtetem Daumen (Austin 1806/1966: Bildtafel 6)

Die Daumengeste kann aber auch das verbale *okay* ersetzen und im Satz dessen Funktion als Prädikativum übernehmen:

(2) *Das war* (Daumengeste).

Da die Geste an derselben Äußerungsposititon wie *okay* auftritt, ist die Geste nicht nur funktional, sondern auch linear als Konstituente in den Äußerungsverlauf integriert. Wie Richter und Wegner (1977: 213) betonen, „steht außer Zweifel, daß verbale und nichtverbale Verhaltenselemente füreinander eintreten können". Jedoch wird diese funktionale Äquivalenz nicht auf die Darstellungsfunktion[1] nach Bühler (1934/1982: 28) bezogen, die Gegenstand des vorliegenden Buches ist, sondern auf die Bühlersche Funktion der Kundgabe (Bühler 1934/1982: 28):

1 In seinem Organonmodell der Sprache unterscheidet Bühler zwischen den Funktionen der „Darstellung", der „Kundgabe" (Ausdruck) und der „Auslösung" (Appell) und ordnet diesen Funktionen die drei Zeichenarten „Symbol", „Symptom" und „Signal" zu: „Es [das sprachliche Zeichen, E.F.] ist Symbol kraft seiner Zuordnung zu Gegenständen und Sachverhalten, Symptom (Anzeichen, Indicium) kraft seiner Abhängigkeit vom Sender, dessen Innerlichkeit es ausdrückt, und Signal kraft seines Appells an den Hörer, dessen äußeres oder inneres Verhalten es steuert wie andere Verkehrszeichen." (Bühler 1934/1982: 28)

> „Wenn es uns auch so erscheint, daß z.B. verunsicherte und ängstliche Individuen die Wichtigkeit von nichtverbalen Verhaltensweisen gegenüber verbalen beträchtlich überschätzen, so scheinen auch unter „normalen" Interaktionspartnern die nichtverbalen Verhaltensweisen bevorzugt als Quelle der Inferenz über den Interaktionspartner benutzt zu werden. Hiermit wird eine funktionale Differenzierung des Kommunikationsprozesses in eine Funktion der Beziehung und eine des Inhalts unumgänglich. Und wir können uns Argyle (1972) anschließen, der sagt, daß in einem kommunikativen Interaktionsprozeß die Sprache allgemein dazu dient, Tatsachen, Meinungen, und Probleme zu diskutieren, und nichtverbale Verhaltensweisen dazu verwendet werden, um Emotionen, Bedürfnisse und interpersonelle Einstellungen auszudrücken." (Richter und Wegner 1977: 216)

Unsere Frage hingegen lautet: Inwieweit ist Sprache allgemein und die deutsche Sprache im Besonderen im Hinblick auf ihre Darstellungsfunktion multimodal verfasst?[2] Um diese Frage zu beantworten, ist es notwendig, den Begriff der Multimodalität zu definieren und ihn von anderen Begriffen wie den der Multimedialität abzugrenzen. Im zweiten Kapitel schlage ich einen Begriff der Multimodalität im engeren und weiteren Sinn vor. Multimodalität im engeren Sinn – der Gegenstand dieses Buches – ist dann gegeben, so der Grundgedanke, wenn nicht nur zwei verschiedene Sinnesmodalitäten oder Kodierungsmedien vorliegen, sondern wenn darüber hinaus zugleich eine strukturelle und/oder funktionale Integration in ein- und denselben Kode geleistet wird.

Zieht man redebegleitende Gesten als diejenigen Körperbewegungen, die man beobachten kann, wenn jemand spricht, in Betracht, dann ist die je spezifische Medialität von lautsprachlicher und gestischer Äußerung zu beachten. Lautsprachliche Äußerungen sind eindimensional im zeitlichen Verlauf gegeben, gestische Äußerungen hingegen sind vierdimensional, da sie räumlich (dreidimensional) und zeitlich sind (vgl. Stetter 2005: 225). Was folgt daraus für die linguistische Beschreibung von Formen und Bedeutungen? Für eine multimodale Syntax wäre zu prüfen, ob das zeitliche Nacheinander die einzige syntaktisch relevante Dimension darstellt. Im dritten Kapitel schlage ich eine „dreidimensionale" Syntaktik[3] der Gesten

---

2 Müller (1998: 104ff.) gehört zu den ersten in der Gestenforschung, die explizit an Bühler anknüpft und den Primat der Darstellungsfunktion für die Untersuchung redebegleitender Gesten herausstellt.

3 Vgl. Morris (1938/1972: 32ff.) sowie Posner und Robering (1997: 14ff.).

vor, die nicht nur das zeitliche Nacheinander, sondern auch simultan auftretende semantisierte Formparameter wie Handform, Handorientierung, Bewegungsform und räumliche Situierung sowie simultan ausgeführte Bewegungen unterschiedlicher Körperteile wie linke Hand, rechte Hand, Kopf, linker Fuß, rechter Fuß usw. für eine syntaktische Kombinatorik berücksichtigt. Die grundlegende Beobachtung ist, dass die syntaktische Relation der Modifikation (mod) (Eisenberg 1998 und 1999) nicht nur zeitlich linear wie in der Lautsprache, sondern für redebegleitende Gesten auch zeitlich simultan erfolgt. Mit anderen Worten: Multimodale Syntagmatik ist nicht mit einem zeitlichen Nacheinander gleichzusetzen.

In Nominalgruppen lautsprachlicher Äußerungen modifiziert beispielsweise ein adjektivisches Attribut das Kernsubstantiv. Betrachten wir die folgende Äußerung, die ich zusammen mit meiner Kollegin Jana Bressem am 28. Juli 2006 in Frankfurt (Oder) beobachtet habe:

(3) A: *Jetzt hole ich mir dieses gelbe Urlaubsformular.* (+ Geste, bei der beide Hände ein Rechteck zeichnen, dessen kurze Seiten vertikal nach oben zeigen.)

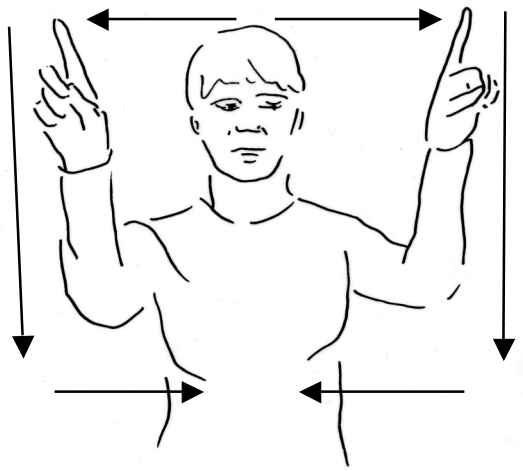

Abbildung 2: Redebegleitende Geste zur Nominalgruppe *dieses gelbe Urlaubsformular*

Das Urlaubsformular, von dem hier die Rede ist, ist ein hellgelber, dünner Karton in DIN-A4-Größe, der im Querformat beschriftet ist. Welche Informationen können wir dieser Äußerung entnehmen? Wenn wir uns auf die Nominalgruppe *dieses gelbe Urlaubsformular* beschränken, dann liegt auf der verbalen Ebene lediglich eine attributive Erweiterung des Kernsubstantivs *Urlaubsformular* durch das Farbadjektiv *gelb* vor. Ziehen wir die gestische Ebene hinzu, dann erfahren wir

nicht nur etwas über die Farbqualität des besagten Formulars, sondern auch etwas über dessen Formqualität, nämlich dass es rechteckig ist. Außerdem erfahren wir, dass es sich bei dem Formular um ein Querformat handelt. Diese Informationen sind keinesfalls redundant und nicht in der verbalen Bezugsgröße *Urlaubsformular* enthalten. Ein nur verbal genanntes Formular könnte beispielsweise auch im Hochformat sein. Andere Formen als die rechteckige Form sind für Verwaltungsformulare eher untypisch, obwohl auch das prinzipiell möglich wäre. Das Farbadjektiv *gelb* fungiert innerhalb der Nominalgruppe *dieses gelbe Urlaubsformular* als Attribut zu Urlaubsformular. Lässt sich für die begleitende Geste eine ähnliche Funktion annehmen? Betrachtet man Attributdefinitionen in Grammatiken des Deutschen, wie z.B. in Peter Eisenbergs „Grundriß der deutschen Grammatik" (1999: 231), dann scheint die ein Rechteck nachahmende Geste in unserem Beispiel unter dessen Definition zu fallen: „Die primäre Leistung der Attribute besteht darin, das von einem Substantiv Bezeichnete ‚näher zu bestimmen'". Damit gehören redebegleitende Gesten zum Gegenstandsbereich der Syntax und Satzsemantik und nicht nur, wie bisher angenommen, zum Gegenstandsbereich der Pragmatik und der linguistischen Gesprächsforschung.[4]

Betrachtet man Beispiel (3) näher, dann fällt außerdem auf, dass die Zuschreibung der Farbeigenschaft verbal, die Zuschreibung der Formeigenschaft hingegen gestisch erfolgt, eine „Arbeitsteilung", die sich, wie wir im fünften Kapitel sehen werden, häufig beobachten lässt. Denn Gesten sind aufgrund ihrer spezifischen Medialität besonders geeignet für die Nachahmung von Handlungen, die Darstellung von Formen und Größenangaben sowie für Lokalisationen im Raum.

Dass die Medialität der Sprache unhintergehbar ist, wird von Christian Stetter in seinem Buch „System und Performanz" (2005) herausgestellt. Anders als Stetter behaupte ich nicht nur die Unhintergehbarkeit der sprachlichen Medialität, sondern darüber hinaus auch die Unhintergehbarkeit der sprachlichen Multimodalität (siehe Kapitel 2). In welcher Relation stehen System und Performanz in multimodalen Äußerungen? Stetter entwirft in seinem Buch eine Linguistik, die das System nicht hinter, sondern in der Äußerung sucht (Stetter 2005). Er bezieht sich dabei auf die von Nelson Goodman vorgeschlagene Beziehung zwischen Typ und Token als „Menge sukzessiver voneinander erzeugter Kopien, für deren keine es ein

4 Zum wissenschaftsgeschichtlichen Hintergrund dieser Annahme siehe Fricke (2007: 4ff.).

Original gibt“ (Stetter 2005: 15). Folgt man diesem Ansatz, dann ist ein System ohne eine medial realisierte Parole, um in Saussures Termini zu sprechen, nicht denkbar.

Im eigentlichen Sinne fruchtbar für die Zielsetzungen dieser Arbeit wird die Adaption des Goodmanschen Typ-Token-Konzepts, wenn es um die Darstellung von Konventionalisierungs- und Semantisierungsprozessen unterhalb der Ebene stabiler Form-Bedeutungsbeziehungen wie sie bei so genannten emblematischen Gesten wie der o.k.-Geste vorliegen. Das Konzept des „Phonästhems“ (Zelinsky-Wibbelt 1983) als submorphematischer, semantisch belasteter lautlicher Einheit lässt sich auf der Grundlage des Stetterschen Systembegriffs für Gesten adaptieren. Damit ist es möglich, zwischen den rein „idiosynkratischen Gesten“ (McNeill 1992), die den äußersten linken Pol mit dem geringsten Konventionalisierungsgrad in Kendons Kontinuum bilden, und den morphemanalogen, lexikalisierten emblematischen Gesten in der Mitte des Kontinuums weitere Zwischenstufen nicht nur anzunehmen, sondern auch präzise zu beschreiben. Die Grundlage für die Beschreibung derartiger Semantisierungsprozesse sind an Eisenberg (1998 und 1999) angelehnte Markierungsstrukturen mit den gestischen Formparametern und ihren Instanziierungen. Letztere gehen als terminale Konstituenten in gestische Konstituentenstrukturen ein.

Mit dem Anliegen einer multimodalen Grammatik, in diesem Buch fokussiert auf die Syntax des Deutschen, betritt dieses Buch Neuland, insofern

- es zum ersten Mal für redebegleitende Gesten (anknüpfend an Kendon 1972) die Möglichkeit syntaktischer Konstituenz im Hinblick auf ihre Medialität systematisch auslotet,
- es intersubjektive Semantisierungsprozesse für redebegleitende Gesten nicht nur annimmt, sondern zum ersten Mal über die Adaption des linguistischen Phonästhem-Konzepts konsistent darstellbar macht,
- es zum ersten Mal nachweist, dass Gesten rekursive Konstituentenstrukturen zugewiesen werden können, die über die Eigenschaft der Selbsteinbettung verfügen,
- es zum ersten Mal am Beispiel der Attribuierung systematisch zeigt, wie Gesten strukturell und funktional in lautsprachliche syntaktische Strukturen integriert werden können,

- es zum ersten Mal systemlinguistische und performanzlinguistische Integrierbarkeit in einen systematischen Zusammenhang bringt
- und damit zum ersten Mal zeigt, dass Gesten zum kernlinguistischen Gegenstandsbereich der Grammatikschreibung gehören.

Warum wurden syntaktische Untersuchungen redebegleitender Gesten allein – seit Kendons wegweisenden Aufsätzen (1972 und 1980) gab es praktisch keine wirklich relevanten Neuerungen – und von Gesten in Relation zur Lautsprache bisher vernachlässigt?

Dafür gibt es wissenschaftsgeschichtliche Gründe. Aktuell lassen sich in der Gestenforschung und in der Bildwissenschaft zwei Hypothesen anführen, die syntaktische Untersuchungen bisher eher behindert haben. Die Gestenforschung wird durch McNeills These dominiert, dass redebegleitende Gesten idiosynkratisch seien und nur unter Bezugnahme auf eine verbale Bezugsgröße interpretierbar. Gesten werden als „Fenster zum Geist" betrachtet, die Beschreibung gestischer Kodierungsformen und ihre Einbettung in linguistische Strukturen rückt dabei im Hinblick auf eher psychologisch orientierte Fragestellungen in den Hintergrund. Es besteht vielmehr die Annahme, dass redebegleitende Gesten einen direkten, nicht durch einen Kode verstellten oder „verfälschten" Zugang zu den bildlichen Vorstellungen bieten.

Die Bildwissenschaft ist durch die Goodmansche These dominiert, dass Bilder und damit zumindest auch ikonische Gesten „syntaktisch dicht" seien. Diese These macht sich auch Stetter (2005: 33) zu eigen und behauptet:

> „[...] Mimik und Gestik haben keine Syntax. Das Produkt ist eine mediale Performanz, die ein Original ist wie ein Bild, auch wenn dieses Bild eine ephemere Erscheinung ist. Dieses Bild ist primär noch vom sprachlichen Text her strukturiert, aber es ist doch kein Text mehr, der gelesen, sondern ein Bild, das wahrgenommen wird und das deswegen das Gefühl als erste Instanz anspricht". (Stetter 2005: 33)

Weiter:

> „Das gestische Artikulationsschema ist offenkundig syntaktisch dicht, also analog konstruiert. Darüber hinaus aber gibt es keine Typik, die hier noch in Anschlag zu bringen wäre. Hier kommt eine Qualität des signifiant ins Spiel, die einer anderen Ordnung von Phänomenen angehört als die, in denen Syntax und Semantik organisiert sind, eine Qualität, die diesen vorgeordnet ist: der Fülle der Artikulation *als Erscheinung*." (Stetter 2005: 35)

Stetter bezieht sich hier auf den Begriff der syntaktischen Dichte nach Goodman (1981: 136):

> "A scheme is syntactically dense if it provides for infinitely many characters so ordered that between each two there is a third." (Goodman 1981: 136)

Synaktische Dichte liegt dann vor, wenn „zwischen zwei Bildelementen, die sich nur bezüglich eines visuellen Parameters (Lage, Größe, Farbton, Helligkeit usw.) unterscheiden, immer noch eine Zwischenstufe möglich ist, weil alle diese Parameter kontinuierlich variieren" (Posner und Schmauks 1998: 23).

Die Tatsache, die von Stetter und Bildwissenschaftlern wie Sachs-Hombach (2003: 103ff.) übersehen wird, ist, dass erstens Gesten nicht nur bildlich sind, sondern beispielsweise auch nichtikonische Zeigegesten umfassen, und dass zweitens, wenn sie ikonisch sind, es sich nicht um statische Bilder handelt, sondern um natürlichsprachliche Bilder in Bewegung. Diese Bilder in Bewegung lassen sich als Gestenphasen notieren, die hierarchisch strukturiert sind (Kendon 1972, 1980, 2004). Darüber hinaus lassen sich ihnen, wie wir in diesem Buch zeigen werden, über eine kontextfreie Phrasenstrukturgrammatik in Anlehnung an die frühen Arbeiten Chomskys rekursive Konstituentenstrukturen zuschreiben. Damit wäre Stetters Annahme der syntaktischen Dichte von Gesten widerlegt.

Fassen wir unsere „Antithesen" zusammen. Anders als Stetter behaupte ich Folgendes:

1. Gesten sind nicht beschränkt auf die Bühlersche Funktion der Kundgabe, sondern werden auch im Hinblick auf ihre Darstellungsfunktion produziert und rezipiert (vgl. Bühler 1934/1982, Müller 1998, Kendon 2004).
2. Gesten sind unterhalb der Ebene der Lexikalisierung typisierbar und als Typisierungen linguistisch beschreibbar.[5]
3. Gesten sind nicht im Goodmanschen Sinn syntaktisch dicht, sondern als rekursive Konstituentenstrukturen beschreibbar.
4. Syntaktischen Strukturen und Funktionen lassen sich multimodal instanziieren.

---

5 Ausgangspunkt ist hier der Begriff der „Gestenfamilie" nach Kendon (2004) und Müller (2004).

1.2 Zum Forschungskontext

Die vorliegende Arbeit bewegt sich innerhalb zweier Forschungskontexte, einerseits in Sprachwissenschaft, dem Schwerpunkt der Arbeit, und andererseits in Gestenforschung. Warum ist bisher das Projekt einer multimodalen Grammatik weder ernsthaft gedacht, geschweige denn praktisch in Angriff genommen worden?[6]. Innerhalb der Gestenforschung besteht das Problem, dass die überwiegende Anzahl der Forscher keine Linguisten sind, sondern Psychologen, Kognitionswissenschaftler, Anthropologen und mit der Untersuchung von redebegleitenden Gesten andere Forschungsziele und Fragestellungen verbinden. Sind einige von ihnen dennoch Linguisten, handelt es sich entweder um Gebärdensprachlinguisten, in deren Gegenstandsbereich die Lautsprache nicht gehört, oder um Linguisten, die primär aus der Gesprächsforschung kommen und Gesten auf der Ebene der Pragmatik als reines Performanzphänomen betrachten. Grammatische Fragestellungen kommen auch bei letzterer Ausrichtung nicht in den Blick, und die Forscher verfügen daher über kein entsprechend elaboriertes linguistisches Methodenwissen. Grammatiker und Syntaktiker hingegen, die über elaborierte grammatische Beschreibungsverfahren verfügen, betrachten Gesten bisher als ihrem Gegenstandsbereich nicht zughörig. Sie verfügen in der Regel über keinerlei Wissen des empirischen Phänomenbereichs und halten allenfalls, wie immerhin z.B. Stetter, nur Gebärden innerhalb von Gebärdensprachen für grammatisch relevant.

An dieser Stelle eine Lücke zu schließen, ist ein Desiderat der Forschung. Ist eine partiell multimodale Grammatik möglich? Und, wenn ja, welches sind die Bereiche der grammatischen Interaktion zwischen Gestik und Lautsprache? Wie ist diese Interaktion beschaffen?

Die vorliegende Arbeit konzentriert sich im Rahmen der syntaktischen Fragestellungen auf drei Problemkreise. Die Lösung des dritten Problems setzt die Lösung des ersten und zweiten voraus:

---

6 Im Forschungsprojekt „Towards a grammar of gesture: evolution, brain, and linguistic structures“, in dessen Kontext das vorliegende Buch entstanden ist, liegt der Schwerpunkt ebenfalls auf Gesten im Sprachgebrauch, und zwar unabhängig von und vor einer möglichen lautsprachlichen Integration. Diese erste Stufe ist methodologisch notwendig, bevor in einem weiteren Schritt die Relation zur lautsprachlichen Ebene beschrieben werden kann.

1. Das Problem der gestischen Bezugnahme: Referieren Sprecher mit ikonischen Gesten unmittelbar auf das von ihnen mit ihrer Äußerung intendierte Referenzobjekt oder nur mittelbar über die Bedeutung der der Geste zugehörigen verbalen Bezugsgröße? Nur unter der Voraussetzung, dass Sprecher mit ihren Gesten auch unmittelbar referieren können, ist das Projekt einer oberflächenorientierten multimodalen Grammatik überhaupt möglich.
2. Das Problem des gestischen Kodes unabhängig von der Lautsprache: Weisen Gesten, anders als bisher angenommen, unabhängig von der Lautsprache eine eigenständige Syntax auf? Nur wenn man zunächst Gesten allein untersucht, ohne von vornherein den Standpunkt der Lautsprache einzunehmen, bekommt man diejenigen Phänomene in den Blick, die die Grundlage für die Untersuchung einer multimodalen Interaktion von Gestik und Lautsprache sein können.
3. Das Problem der Integration von Gesten in den lautsprachlichen Kode: Können Gesten syntaktische Strukturen und Funktionen innerhalb der Lautsprache instanziieren?

Die Möglichkeit einer multimodalen Grammatik sondiere ich aus primär linguistischer Perspektive mit linguistischen Methoden, die, soweit erforderlich, um semiotisch orientierte Beschreibungsverfahren ergänzt werden. Dabei schließe ich mich auf dieser ersten Stufe der Sondierung und des Entdeckens keiner expliziten grammatischen Schule an wie der Generativen Grammatik nach Chomsky, der Integrativen Sprachwissenschaft nach Lieb und Eisenberg oder der kognitiven Linguistik (z.B. Lakoff 1987, Lakoff/Johnson 1980 und 1999, Langacker 1987 und 1991, Fauconnier 1985 und 1997, Fauconnier/Turner 2002, Goldberg 1995, Croft/Cruse 2004) an, sondern ich verfahre „methodeneklektisch", um bestimmte Phänomene wie beispielsweise die Rekursivität von Gestenphasen überhaupt zum ersten Mal sichtbar zu machen. So benutze ich dafür das Werkzeug einer kontextfreien Phrasenstrukturgrammatik, ohne die Chomskyschen Implikationen einer Universalgrammatik zu teilen. Ganz allgemein verorte ich mich in der Tradition des Europäischen Strukturalismus, der auf die so genannten einzelsprachlichen „Oberflächenstrukturen" hin ausgerichtet ist. Diese Ausrichtung wird auch in Eisenbergs „Grundriß der deutschen Grammatik" explizit vertreten, der einen weiteren Ausgangspunkt für unsere syntaktischen Beschreibungen bildet. Multimodale Integration heißt in diesem Kontext zunächst Integration in Strukturen und Funktionen

der syntaktischen „Oberfläche“. Dies ist aber nicht die einzige Art möglicher Integration von Geste und Lautsprache. Sondern darüber hinaus können diese auch kognitiv im Sprachproduktionsprozess selbst integriert sein. Diese beiden Integrationstypen, syntaktische Oberflächenintegration und kognitive Integration, korrespondieren mit den zwei wichtigsten „Schulen“ innerhalb der Gestenforschung, der McNeill-Schule und der Kendon-Schule.

McNeill (1992 und 2005) ist primär an Gesten als einem „Fenster zum Geist“ interessiert, das erlaubt, Denken und Äußerungsprozess in ihrem Werden zu beobachten. In seinem Buch „Gesture and Thought“ (2005) grenzt er seinen Ansatz als dynamisch von den von ihm als statisch klassifizierten linguistischen Ansätzen in der Tradition Chomskys oder Saussures ab (McNeill 2005: 17).

> “The new step is to emphasize the ‘dynamic dimension’ of language—how linguistic forms and gestures participate in a real-time dialectic discourse, and thus propel and shape speech and thought as they occur moment to moment.” (McNeill 2005: 3)

Er beschreibt die kognitiven Grundlagen sprachlich-gestischer Äußerungen in Anlehnung an Wygotsky (1934/1971) als selbstorganisatorischen Prozess, der durch die Dialektik von bildlichen und propositionalen Repräsentationen angetrieben wird. Der „growth point“ (McNeill 1992: 219) bildet den kognitiven Ausgangspunkt dieses Prozesses. Er birgt gestalthaft Elemente des Ganzen in sich.

> “A growth point, or GP, is a minimal unit of dialectic in which imagery and linguistic content are combined. A GP contains opposite semiotic modes of meaning capture—instantaneous, global, nonhierarchical imagery with temporally sequential, segmented, and hierarchical language.” (McNeill 2005: 18)

Im kognitiven Entstehungsprozess interagieren beiderlei Arten von Repräsentation, sie manifestieren sich in Gestik und Rede: “Gestures and speech considered jointly, reveal a process in which holistic and imagistic representations interact with analytic and linguistic representations” (McNeill 1992: 218). Der gesamte Äußerungsprozess wird von Gestik und Rede gemeinsam durchlaufen, “gesture and speech arise from a single process of utterance formation” (McNeill 1992: 30), wobei sie gemeinsam, wenn auch in unterschiedlicher Weise, die zugrunde liegende gedankliche Einheit zum Ausdruck bringen.

Der entscheidende Punkt ist nun, dass McNeill erstens Bedeutung mit „imagery“ gleichsetzt, und zweitens behauptet, dass Gesten Bedeutungen ikonisch materialisieren:

> "To make a gesture, then, is to iconically materialize a meaning in actional and spatial form." (McNeill 2005: 56)

Weiter:

> "Images vary materially from no apparent gesture at all to elaborate multidimensional displays; but, hypthothetically, imagery is ever present. What varies is the amount of materialization." (McNeill 2005: 18)

Wenn nun Gesten Bedeutungen als Bilder materialisieren, dann repräsentieren sie diese nicht. Genau dieses Argument verwendet McNeill, um Goodmans Kritik der ikonischen Repräsentation in „Languages of Art" (1969/1981) als für Gesten irrelevant zu erklären. Goodmans These ist, dass die Repräsentationsfähigkeit eines Bildes unabhängig ist von seiner Ikonizität. Bilder repräsentieren, aber sie tun dies nicht weil sie ikonisch sind. Ikonizität ist nach Goodman für die Funktion der Repräsentation weder notwendig noch hinreichend. Betrachten wir die drei Argumente, die Goodman (1968/1981) bringt, genauer. Das erste Argument ist, dass ein bestimmtes Objekt sich selbst am ähnlichsten ist, sich aber dennoch nicht selbst repräsentiert. Das zweite Argument besteht darin, dass Ikonizität und Repräsentation zwei unterschiedliche Relationstypen konstituieren: Ikonizität ist symmetrisch, Repräsentation asymmetrisch, d.h. dass beispielsweise eine Person und ihre Darstellung auf einem Gemälde sich wechselseitig ähnlich sehen, wohingegen das Gemälde zwar die abgebildete Person repräsentiert, aber nicht umgekehrt. Das dritte Argument lautet, dass Ähnlichkeit für eine bildliche Repräsentation weder notwendig noch hinreichend sei, weil beispielsweise das Gemälde eines Gebäudes einem anderen Gemälde desselben Gebäudes viel ähnlicher sei als dem abgebildeten Gebäude selbst. Aufschlussreich ist nun, wie McNeill auf diese Argumente Goodmans reagiert:

> „We are compelled to accept these arguments in general. However, they do not apply to gestural images. Gestures differ from photos, paintings, etc. in two ways that remove them from Goodman's arguments.
>
> First, iconicty adds reality as a *material carrier*. By moving the hand up, upness is materialized. Iconicity has a different explanatory status in first- and third-person perspective. It can be a causal force in a first party's gesture and, for this person more iconicity equals greater materilization or reality of meaning, even though iconicity is unable to explain how a gesture represents something to a third party.

> Second, for the party performing a gesture, a gesture does not represent at all; the gesture is created by the speaker as a materialization of meaning. This fact alone removes it from Goodman's critique.
>
> Putting these properties together—images are global, synthetic, instantaneous, noncombinatoric, and dynamically shaped by context; they are actions nonrepresentational for the one creating them, and vary from simple to elaborate depending on the need for materialization; and we have a creation of meaning in action that is imagistic but not the experience of photo realism." (McNeill 2005: 58f.)

Zunächst geht es mir vor allem darum, McNeills Position zu verdeutlichen. McNeill, so können wir diesem Zitat entnehmen, behauptet, dass Gesten für den jeweiligen Sprecher Bedeutungen nicht repräsentieren, sondern materialisieren und so in gewisser Weise die Bedeutungen selbst sind. Je höher der Grad der Ikonizität, desto höher der Materialisierungsgrad der Bedeutung. Die Annahme eines gestischen Kodes verbunden mit der Annahme, dass Gesten repräsentieren, widerspräche McNeills Behauptung, dass Gesten einen direkten, gleichsam „unverfälschten" Zugang zur bildlichen Bedeutung des von ihm angenommenen *growth points* geben. Gesten sind für McNeill also keine Zeichen und können damit, so wäre die Konsequenz, nicht Gegenstand linguistischer und semiotischer Beschreibungsmethoden sein.[7] Der Sprecher ist in McNeills Modell, wie bei fast allen psychologisch orientierten Sprachproduktionstheorien, allein mit der Welt und spricht. Was passiert nun aber, wenn der Sprecher nicht für sich selbst, sondern zu jemandem spricht, und Gesten vom jeweiligen Adressaten der Kommunikationssituation interpretiert werden? Negiert McNeill auch in diesem Fall die Repräsentationsfunktion redebegleitender Gesten? Betrachten wir das folgende Zitat, in dem er das „H-Modell" (H wie Heidegger) der Kommunikation vorstellt:

> "The concept of a 'material carrier' of meaning is also taken from Vygotsky. It is extended via the 'H-model', after Heidegger, according to which the materialization of one's meaning in a gesture (and speech) is, for the one speaking, not a representation

7 McNeill gibt nirgendwo Kriterien an, wie man Gesten als Zeichen von Gesten als Nichtzeichen unterscheiden kann. Das Gesten als Medium grundsätzlich als Zeichen auf der Basis eines gestischen Kodes fungieren können, ist durch die Gebärdensprachen der Gehörlosen nachgewiesen. Von daher bleibt McNeills Behauptung, redebegleitende Gesten seien keine Zeichen und verfügten über keinen Kode, auch nach mehrfacher Wiederholung lediglich eine Behauptung.

> but an updating the speaker's momentary state of cognitive being. The greater this contribution, the more the materialization, hence the more developed the gesture. The listener, in turn, inhabits the same meaning by updating in parallel his or her own momentary being, communication being a matter not only of signal exchange but of social resonance and inhabitance in the same 'house of being'." (McNeill 2005: 19)

Auch Adressaten interpretieren redebegleitende Gesten laut McNeill als nicht repräsentierend, die in dieser Weise stattfindende Kommunikation wird als ein Phänomen sozialer Resonanz beschrieben. Wenn nun Gesten weder für Sprecher noch Adressat etwas repräsentieren, dann können Sprecher mit ihnen auch keine Dinge in der Welt bezeichnen. Deshalb können Sprecher nach McNeill auch nicht unmittelbar, sondern nur mittelbar über eine verbale Bezugsgröße („lexical affiliate") referieren.[8] Redebegleitende Gesten setzen nach McNeill das Gesprochene voraus. Ich werde gegen die These McNeills, dass Gesten nicht repräsentieren und über keinen Kode und damit auch über keine Syntax verfügen, an anderer Stelle Beispiele anführen. Hier möchte ich hervorheben, dass bei allen Bedenken zwei Dinge für unsere weitere Argumentation wichtig sind: erstens können Gesten sich auf die Bedeutung der Rede beziehen und nur mittelbar über die Bedeutung der Rede referieren, zweitens kann die Interpretation von Gesten für Sprecher und Adressat auseinanderfallen (siehe Kapitel 5).

Adam Kendon ist anders als McNeill nicht Psychologe, sondern Anthropologe, und damit sehr viel stärker an Gesten in der Interaktion und Kommunikation und an der Beschreibung gestischer Kodes interessiert. In seinem neuesten Buch „Gesture: Visible action as utterance" (2004) führt Kendon den Terminus „gesture-speech ensemble" ein (Kendon 2004: 128):

> "In creating an utterance that uses both modes of expression, the speaker creates an e n s e m b l e in which gesture and speech are employed together as p a r t n e r s in a single rhetoric enterprise." (Kendon 2004: 127)

Weiter:

> "Speech and gesture are partnered in the common enterprise of discourse construction. Neither is the cause or auxiliary of the other, nor is there any obligatory link between them." (Kendon 2004: 128)

---

8 Diese kann nach McNeill (1992) implizit sein.

Die Art und Weise, wie Geste und Rede sich zueinander verhalten, stimmt der Sprecher mit den von ihm intendierten kommunikativen Zielen ab. Geste und Rede stehen nach Kendon in keinem kausalen Verhältnis zueinander, beide Komponenten sind voneinander unabhängig (Kendon 2004: 128), werden vom Sprecher kontrolliert und in einer multimodalen Äußerung, dem „gesture-speech ensemble", integriert. Kendon kontrastiert seine Position, die die Integration von Geste und Rede an ein gemeinsames kommunikatives oder in Kendons Terminologie „rhetorisches" Ziel bindet, mit einer Position, die diese Integration in zugrunde liegenden Prozessen des Denkens und der Sprachproduktion verortet:

> „As a close examination of the coordination of gesture with speech suggests, these two forms of expressions are integrated, produced together under the guidance of a single aim. Is this because they are expressions of two different forms of thought that originate jointly in a single, 'deeper' process? Or are they integrated as a consequence of how a person, engaged in producing an utterance, adapts two separate modes of expression and conjoins them in a single rhetorical aim? Do the gestural expressions that so often are integrated with spoken expressions provide insight into the processes of thought that lead up to the organization and pronunciation of sequences of words? Or do they, rather, contribute in their own right to what is being said, and so enrich an expression that would otherwise be poorer if constructed out of words alone?" (Kendon 2004: 3)

Vergleicht man Kendons Position mit derjenigen McNeills, dann lässt sich feststellen, dass Gesten Bedeutungen nicht direkt materialisieren, sondern Formen sind, die bedeuten. Daraus folgt, dass Sprecher mit Gesten innerhalb eines „gesture-speech ensemble" auch auf außersprachliche Objekte referieren können. Anders als McNeill schließt Kendon für redebegleitende Gesten das Vorliegen eines in Ansätzen eigenständigen Kodes nicht aus. Es gibt nach Kendon keine strikte Trennung zwischen spontanen redebegleitenden Gesten und konventionalisierten Gesten wie Emblemen, die über eine stabile Form-Inhalts-Beziehung verfügen. Redebegleitende Gesten können sich über Konventionalisierungs- und Lexikalisierungsprozesse hin zu emblematischen Gesten entwickeln und sogar den Status einer Zeichensprache erlangen. Dies geschieht beispielsweise, wenn redebegleitende Gesten zunehmend sprachersetzend gebraucht werden.

> „It seems that the more gesture is used for the communicative tasks for which speech is used, the more it will develop properties that are linguistic in character. On this view, the sign languages that are today the objects of linguistic inquiry are the products of

> cultural evolutionary processes which occur whenever gesture is adapted to wider and ever more unspecialized utterance uses." (Kendon 2004: 285f.)

Mit anderen Worten: Je höher bei sprachlichen Äußerungen der Grad der Abwesenheit des vokalen Sprechens, desto verstärkter werden sprachliche oder sprachähnliche Konventionen bei Gesten ausgebildet. Diese Tendenz wird durch experimentelle Untersuchungen bestätigt, in denen Sprecher aufgefordert werden, eine Geschichte nur mit Hilfe von Gesten zu erzählen:

> "When the subjects had to describe the scenes using gestures alone, in most cases they invented gestures that named the object observed in the scene and they were more likely to use a hand shape that represented an aspect of the moving object in the movement or action gesture. Not very surprisingly, the subjects produced gestures in strings, creating gesture sentences. In these sentences, the objects in the scene were referred to first, the gesture depicting the action being given last." (Kendon 2004: 286)

Kendon nimmt also für redebegleitende Gesten potentiell sprachähnliche Eigenschaften an, deren spezifische Ausprägung von den vorliegenden kommunikativen Bedingungen, den Intentionen des Sprechers und der Einbettung in eine „Sprachgemeinschaft" abhängen. Dadurch dass Kendon die jeweilige Interaktion von Gestik und Rede in einer Äußerung als Resultat der kommunikativen Ziele des Sprechers auf der Basis eines gestischen Kodes ansieht, ist unsere Idee der multimodalen Instanziierung grammatischer Funktionen prinzipiell an Kendons Konzeption anschließbar. Zwar erlaubt Kendons Konzeption die Referenz auf außersprachliche Sachverhalte, doch unterscheidet er nicht explizit zwischen Bedeutung und Bezeichung. Während McNeill ganz klar Gesten nur einen Bezug auf die Bedeutung der verbalen Bezugsgröße zugesteht und damit keine unmittelbare Referenz auf das vom Sprecher intendierte Referenzobjekt zulässt, bleibt bei Kendon die Möglichkeit von mittelbarer und unmittelbarer Referenz unentschieden:

> "In some cases it may seem as if a gesture provides an expression parallel to the meaning that is provided in words. In other cases gesture appears to refine, qualify or make more restricted the meaning conveyed verbally, and sometimes we encounter the reverse of this. In yet other cases gesture provides aspects of the reference that are not present at all in the verbal component. In other cases, again, gesture may serve to create an image of the object that is the topic of the spoken component." (Kendon 2004: 161)

Für die Annahme gestischer Attribute hat die Frage, ob Gesten nur mittelbar oder auch unmittelbar referieren können, gravierende Konsequenzen. Beziehen sich

Gesten ausschließlich auf die Bedeutung einer verbalen Bezugsgröße in einer lautlichen Äußerung, dann kann es keine strukturelle oder funktionale Integration auf der „Oberfläche“ geben, denn Attribute spezifizieren nach der oben gegebenen Definition Eisenbergs nicht die Bedeutung, sondern das von einem Kernsubstantiv in einer Nominalgruppe Bezeichnete. Eine Integration wäre dann nur auf kognitiver Ebene möglich. In Abschnitt 1.5 werde ich die beiden Integrationstypen mit Hilfe des Peirceschen Zeichenbegriffs genauer charakterisieren und auf dieser Grundlage die Frage nach einer strukturellen und funktionalen syntaktischen Integration von Gestik und Lautsprache neu formulieren. In Abschnitt 1.4 werden u.a. die dafür notwendigen zeichentheoretischen Grundbegriffe eingeführt.

1.3 Datenerhebung, Transkription und Analyse

Die Beispiele, die wir verwenden, basieren entweder auf empirischen Belegen oder sind, wenn es sich um Konstruktbeispiele handelt, zumindest insofern empiriegeleitet, als dass diese datenorientiert konstruiert wurden. Datenorientiert heißt zum einen, dass das Beobachtungsvermögen und die Analysefähigkeit des Beispielkonstrukteurs eine Schulung durch die Transkription, Analyse und Kodierung empirischen Beispielmaterials durchlaufen hat, zum anderen werden, um der argumentativen Klarheit willen, belegte multimodale Äußerungen klassischen linguistischen Testverfahren wie der Substitutionsprobe unterworfen. Vielfach wird durch die empirischen Belege zum ersten Mal gezeigt, dass bestimmte Phänomene überhaupt auftreten können. Empirisches Arbeiten in dieser Form beinhaltet immer eine Entdeckungsprozedur. Darüber hinaus ist insbesondere für redebegleitende Gesten der Rückgriff auf die Kompetenz eines „native speaker“ wenig verlässlich, da redebegleitende Gesten sehr viel schwerer zu zitieren und memorieren sind als lautsprachliche Äußerungen. Die in dieser Arbeit verwendeten Beispiele sind überwiegend einem Korpus von Wegbeschreibungen entnommen, welche im Dezember 2000 am Potsdamer Platz und in meinem damaligen Büro an der Technischen Universität Berlin aufgezeichnet wurden.

Insgesamt nahmen 33 Informanten, vorwiegend Erstsemestler, an der Datenerhebung teil. Es gab drei Informantengruppen. Die Mitglieder der Informantengruppe A gingen jeweils einzeln in Begleitung des Versuchsleiters einen bestimmten, zuvor festgelegten Weg am Potsdamer Platz, der am U-Bahnhof Potsdamer

Platz anfing und am Sonycenter endete (siehe die Fotos des „Parcours“ im Anhang). Jeder Informant dieser Gruppe bekam die Instruktion, diesen Weg einem anderen Informanten aus der Informantengruppe B, der den Weg nicht kannte, so genau zu beschreiben, dass dieser sich in die Lage versetzt fühlt, diesen Weg einer dritten Person aus der Informantengruppe C wiederum so genau zu beschreiben, dass diese den Weg eigenständig findet. An zwei Stellen sollte zudem ein Foto gemacht werden: Zum einen sollte ausgehend von einem Durchgang im Stella-Musicaltheater der Marlene-Dietrich-Platz fotografiert werden, zum anderen das Dach des Sonycenters am Ende des Parcours. Mit dieser Aufgabe sollte die Aufmerksamkeit auf bestimmte Stellen gelenkt werden. Das grundlegende Prinzip der Instruktionen ähnelt dem bekannten Spiel „Stille Post“. Die Instruktionen wurden des Weiteren nach dem Ort der Wegbeschreibung und der Art der verwendeten Hilfsmittel variiert: In der ersten Variante erfolgten die Wegbeschreibungen am Potsdamer Platz selbst, und zwar am Ausgangspunkt des Parcours ohne weitere Hilfsmittel, in den anderen drei Varianten fanden die Wegbeschreibungen in einem Büro der Technischen Universität Berlin statt. In der zweiten Variante wurde wie in der ersten der Weg ohne weitere Hilfsmittel beschrieben, in der dritten Variante stand für die Wegbeschreibung eine Karte des Potsdamer Platzes zur Verfügung, in der vierten ein großes Blatt Papier (DIN A2) und mehrere Filzstifte, um den Weg begleitend zu zeichnen. Wenn Informanten nach dem Zweck der Datenerhebung fragten, wurde angegeben, es handle sich um eine Untersuchung zur Informationsabweichung in der Kommunikation. Dass Gesten bei der Untersuchung eine Rolle spielen, wurde nicht erwähnt. Im Anschluss an ihre Wegbeschreibung wurden die Informanten in einem Fragebogen gebeten, ihre Vorkenntnisse des Potsdamer Platzes einzuschätzen und Beobachtungen und Kommentare zum Versuchsablauf zu notieren. Außerdem wurden sie gebeten, mir zu erlauben, das von ihnen aufgezeichnete Bild- und Tonmaterial in wissenschaftlichen Veröffentlichungen zu verwenden.

Die Wegbeschreibungen wurden ursprünglich erhoben, um Zeigegesten zu untersuchen (vgl. Fricke 2007). Für die Beschreibung multimodaler Attribuierungen sind sie deshalb besonders geeignet, weil zahlreiche Wegmarken mit ihren besonderen Eigenschaften durch attributiv erweiterte Nominalgruppen beschrieben werden. Diese Wegmarken sind am Potsdamer Platz besonders markant wie z.B. die

im Jahr 2000 noch existierende Infobox oder das halbrunde Gebäude des Sonycenters.

Abbildung 3: Die Infobox am Potsdamer Platz

In der folgenden Abbildung befindet sich der Glasturm des Sonygebäudes auf der rechten Seite, der obere Teil ist vom Bildrand abgeschnitten. Im Vordergrund ist die Infobox zu sehen.

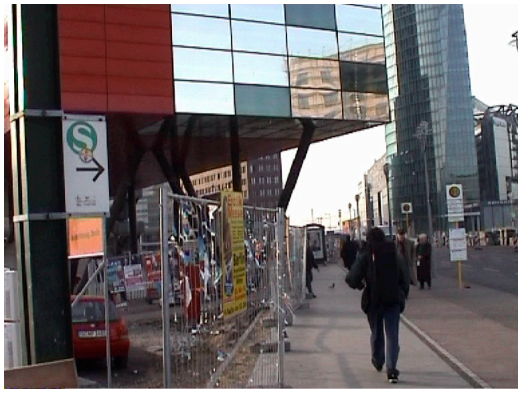

Abbildung 4: Das halbrunde Sony-Gebäude am Potsdamer Platz

Ein großer Vorteil des vorliegenden quasi-experimentellen Verfahrens ist, dass es für alle Wegbeschreibungen ein tertium comparationis derselben beschriebenen Situation gibt, den Parcours am Potsdamer Platz. Das heißt, für Gesten verschiedener Sprecher ist das vom jeweiligen Sprecher intendierte Referenzobjekt bekannt. Ohne ein derartiges tertium comparationis wären im fünften Kapitel einige wichtige Beobachtungen zur attributiven Funktion von Gesten nicht möglich gewesen. Die folgende Abbildung zeigt den Parcours am Potsdamer Platz mit der Infobox als Ausgangspunkt und dem Sonycenter als Endpunkt:

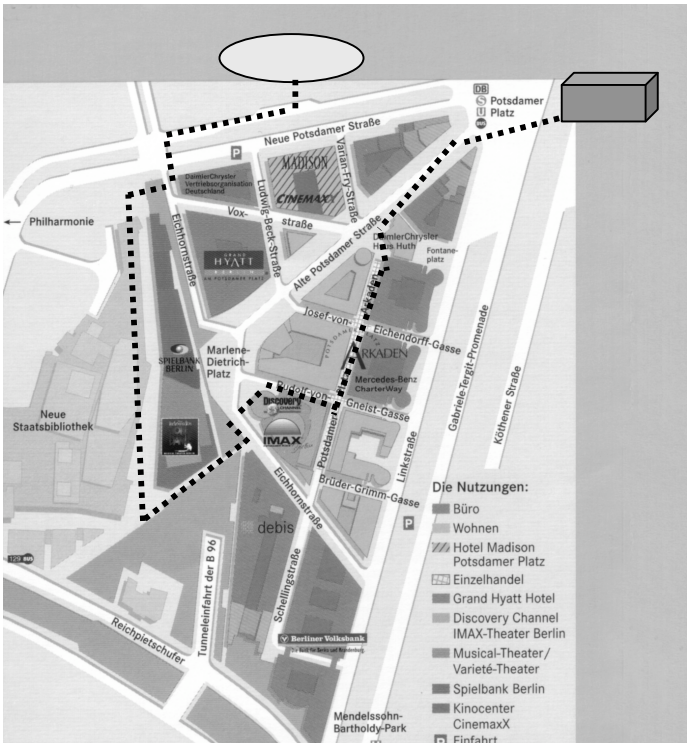

Abbildung 5: Der Parcours am Potsdamer Platz

Die Transkriptionskonventionen der in der vorliegenden Arbeit angeführten verbalen und gestischen Äußerungen orientieren sich weitgehend an den in Müller (1998), McNeill (1992) und Fricke (2007) verwendeten Transkriptionskonventionen. Sie sind im Anhang unter der Rubrik „Abkürzungen und Notationskonventionen“ aufgelistet (siehe Kapitel 7.1).

1.4 Redebegleitende Gesten: eine kurze Übersicht aus semiotischer Perspektive

1.4.1 Gestische und lautsprachliche Zeichen

Sowohl bei gestischen als auch lautsprachlichen Äußerungen handelt es sich um Zeichen nach Charles Sanders Peirce. Ein Zeichen ist eine triadische Relation in der etwas (das Zeichenmittel oder Repräsentamen) als für etwas anderes stehend (das Objekt) von einem Dritten interpretiert wird (Peirce 1983: 64), und zwar unabhängig davon, ob das Repräsentamen konventionalisiert ist oder nicht. In Abbildung 6 ist die Zeichenrelation nach Peirce schematisch dargestellt.

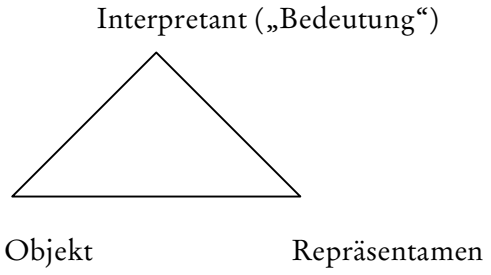

Abbildung 6: Die triadische Zeichenrelation nach Peirce

Nehmen wir die lautsprachliche Nominalgruppe *dieses gelbe Urlaubsformular* in Beispiel (3):

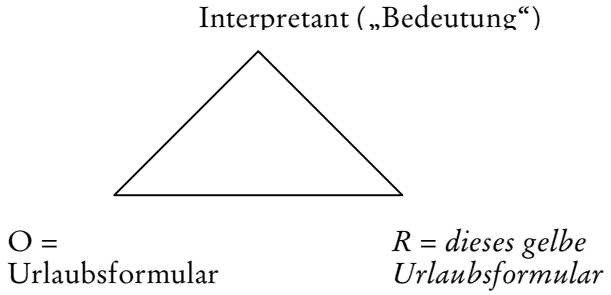

Abbildung 7: Die Instanziierung des Repräsentamens durch eine lautsprachliche Äußerung

Der Äußerung *dieses gelbe Urlaubsformular* wird vom Interpreten gemäß des im Interpretanten niedergelegten Kodes, hier der deutschen Sprache, eine Bedeutung zugeordnet, die es ihm ermöglicht, das vom Sprecher intendierte Referenzobjekt zu identifizieren. Sprachliche Bedeutungen charakterisiert Peirce als gewohnheitsmäßige Assoziationen zwischen Wortformen und Vorstellungen:

> “Any ordinary word as ‘give’, ‘bird’, ‘marriage’, is an example of a symbol. It is applicable to whatever may be found to realize the idea connected with the word; it does not, in itself, identify those things. It does not show us a bird, nor enact before our eyes a giving or a marriage, but supposes that we are able to imagine those things, and have associated the word with them.” (Peirce 1931–58, CP 2.298)

Die Unterscheidung zwischen Systembedeutung und Gebrauchsbedeutung geht mit einer Differenzierung unterschiedlicher Interpretanten einher (vgl. Fricke 2007: 193ff.). Der finale Interpretant lässt sich mit der konventionalisierten Bedeutung innerhalb einer Sprachgemeinschaft vergleichen. Er ist nach Peirce „die Gewohnheit, in deren Hervorbringung sich die Funktion des Zeichens erschöpft“ (Peirce

2000, Bd. 3: 224), etwa die Gewohnheit einer Sprachgemeinschaft, einem bestimmten Ausdruck regelhaft einen bestimmten Inhalt zuzuordnen (vgl. Peirce 2000, Bd. 3: 284). Es handelt sich dabei um „Gewohnheiten des inneren oder imaginierten Handelns, die von jeder Bezugnahme auf einen individuellen Geist abstrahiert wurden, in dem sie vielleicht zufällig verkörpert sind und dessen zukünftige Handlungen sie vielleicht leiten" (Peirce 2000, Bd. 3: 289).

Im Gegensatz zum finalen Interpretanten besteht der dynamische Interpretant in der tatsächlichen Wirkung auf den Geist eines Interpreten. Er entspricht in bestimmter Hinsicht der Äußerungsbedeutung eines Zeichens:

> „Der dynamische Interpretant ist die tatsächliche Wirkung, die in einem gegebenen Interpreten bei einem gegebenen Anlass bei einer gegebenen Phase beim Erwägen des Zeichens erzeugt wird." (Peirce 2000, Bd. 3: 225)

Dynamischer und finaler Interpretant beeinflussen sich wechselseitig. Der finale Interpretant bestimmt über Typen die „Verkörperung" der Gewohnheit als Token im dynamischen Interpretanten. Zugleich kann sich im dynamischen Interpretanten schrittweise eine neue oder veränderte individuelle Gewohnheit etablieren, die von anderen Interpreten der Interpretengemeinschaft übernommen wird und in einer Veränderung des finalen Interpretanten resultiert. Damit erlaubt der Peircesche Zeichenbegriff die Darstellung von Sprachwandel- und Konventionalisierungsprozessen, wie wir sie im dritten Kapitel zur Semantisierung und Typisierbarkeit syntaktischer Komponenten vorschlagen.

Redebegleitende Gesten verfügen, da sie im Gegensatz zu lautsprachlichen Ausdrücken nicht konventionalisiert sind, über keinen finalen Interpretanten im Sinne einer sprachlichen Systembedeutung. Sie können aber zunehmenden Konventionalisierungsprozessen ausgesetzt sein und sich beispielsweise über Zwischenstufen hin zu emblematischen Gesten entwickeln. Ein Sprecher gewöhnt sich in seiner Rede an, einer bestimmten gestischen Ausführung eine bestimmte Bedeutung zuzuordnen. Er bildet eine Gewohnheit aus, die von anderen Sprechern übernommen wird und schließlich eine größere Gemeinschaft von Gestikulierenden umgreift.[9]

---

9 Unter einer Gewohnheitsveränderung versteht Peirce „eine Veränderung in den Neigungen einer Person zum Handeln [...], die sich aus vergangenen Erfahrungen oder aus vergangenen Willensanstrengungen oder Handlungen oder aus Mischungen beider Ursachen ergeben. [...] sie

Betrachten wir die Geste in Beispiel (3), die die Äußerung *dieses gelbe Urlaubsformular* begleitet. Die Sprecherin benutzt ihre Finger wie einen Stift und zeichnet den Umriss eines Rechtecks, dessen Längsseite parallel zum Fußboden ist. Die folgende Abbildung zeigt die Rechteckgeste als Peircesche Zeichenkonfiguration:

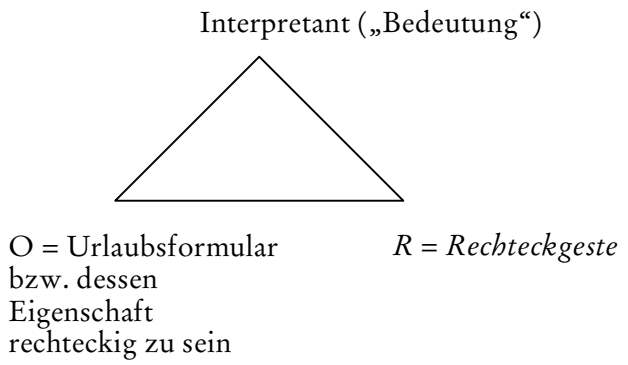

Abbildung 8: Die Instanziierung des Repräsentamens durch eine Geste

Anders als beispielsweise die lautsprachliche Äußerung von *Urlaubsformular*, verfügt die Rechteckgeste über keine lexikalisierte Bedeutung. Sie ist kein Token eines (Bedeutungs-)Typs. Dennoch spielen auch bei der Rechteckgeste Konventionen eine Rolle, keine sprachlichen Konventionen, sondern Konventionen der Interpretation von bildlichen Darstellungen, die es erst erlauben, die Geste als Darstellung eines zweidimensionalen Umrisses zu interpretieren. So betont Scholz (1991/2004: 43) in seiner Kritik der Ähnlichkeitstheorien des Bildes, dass das „Lesen“ von Bildern gelernt werden muss:

> „Dass manche Bilder einen einzelnen Gegenstand darstellen, andere jeden einzelnen von vielen Gegenständen, wieder andere gar keinen, versteht sich ebenfalls nicht von selbst. Es muss also in irgendeiner Weise gelernt werden, dass Bilder als singulär oder generell bezugnehmende oder im Sachbezug leere Zeichen dienen können. [...] Wir lernen ferner verschiedene Methoden, Figur und Grund voneinander abzusetzen. Auch werden die abgebildeten Gegenstände in aller Regel zurecht als dreidimensionale Dinge aufgefasst, obwohl die meisten Bilder (annähernd wenigstens) zweidimensionale Gebilde sind. Bekanntlich gibt es zahlreiche Projektionsmethoden, nach denen Dreidimensionales durch zweidimensionale Konfigurationen wiedergegeben werden kann. Mit solchen Verfahren muss man vertraut sein.“ (Scholz 1991/2003: 43).

---

schließt neben Assoziationen das ein, was man ‚Transsoziationen«‘ oder Veränderungen von Assoziationen nennen könnte [...]“ (Peirce 2000, Bd. 3: 283f.).

Wenn wir nun in Beispiel (3) die lautsprachliche Äußerung und die begleitende Geste gemeinsam betrachten, lässt sich die Relation zwischen beiden durch die folgende Zeichenkonfiguration beschreiben:

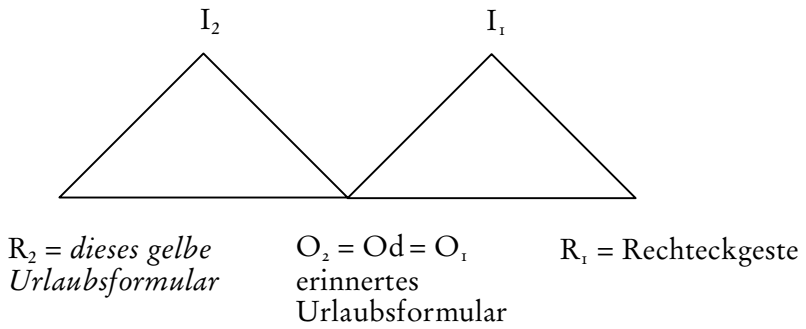

Abbildung 9: Die Zeichenkonfiguration von verbaler und gestischer Äußerung in Beispiel (3)

Vorsprachliche Wahrnehmungen oder Vorstellungen determinieren ein Repräsentamen, das als für diese vorsprachlichen Wahrnehmungen oder Vorstellungen stehend interpretiert wird. In Beispiel (3) referiert die Sprecherin zum einen mittels ihrer Äußerung *dieses gelbe Urlaubsformular* auf eine erinnerte Wahrnehmung oder Vorstellung eines bestimmten Formulars, zum anderen referiert sie auf dieselbe Erinnerung oder Vorstellung mittels einer ikonischen Geste, die die rechteckige Form des Formulars zeichnend nachahmt. Verbale Äußerung und begleitende Geste erfassen jedoch nicht dieselben Aspekte des erinnerten oder vorgestellten Formulars: Nach Peirce liegt mit der erinnerten Wahrnehmung oder der Vorstellung des Formulars zunächst ein dynamisches Objekt (Od) vor, das in zwei Zeichenrelationen, die verbale und die gestische, eintritt.[10] Die beiden Zeichen-

10 In Fricke (2007: 192) wird die Unterscheidung zwischen dynamischem und unmittelbarem Objekt folgendermaßen gefasst: „Das Objekt als zweites Korrelat der Zeichentriade lässt sich in etwa mit dem vergleichen, was in anderen zeichentheoretischen Ansätzen unter ‚Referent' oder ‚Bezeichnetes' verstanden wird. Peirce unterscheidet zwei Aspekte des Objekts: das dynamische Objekt und das unmittelbare Objekt. Das dynamische Objekt, d.h. das ‚Objekt, dessen Sein nicht von dem Zeichen abgeleitet ist, das aber das Zeichen beeinflußt, es darzustellen' (Peirce 2000, Bd. 2: 284) existiert außerhalb und unabhängig von der Zeichentriade. Das unmittelbare Objekt hingegen, d.h. dasjenige Objekt wie es ‚das Zeichen selbst darstellt und dessen Sein also

relationen erfassen unterschiedliche Aspekte des dynamischen Objekts. Diese Aspekte stehen als unmittelbare Objekte (O_1, O_2) in der jeweiligen verbalen und gestischen Zeichentriade. Der intendierte Referent von Geste und verbaler Bezugsgröße ist jedoch derselbe.

Die Unterscheidung zwischen dynamischem und unmittelbarem Objekt legt das Missverständnis nahe, dass es sich um zwei verschiedene Objekte handele: das Objekt, so wie es repräsentiert wird, und das Objekt an sich. Es handelt sich jedoch, wie Schönrich betont, „nicht um zwei Objekte, sondern um zwei Thematisierungsweisen ein und desselben Objekts [...]" (Schönrich 1990: 129). Innerhalb der Zeichentriade wird das Objekt thematisiert als etwas, das sich allein aufgrund der Aktivität der Semiose einstellt.

1.4.2 Gestentypen: Ikon, Index, Symbol

Ein Repräsentamen ist zunächst einmal dadurch gekennzeichnet, dass es vom Objekt verschieden sein muss, um es repräsentieren zu können. Hinsichtlich des Objektbezugs des Repräsentamens unterscheidet Peirce zwischen Ikon, Index und Symbol. Diese Verschiedenheit vom bezeichneten Objekt bedingt jedoch nicht, dass die Repräsentamen eine Klasse von Entitäten außerhalb und unabhängig von der jeweiligen Zeichentriade bilden, vielmehr gilt nach Peirce: „Ein Zeichen fungiert nicht als Zeichen, wenn es nicht als Zeichen verstanden wird" (Peirce 2000, Bd. 1: 424).

Als Ikon besteht zwischen einem Repräsentamen und seinem Objekt eine Beziehung der Ähnlichkeit (z.B. ein Bild, Schema oder Diagramm), als Index steht das Repräsentamen nicht in einer abbildenden, sondern als Hinweis oder Anzeige in einer realen, raumzeitlich bestimmten Beziehung zu seinem Objekt (z.B. Wegweiser, Krankheitssymptom oder Thermometer). Als Symbol ist ein Repräsentamen von einer Ähnlichkeit oder physischen Verbindung mit seinem Objekt unabhängig. Es ist im Wesentlichen in seinem Objektbezug wie ein Sprachzeichen durch Konventionalität und Arbitrarität gekennzeichnet (vgl. Peirce 2000, Bd. 2: 273). Peirces Zeichentypologie lässt sich für eine Klassifikation redebegleitender

---

von seiner Darstellung im Zeichen abhängig ist [...]' (Peirce 2000, Bd. 3: 145), ist ein Korrelat der Zeichentriade und abhängig von dieser.

Gesten und darüber hinaus auch anderer Gestentypen wie den emblematischen Gesten nutzbar machen. Abbildung 10 stellt einen Ausschnitt aus dem Klassifikationsschema redebegleitender Gesten nach Fricke (2007: 222) dar:

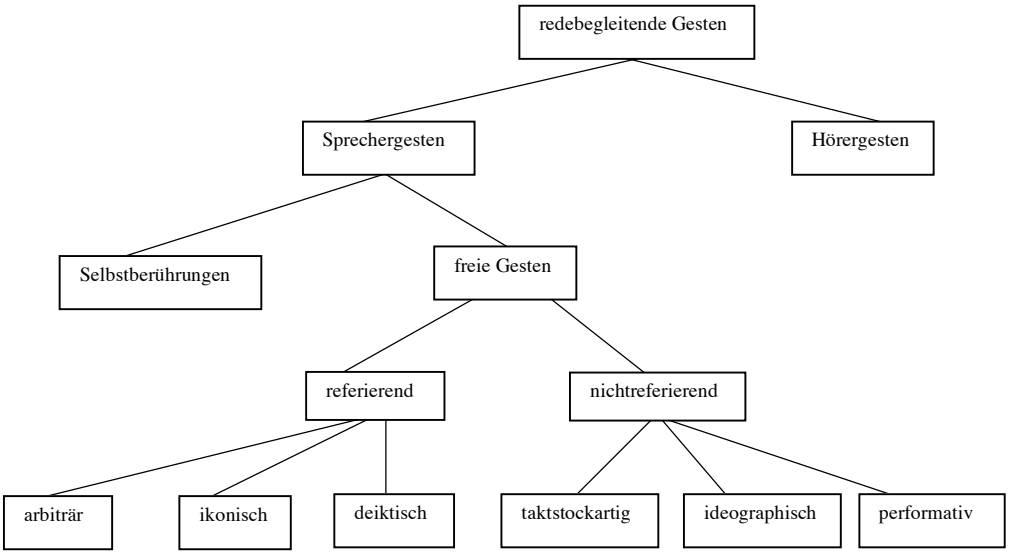

Abbildung 10: Ausschnitt aus dem Klassifikationsschema redebegleitender Gesten nach Fricke (2007: 222)

Redebegleitende Gesten lassen sich in Anlehnung an Müller (1998: 107) zunächst in „Sprechergesten“ und „Hörergesten“ unterscheiden. Bei den Hörergesten handelt es sich überwiegend um Selbstberührungen, bei den Sprechergesten hingegen um „freie Gesten“ (Müller 1998: 107), die vor dem Körper ausgeführt werden. Sprecher- und Hörergesten im Wechsel ausgeführt tragen zur Gesprächsorganisation bei:

> „Für alle Gesprächsteilnehmer wird durch das Wechselspiel von Selbstberührungen und freien Gesten der Status quo der Gesprächsrollen und damit die Verteilung des Rederechts auch gestisch fortlaufend erzeugt und sichtbar gemacht.“ (Müller 1998: 107)

Die freien Sprechergesten werden nach ihrem Referenzpotential unterschieden, nämlich danach, ob der Sprecher in einer Äußerung mit ihnen auf etwas referieren kann oder nicht. Nichtreferierende Gesten verfügen lediglich über eine parasprachliche Funktion. Ideographische Gesten im Sinne Efrons (1941/1972) bilden Aspekte des Ideations- und internen Sprachproduktionsprozesses ab, Taktstockgesten zeigen den Wechsel von Diskursebenen an oder heben besonders relevante Äuße-

rungsbestandteile hervor (McNeill 1992). Das folgende Beispiel zeigt eine Taktstockgeste. Die flache Hand bewegt sich in drei „Schlägen“ auf und ab[11]:

(4) B: [**noch**mal zur | **In**fo | **ähm**/]

Abbildung 11: Die Taktstockgeste in Beispiel (4) (Fricke 2007: 219)

Performative Gesten sind dadurch gekennzeichnet, dass ihr jeweiliges Bezeichnungspotential für die Äußerung als Sprechhandlung nicht konstitutiv ist. So wird mit der emblematischen Geste des Schwörens ein Schwur vollzogen, nicht jedoch auf diesen referiert. Zwar ist es prinzipiell auch möglich mit der Geste des Schwörens auf einen Schwur zu referieren, aber nur in einer Äußerung, die selbst keine Schwurhandlung ist, z.B. bei der Beschreibung eines Schwurs. Ähnliches gilt für performative redebegleitende Gesten. Eine Geste des „Wegwischens“ (Müller 1998: 111, Teßendorf 2005, in Vorb. a und b) referiert nicht, wenn zum Beispiel ein Argument „weggewischt“ und damit abqualifiziert wird, denn der Vollzug der negativen Bewertung findet durch die Geste selbst statt.

Referierende Gesten werden in Anlehnung an Peirce in einem ersten Klassifikationsschritt nach dem Objektbezug des gestischen Zeichens unterschieden, je nachdem, ob dieser auf Arbitrarität, Ähnlichkeit oder auf einer hinweisenden, origorelationalen Beziehung beruht. Im ersten Fall liegt eine arbiträre, im zweiten eine ikonische und im dritten eine deiktische Geste vor. Arbitrarität kommt bei redebegleitenden Gesten nur ansatzweise vor (Fricke 2007: 218), die beiden Hauptgestenklassen bilden Zeigegesten und ikonische Gesten. Auf letzteren vor allem

11 Zum Begriff des Stroke siehe Kapitel 3.

liegt der Schwerpunkt dieses Buches. Das folgende Beispiel zeigt zunächst eine Zeigegeste, die als Repräsentamen ihr Objekt nicht charakterisiert, sondern lediglich auf es hinweist. Die Adressatin B ist in der Lage von einem Ursprung (Origo) aus, der bei der Sprecherin A liegt, das von der Sprecherin intendierte Referenzobjekt, eine von beiden Kommunikationspartnern wahrnehmbare Straße, durch die Verlängerung der durch den Arm und den Zeigefinger markierten Geraden aufzufinden:

(5) A: [du kommst hier **vor**ne raus an dieser Straße (.)]

Abbildung 12: Die Zeigegeste in Beispiel (5)

Im nächsten Beispiel versucht Sprecherin B mit einer ikonischen Geste, die einen Quader modelliert, ein Modell des Potsdamer Platzes aufzubauen und lokalisiert diesen Quader vor sich im Gestenraum.

(6) B: also [(.) **hier** iss die Infobox]

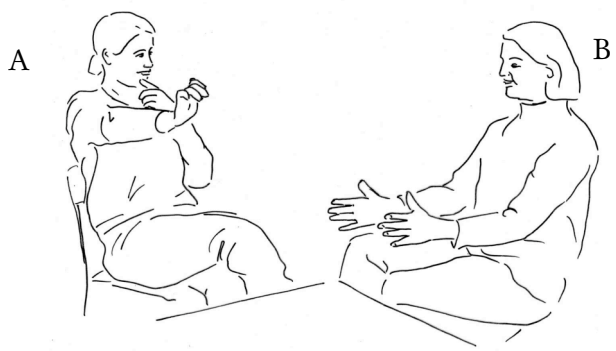

Abbildung 13: Die ikonische Geste in Beispiel (6)

Der Quader stellt die Infobox am Potsdamer Platz dar (siehe Abbildung 3). Dieselbe Infobox hätte jedoch auch als zweidimensionaler, rechteckiger Grundriss repräsentiert werden können. In ihrem Buch „Redebegleitende Gesten: Kulturge-

schichte – Theorie – Sprachvergleich“ (1998) unterscheidet Müller für ikonische Gesten zwischen unterschiedlichen Weisen der gestischen Darstellung. Der folgende Abschnitt fasst die wesentlichen Unterscheidungsmerkmale zusammen.

1.4.3 Gestische Darstellungsweisen nach Müller: die Hand agiert, modelliert, zeichnet, verkörpert

Ikonische Gesten können weiter danach subklassifiziert werden, in welcher Weise sie als Repräsentamen das Objekt repräsentieren. Müller (1998: 114ff.) unterscheidet vier verschiedene gestische Darstellungsweisen: „die Hand agiert“, „die Hand modelliert“, „die Hand zeichnet“, „die Hand repräsentiert“ bzw. in einer neueren Fassung „die Hand verkörpert“.

Bei der Darstellungsweise „die Hand agiert“ tut die Hand so, als vollzöge sie eine Handlung. Häufig werden dabei Gegenstände imaginiert, die an der Handlung beteiligt sind. Die dargestellten Handlungen werden ausschnittsweise pantomimisch nachvollzogen wie beispielsweise das Abnehmen eines Telefonhörers (vgl. Müller 1998: 115). In dem folgenden Beispiel Müllers greift die Sprecherin eine Handschaltung und zieht sie in Richtung ihres Körpers.

Bei der Darstellungsweise „die Hand modelliert“ wird ein Objekt oder Geschehensverlauf dadurch dargestellt, dass „die Hände es in der Luft nachmodellieren“ (Müller 1998: 117). In dem Beispiel (6) oben wird die Infobox als flüchtige Skulptur nachmodelliert.

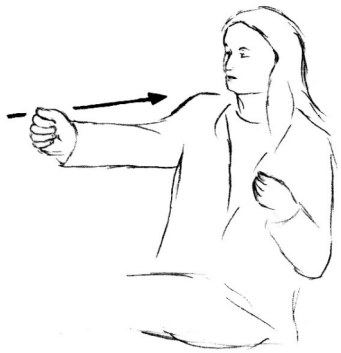

Abbildung 14: Der Darstellungsmodus „die Hand agiert“ in einem Beispiel von Müller (vgl. Müller 1998: 115)

Im Gegensatz zu den ersten beiden Darstellungsweisen, wird bei der Darstellungsweise „die Hand zeichnet“„keine dreidimensionale Gestalt geformt, sondern eine

zweidimensionale Form“ geschaffen (Müller 1998: 118). „Sie [die Hände] werden wie ein Zeichenstift verwendet, um den Umriß von Objekten oder den Bewegungsverlauf eines Ereignisses nachzuzeichnen“ (Müller 1998: 118). In Beispiel (3) wird das Urlaubsformular in seiner rechteckigen Gestalt gezeichnet. Die Sprecherin des folgenden Beispiels zeichnet mit ihrer rechten Hand auf einer imaginären Karte den Umriss eines durch *hier* bezeichneten Raumgebiets nach:

(7) A: [**hier iss das Bächlein/ (..)**]

Abbildung 15: Die Hand zeichnet in Beispiel (7)

Das Darstellungsmittel „die Hand verkörpert““ „unterscheidet sich von den drei erstgenannten insofern, als hier das dargestellte Objekt als Ganzes repräsentiert wird. [...] Die gestikulierenden Hände verwandeln sich in eine Skulptur“ (Müller 1998: 119f.). So werden im folgenden Beispiel (8) einzelne Gebäude am Potsdamer Platz durch auf einer imaginären Karte platzierte Handflächen verkörpert. Die linke Hand repräsentiert die Arkaden, die rechte Hand das Stella-Musicaltheater am Potsdamer Platz.

(8) [1][**hier** sind die Arkaden/ [2][**hier** iss das Musicaltheater\]$_1$ xxx]$_2$

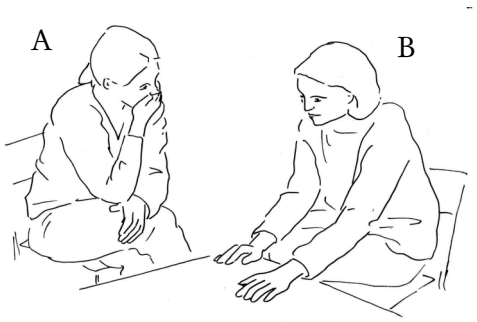

Abbildung 16: Die Hand verkörpert (Geste 1)

Abbildung 17: Die Hand verkörpert (Geste 2)

Der hier eingeführte Zeichenbegriff sowie die Unterscheidungen nach Gestentypen und die Subklassifikation der ikonischen Geste nach gestischen Darstellungsweisen sind Voraussetzung für die Untersuchung der gestischen Attributfunktion im fünften Kapitel.

1.5 Das Problem der gestischen Bezugnahme: mittelbare und unmittelbare Referenz bei ikonischen Gesten

Will man den Nachweis für die syntaktische Integration von Gesten in lautsprachliche Strukturen erbringen, dann stellt sich zum einen die Frage, ob und inwieweit Gesten unabhängig von der Lautsprache über einen eigenständigen syntaktischen Kode verfügen. Und wenn dies so ist, dann stellt sich weiter die Frage, wie gestische und lautsprachliche Strukturen interagieren und aufeinander beziehbar sind. Unabhängig von der Frage, ob Gesten über einen eigenständigen syntaktischen Kode verfügen oder nicht, stellt sich das Problem der gestischen Bezugnahme.

Folgt man McNeill, dann kann der Sprecher mit Gesten nicht referieren, sondern nur mittels der lautsprachlichen Bedeutungen, die die Gesten nach McNeill „materialisieren". Die Geste wird bezogen auf die Äußerung und die vom Sprecher mit der Äußerung intendierten Referenten zum Epiphänomen. Damit ist zugleich eine strukturelle und funktionale Integration von Gesten in die „Oberflächen"-Syntax der Lautsprache ausgeschlossen. Kendon hingegen bindet das „gesture-speech ensemble" an ein gemeinsames rhetorisches Ziel. Anders als McNeill gesteht er Gesten einen eigenständigen Kode zu sowie die Fähigkeit zur Repräsentation. Dennoch bindet auch Kendon tendenziell die Möglichkeit der Referenz nichtkonventionalisierter redebegleitender Gesten an den Umweg über die entsprechende verbale Bezugsgröße:

> "As far as can be seen at the moment, gesticulation is not composed of elements which are formed into constructions according to a syntax. They occur, rather, as a succession of enactments which express in presentational fashion the ideas which may also be given in discursive speech. The sequencing of gesticular phrases and units is governed by the order of presentation of ideas in the discourse." (Kendon 1975: 365)

Die Diskussion um das Vorliegen eines gestischen Kodes verstellt den Blick auf die grundlegende Dichotomie zwischen mittelbarer und unmittelbarer Referenz.

Wir nehmen hier den Grundgedanken Nelson Goodmans auf, nämlich dass Referenz der Kern der Repräsentation ist, unabhängig von der Ähnlichkeit:

> „Symbol ist nicht gleich Symbol, Darstellung nicht gleich Darstellung. Der Raum, der von pikturalen, gestischen, mimischen, verbalen, musikalischen, von artifiziellen Repräsentationen jedweder Art ausgefüllt wird, ist unübersehbar. Und doch lassen sich alle Weisen, in denen Symbole etwas bedeuten können, auf eine einzige Grundfigur zurückführen: die Bezugnahme auf Weltausschnitte, die von Menschen in ihrer Produktion und in ihrem Verständnis von Symbolen je wieder geleistet wird. Dies ist der Grundgedanke von Nelson Goodman *Languages of Art*." (Stetter 2005: 21)

In Fricke (2007) wurde für Zeigegesten nachgewiesen, dass sie einerseits unmittelbar auf das vom Sprecher intendierte Referenzobjekt referieren können (Zeigen auf Nichtzeichen), andererseits aber auch nur mittelbar über ein zwischengeschaltetes Demonstratum, welches als Zeichen für das vom Sprecher intendierte Referenzobjekt interpretiert wird (Zeigen auf Zeichen). Nehmen wir zum Beispiel einen Stift, der auf einem Tisch liegt. Der Sprecher zeigt mit einer Zeigegeste auf den Stift und äußert: „Da ist der Bach!". In diesem Fall fungiert der Stift nicht als Stift, sondern als Zeichen – beispielsweise innerhalb einer vorgestellten Karte – für einen in der Äußerungssituation nicht anwesenden Wasserlauf. Der Stift selbst ist also nicht das vom Sprecher intendierte Referenzobjekt, sondern lediglich das Demonstratum der Zeigegeste, welches als Zeichen für das eigentliche Referenzobjekt Bach interpretiert wird.

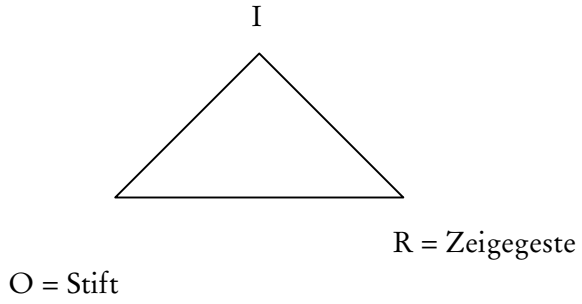

Abbildung 18: nicht-zeichenbezogenes Zeigen (unmittelbare Referenz)

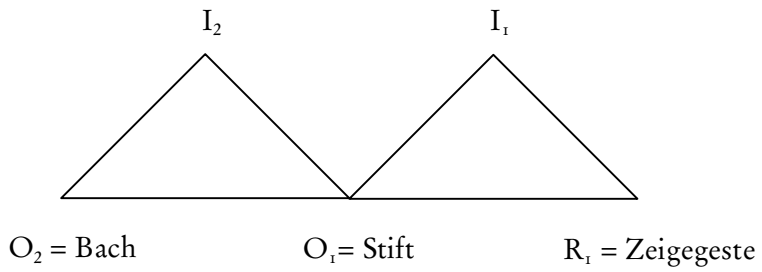

Abbildung 19: zeichenbezogenes Zeigen (mittelbare Referenz)

Die an Zeigegesten gewonnene Unterscheidung zwischen mittelbarer und unmittelbarer Referenz lässt sich auf ikonische Gesten übertragen. Betrachten wir das folgende Beispiel, das in Kapitel 5 ausführlich beschrieben wird. Die Sprecherin gehört zur Informantengruppe A und ist mit dem Versuchsleiter den in Kapitel 7.2 abgebildeten Parcours entlanggegangen. In ihrer verbalen Äußerung beschreibt sie eine Öffnung, durch die man von der Rückseite des Theatergebäudes auf den Marlene-Dietrich-Platz schauen kann. Diese Öffnung ist eckig. Interessant an dieser Äußerung ist nun, dass sie gestisch auf die Öffnung in unterschiedlicher Weise Bezug nimmt: In ihrer Beschreibung verwendet sie begleitend zur verbalen Äußerung der Wortform *Tor* einmal eine Geste, die ein Rechteck zeichnet, das andere Mal begleitend zur Äußerung der Wortform *Loch* eine Geste, die einen Kreis zeichnet. Die Form des Kreises widerspricht der Form der wahrgenommenen Öffnung im Theatergebäude. Wie kommt dieser Widerspruch zustande? Mit ihrer Rechteckgeste referiert die Sprecherin unmittelbar auf Aspekte des von ihr intendierten Referenzobjekts, die Geste ist also ohne ein weiteres zwischengeschaltetes Zeichen direkt auf das dynamische Objekt bezogen.

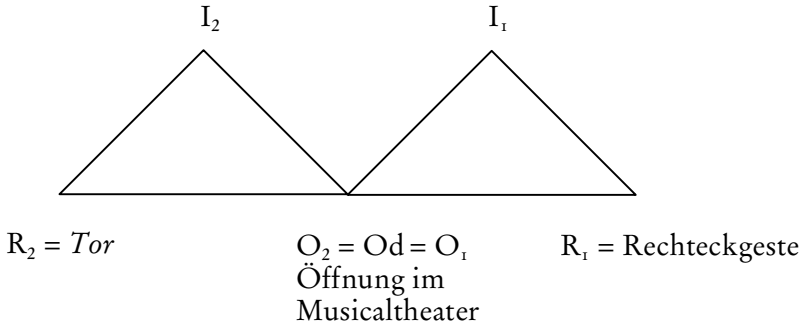

Abbildung 20: Die Zeichenkonfiguration einer objektbezogenen ikonischen Geste

Im zweiten Fall nimmt die Geste einen Umweg über die mit der Wortform *Loch* verbundene Bedeutung, z.B. im Sinne McNeills. Ich möchte an dieser Stelle offen lassen, ob es sich tatsächlich um die Bedeutung der Wortform *Loch* oder um ein mit dieser Wortform verbundenes Konzept handelt. Der Peircesche Begriff des Interpretanten schließt beide Möglichkeiten ein. Wenn wir unserer Darstellung wiederum die Peircesche Zeichenkonzeption zugrunde legen, dann handelt es sich um eine Geste, die sich auf den Interpretanten der Wortform *Loch* bezieht.[12] Die folgende Abbildung stellt diese Beziehung als Peircesche Zeichenkonfiguration dar.

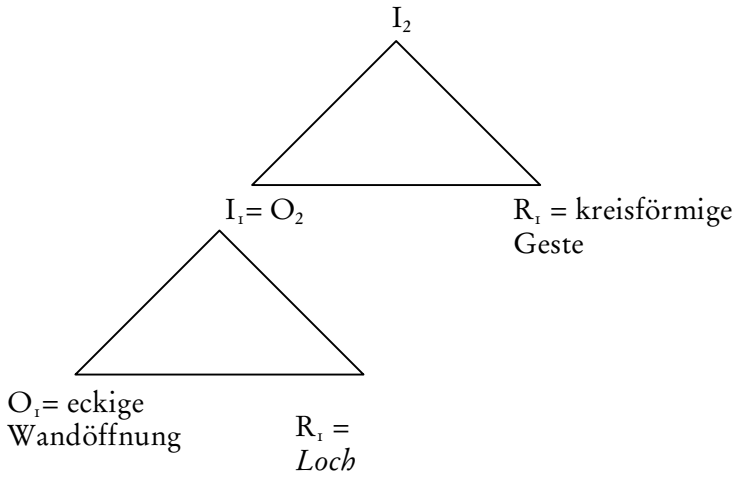

Abbildung 21: Gestischer Bezug auf den Interpretanten der verbalen Bezugsgröße

In Kapitel 5 wird die Unterscheidung zwischen objektbezogenen und interpretantenbezogenen Gesten weiter ausdifferenziert und mit weiteren Beispielen belegt. In diesem Abschnitt geht es zunächst darum, die Unterscheidung zwischen unmittelbarem und mittelbarem Referieren und zwischen gestischem Objekt- und Interpretantenbezug einzuführen.

Mit dieser Unterscheidung ergeben sich auch zwei unterschiedliche Typen möglicher Integration von Gestik und Lautsprache:

1. objektbezogene Integration, die der Ebene der Oberflächensyntax entspricht,
2. interpretantenbezogene Integration, die einer der Oberflächensyntax vorgängigen kognitiven Ebene entspricht.

12 Zur Konzeption des verbalen und gestischen Interpretanten siehe Fricke (2007: 200ff.).

Wir müssen uns also fragen, auf welcher Ebene gestisch-verbale Integrationen anzusiedeln sind:

1. Sind Gesten auf das Objekt ihrer verbalen Bezugsgröße bezogen?
2. Sind sie auf den Interpretanten ihrer verbalen Bezugsgröße bezogen?
3. Können beide Möglichkeiten vorkommen? Wenn ja, wie lässt sich entscheiden, welcher Fall vorliegt?

In Kapitel 5 werden wir zeigen, dass

1. sowohl objekt- als auch interpretantenbezogene Gesten bei Sprechern auftreten,
2. über die qualitativen Deiktika *so, solch* und *son* innerhalb von Nominalgruppen an der syntaktischen „Oberfläche" eine strukturelle Integration von redebegleitenden Gesten mit Attributfunktion geleistet wird,
3. sprecherseits interpretantenbezogene Gesten vom Adressaten als objektbezogen interpretiert und damit als oberflächensyntaktisch integriert betrachtet werden.

1.6 Zum Aufbau der Arbeit

Das Argument für die Möglichkeit einer multimodalen Syntax gleicht einem indirekten „Beweis": In Kapitel 2 grenzen wir zunächst den Begriff der Multimodalität von dem der Multimedialität ab und geben methodologische Argumente für eine multimodale Grammatik. In Kapitel 4 zeigen wir, dass Gesten allein (ohne Lautsprache) rekursive Konstituentenstrukturen zugewiesen werden können, also über einen syntaktischen Kode verfügen. Darüber hinaus zeigen wir in Kapitel 3, dass Instanziierungen einzelner Formparameter wie Handform, Orientierung der Handfläche, Position im Gestenraum, Bewegungsform und -richtung semantisierbar sind und sich wiederum zu komplexen Einheiten kombinieren können. Damit wird der These widersprochen, dass Gesten im Sinne Goodmans syntaktisch dicht und nicht syntaxfähig seien. Die Frage, ob bei diesen Formen und Strukturen ein gestischer Objekt- oder Interpretantenbezug vorliegt, ist auf dieser Ebene nicht entscheidbar. In Kapitel 5 wird am Beispiel des Attributs in der Nominalgruppe gezeigt, dass Gesten syntaktische Funktionen in lautsprachlichen Strukturen instanziieren können. Dass Gesten nicht aussschließlich auf den verbalen Interpretanten bezogen sind, wird durch die Analyse von Nominalgruppen mit dem Deiktikon *son* belegt. *Son* verweist kataphorisch auf eine Position in der lautsprach-

lichen Nominalgruppe, an der eine Qualitätsbeschreibung erfolgen muss. Diese kann auch gestisch instanziiert sein.

Da nun aber über die strukturelle Einbettung durch ein verbales Element der Objektbezug von ikonischen Gesten nachgewiesen ist und diese ikonischen Gesten ihrerseits Teile gestischer Konstituentenstrukuren sind, wie sie in Kapitel 4 beschrieben werden, ist über diesen Umweg auch die Möglichkeit des Objektbezugs für gestische Konstituentenstrukturen ohne Lautsprache gegeben. Damit ist eine wichtige Grundbedingung für die Möglichkeit einer multimodalen Grammatik erfüllt.

Der Objekt- und Interpretantenbezug von Gesten ist nun für Sprecher und Adressat nicht notwendigerweise derselbe, sondern sprecherseits interpretantenbezogene Gesten können von Adressaten als objektbezogen interpretiert werden. Für eine multimodale Grammatik bedeutet dies, dass die herkömmlichen zugrunde gelegten Kommunikationsmodelle wie z.B. das Bühlersche Organonmodell oder an Shannon-Weaver orientierte Modelle nicht hinreichend sind.

Mit dem Nachweis syntaktischer Strukturen und Funktionen von redebegleitenden Gesten, mit dem Nachweis ihrer Typisierbarkeit sowie mit dem Nachweis ihres Objektbezugs sind die Grundlagen gelegt für weiterführende Untersuchungen. Wir können zeigen, dass für Gesten und Lautsprache dieselben zugrunde liegende Prinzipien im Bereich der Semantisierung und Typisierung syntaktischer Komponenten (Kapitel 3), im Bereich der syntaktischen Strukturen (Kapitel 4) und im Bereich syntaktischer Funktionen (Kapitel 5) wirksam sind. Mit anderen Worten: Dieselben grammatischen Prinzipien wirken in Gestik und Lautsprache und leisten eine strukturelle Integration auf der Ebene des Sprachsystems. Wenn man nun dieselben Strukturen und Funktionen in Gestik und Lautsprache findet, dann lässt eine Umkehrung der Betrachtungsweise den Schluss zu, dass dieselben sprachlichen Strukturen und Funktionen sich zugleich in Gestik und Lautsprache manifestieren. Die Aufgabe einer multimodalen Grammatik ist dabei zu untersuchen, wo – das heißt z.B. auf welchen sprachlichen Beschreibungsebenen – die Integrationsbereiche von Gestik und Lautsprache liegen und wie diese beschaffen sind.

Kapitel 6 gibt eine Zusammenfassung der Untersuchungsergebnisse, stellt die Frage nach deren Implikationen für eine multimodale Grammatik und gibt einen Ausblick auf weiterführende Forschungsfragen und -projekte.

In Kapitel 7 befindet sich ein Verzeichnis der Abkürzungen und der Notationskonventionen, welche im Text und für die Analyse der redebegleitenden Gesten verwendet werden. Ferner sind dort die Abbildungen des Parcours am Potsdamer Platz aufgeführt sowie jeweils ein Verzeichnis der Abbildungen und Tabellen im Text.