Arnulf Deppermann

3 Sprache in der multimodalen Interaktion

Abstract: Der Beitrag plädiert für eine Untersuchung der gesprochenen Sprache als integralem Bestandteil multimodaler Interaktionspraktiken. Das leibliche Handeln bildet die Infrastruktur für die Verwendung von Sprache, es schafft Bedingungen, Möglichkeiten und Motivationen für die Verwendung spezifischer sprachlicher Strukturen; umgekehrt wird es seinerseits durch sprachliches Handeln organisiert. Zunächst werden in dem Beitrag grundlegende Eigenschaften multimodaler Interaktion dargestellt: die Vielfalt der leiblichen Handlungsressourcen und ihre Koordination, Sequenzialität und Simultaneität von Aktivitäten, multimodale Beteiligung an der Interaktion, der Stellenwert von Raum, Objekten, Multiaktivität und Bewegung. Ebenso wird kurz auf die methodischen Grundlagen der Untersuchung eingegangen: Videoaufnahme und multimodale Transkription. An drei sprachlichen Phänomenbereichen wird dann exemplarisch gezeigt, wie sprachliche Praktiken durch ihr Zusammenspiel mit anderen leiblichen Ressourcen der Kommunikation geprägt sind. Im Einzelnen geht es um die Disambiguierung sprachlicher Praktiken durch ihre Koordination mit anderen Ressourcen, die Erweiterung sprachlicher Strukturen, die aufgrund von Rezipientenreaktionen simultan zur Turn-Produktion stattfindet, und die Verwendungen minimaler Referenzformen, die sich auf die multimodale Ko-Orientierung der Beteiligten stützt.

Keywords: Deixis, Diskurspartikeln, Ellipse, Interaktionale Linguistik, Konversationsanalyse, multimodale Interaktion, Referenz, Videoanalyse

1 Einführung

Sprache kommt im wirklichen Leben nie abstrakt, als solche, und nur selten allein vor. Immer produzieren und begegnen wir ihr in einer spezifischen materialen Form (als gesprochene oder geschriebene Sprache). Meistens wirkt sie

Anmerkung: Für Kommentare zu einer früheren Version des Textes danke ich Heiko Hausendorf, Axel Schmidt und Reinhold Schmitt.

Arnulf Deppermann, Institut für Deutsche Sprache, R 5, 6–13, D-68161 Mannheim,
E-Mail: deppermann@ids-mannheim.de

https://doi.org/10.1515/9783110538601-004

zusammen mit anderen kommunikativen Ressourcen (dem Körper, Bildern, Filmen, Objekten). Welche sprachlichen Praktiken angemessen und für die Verständigung hinreichend sind, hängt von den InteraktionsteilnehmerInnen, der Räumlichkeit , akustischen Bedingungen, sichtbaren Objekten und vielem anderen mehr ab. Sprache ist also, zumal in der Interaktion unter Anwesenden, eingebettet in eine Infrastruktur leiblichen Handelns. Am folgenden Ausschnitt einer sozialen Interaktion können wir grundlegende Eigenschaften dieser Verwobenheit von Sprache mit dem leiblichen Handeln erkennen. Wir befinden uns bei einer Fortbildung von professionellen Rettungssanitätern, in der diese Notfälle durchspielen, um die Teamkooperation zu verbessern. In diesem Fall liegt eine Motorradfahrerin nach einem Zusammenstoß mit einem Auto am Boden. Die Sanitäter suchen die Ringerlösung (eine isotone Elektrolytlösung), mit der die Patientin versorgt werden soll, während sie sie am Arm verbinden.

(1) FOLK_E_00138_SE_01_T_01, c842, 11:12–11:28: Rettungssanitäterübung[1]

```
01 EL:   aber (.) wo_s jetz die RINger-
02 AS1:  (0.5) d_RINger liegt daneben da [unten.]
03 EL:                                   [ah-   ]
04       (1.4)
05 AS1:  dort JETZT,
06       (2.6)
07 EL:   also (.) halt mal,=
08 AS2:  =DRAN machen oder nich?=
09 EL:   =und halt mal HOCH und guck ob sie läuft.
```

Abb. 3.1: Rettungssanitätereinsatz.

1 Transkriptionen nach GAT 2 (Selting et al. 2009), die verwendeten Konventionen werden im Anhang dieses Bandes aufgeführt. Die hier verwendeten Datenbeispiele sind über die Daten-

Die Lektüre des Transkripts allein lässt zahlreiche Fragen offen: Was ist „die ringer“? Wo ist „daneben“ (Z. 02)? Warum muss „da“ in Zeile 02 noch ergänzt werden? Was ist mit „dort jetzt“ gemeint (Z. 05)? Was soll in den Zeilen 07–09 gehalten werden und wo soll es „dran“ gemacht werden? Abb. 3.1 vermittelt uns zwar eine grobe Idee der Interaktionssituation. Es hilft aber wenig, um die konkreten sprachlichen Wahlen, die intendierten Referenten und ihre Relevanz für die Erbringung der Erste-Hilfe-Leistung zu verstehen. Dazu ist es notwendig zu analysieren, wie die sprachlichen Äußerungen mit anderen multimodalen Ressourcen koordiniert werden.

Bereits Bühler ([1934] 1982: 155) meinte, dass sprachliche Zeichen oft wie Wegweiser eingesetzt werden, die die Synchronisation ansonsten stummer körperlicher Verhaltensströme regeln. Während Sprache in manchen Episoden sozialer Praxis keine bzw. nur eine untergeordnete Rolle spielt (z. B. beim Musizieren oder Schachspielen), ist sie dagegen in anderen dominant (z. B. in einer politischen Diskussion oder einem Vortrag). In jedem Falle ist Sprache aber kein autonomes Medium, sondern eine Ressource des leiblichen Handelns, die im Zusammenspiel mit anderen Bedeutung vermittelt, soziale Beziehungen herstellt und Handlungsprozesse organisiert.

In diesem Beitrag wird zunächst diskutiert, wie sich die multimodale Forschungsperspektive entwickelt hat (Unterkapitel 2 und 3). Ich gehe dann kurz auf die methodischen Grundlagen und Konsequenzen des veränderten Gegenstandsverständnisses ein (Unterkapitel 4). Sodann werden grundlegende Strukturen der multimodalen Interaktion dargestellt (Unterkapitel 5). Sie bilden die Infrastruktur für den Gebrauch von Sprache in der sozialen Interaktion zwischen Anwesenden. Schließlich wird an einigen sprachlichen Strukturen gezeigt, warum eine multimodale Analyse notwendig ist, um die schrittweise Konstitution, die Art und Weise des Einsatzes und die Motivation sprachlicher Strukturen in der sozialen Interaktion zu erklären (Unterkapitel 6).

2 Von der Welt als Text zur leiblichen, multimodalen Interaktion

In den Wissenschaften herrschte bis in die 1990er Jahre ein intellektualistisches Verständnis von Sprache und Bedeutung vor. Im strukturalistischen wie

bank für Gesprochenes Deutsch (DGD, dgd.ids-mannheim.de) verfügbar. Im Transkriptkopf wird jeweils das Gesprächsereignis (z. B. „FOLK_E_00 138_SE_01“), das zugehörige Transkript (z. B. „T01“) sowie die Nummer des ersten im Ausschnitt wiedergegebenen Beitrags (*contribution*, z. B. „c842“) angegeben; es folgt der Zeitausschnitt im Video.

im generativen Paradigma wurde Sprache als abstraktes System begriffen, welches unabhängig von seiner medialen Realisierung zu beschreiben ist. Auch in der Pragmatik, die ja einem weit verbreiteten Verständnis nach die Verwendung von Sprache im Kontext zu untersuchen hat, wurde Sprachgebrauch als Austausch abstrakter Symbole verstanden. Kontext kam hier als Kotext, als soziale oder als kognitive Faktoren, die den Sprachgebrauch bedingen, ins Spiel. Die Formel „Welt als Text" (Garz & Kraimer 1994) bringt die entsprechende epistemologische Position griffig auf den Punkt: Jede bedeutungsvolle Struktur ist textförmig organisiert, Bedeutung ist an Versprachlichung gebunden. Seit Ende der 1990er Jahre lässt sich dagegen ein *visual turn* (Bachmann-Medick 2006) konstatieren, der gegenläufig zur Fetischisierung des Texts die (kulturgeschichtlich wohl zunehmende) Rolle des Visuellen in der Gesellschaft der Gegenwart und seine Wichtigkeit für Wahrnehmung, Handeln und Sinnkonstitution betont. Die Entdeckung der Bedeutung des Visuellen für die menschliche Kommunikation geht Hand in Hand damit, den Stellenwert von Materialität und Medialität für die Kommunikation zu erkennen (z. B. Schneider & Albert 2013). Damit rücken der Körper und mit ihm die Leiblichkeit des Handelns ins Zentrum des Verständnisses sprachlicher Kommunikation. Darum wird es in diesem Text gehen. Multimodalität wird hier in der mündlichen Interaktion kopräsenter InteraktionsteilnehmerInnen behandelt.[2]

3 Kurzer Abriss der Geschichte der Erforschung multimodaler Interaktion

Die frühesten wissenschaftlichen Filmaufnahmen (bzw. Serienbilder) menschlicher Aktivitäten (Muybridge 1901) richteten sich auf motorische Abläufe und (etwas später) die ethnologische Dokumentation außereuropäischer Kulturen (de Brigard 1995). Das Interesse an der multimodalen Organisation sozialer Interaktion begann in den 1950er Jahren in der *context analysis* der Palo Alto Group, einer Gruppe von Psychiatern, Anthropologen und Soziologen, die die behavioralen Grundlagen von Beziehungssystemen auf der Basis von Videoaufnahmen von Therapiesitzungen untersuchten (Birdwhistell 1970; Scheflen 1972; Leeds-

2 Es interessiert uns in diesem Beitrag also weder die Multimodalität bzw. Multimedialität von Texten, Bildern und Filmen (Baldry & Thibaut 2005) oder von computervermittelter Kommunikation, etwa in *social media*, die zunehmend ins Zentrum der linguistischen Internetforschung rückt (vgl. Marx & Weidacher 2014), noch die Multimodalität sensorischer und motorischer Systeme von künstlichen Agenten in der Robotik (Grifoni 2009).

Hurwitz 2010). Aus diesen Anfängen entwickelte sich einerseits die experimentelle psychologische Erforschung nonverbaler Kommunikation, vor allem der emotionalen Valenzen des Gesichtsausdrucks (Ekman, Friesen & Ellsworth 1972), andererseits die Gestenforschung (Kendon 1972). Erving Goffmans Soziologie der rituellen Organisation alltäglicher Interaktionen (z. B. Goffman 1967) beruht wesentlich auf Analysen nonverbaler Ausdrucksformen; allerdings stützte er sich nicht auf Filmaufnahmen, sondern auf teilnehmende Beobachtung. Die Untersuchung verbaler Interaktionen war seit Anfang der 1960er Jahre Gegenstand der Konversationsanalyse (Sacks 1963, [1964–1972] 1992). Von Beginn an (und teilweise bis heute) arbeitete die Konversationsanalyse mit Audioaufnahmen und Verbaltranskripten, obwohl es auch Ansätze zur Beschäftigung mit Gestik auf Grundlage von Videodaten gab (Sacks & Schegloff [1974] 2002; Schegloff 1984). Die erste systematische Untersuchung konversationeller Strukturen auf Basis von Videoaufnahmen legte Charles Goodwin in seiner Dissertation 1981 vor. Die Goodwins (Goodwin & Goodwin 1986; Goodwin 1980) und Heath (1986) waren die Pioniere der multimodalen Analyse sozialer Interaktionen. Sie haben seither Interaktionen in einer Vielzahl von Settings untersucht (Goodwins: z. B. Tischgespräche, Kinderspiele, archäologische Ausgrabungen, Flughafenüberwachung, ozeanographische und geographische Expeditionen; Heath: z. B. Arzt-Patient-Gespräche, U-Bahn-Überwachung, Börsenhandel, Kunstauktionen, Museumsbesuche). Während sich die psychologische Untersuchung nonverbaler Kommunikation auf einzelne Kommunikationskanäle in Isolation bezieht und in Experimenten ihre kontextfreie expressive bzw. semiotische Funktion zu ermitteln sucht (z. B. Argyle 1975), richtet sich der soziologische Ansatz auf die Organisation des multimodalen Handelns in authentischen Kontexten sozialer Interaktion (Heath & Luff 2013). Gearbeitet wird hier mit Feldaufnahmen, um zu analysieren, wie kontextspezifische Handlungsaufgaben multimodal bewältigt werden (Heath, Hindmarsh & Luff 2010). Es werden multimodale Praktiken (Mondada 2014a, 2016a; siehe auch Deppermann, Feilke & Linke 2016) identifiziert, mit denen bestimmte Handlungen in bestimmten Kontexten vollzogen werden.

4 Methodische Erfordernisse der Untersuchung multimodaler Interaktion

Wenn wir soziale Interaktion zwischen räumlich ko-präsenten TeilnehmerInnen untersuchen wollen, sind Videoaufnahmen unerlässlich (Heath, Hindmarsh & Luff 2010; Mondada 2013). Dies ergibt sich aus dem Imperativ der

‚konstitutionslogischen Vollständigkeit des Datums': Für eine gegenstandsangemessene Analyse ist es erforderlich, dass wir als AnalytikerInnen Zugang zu den gleichen Handlungen und Ereignissen wie die Interaktionsbeteiligten haben, um nachvollziehen zu können, worauf sie in der untersuchten Situation reagieren und welche Ressourcen sie zur Organisation ihres Handelns benutzen. Können die InteraktionsteilnehmerInnen einander visuell wahrnehmen, benötigen wir also Videodaten, wenn wir nicht Fehlanalysen, Verständnislücken und das Übersehen relevanter Konstitutionsbedingungen und Vollzüge des Handelns riskieren wollen.[3]

Im Unterschied etwa zu Aufnahmen für psychologische Untersuchungszwecke oder zu massenmedialen Filmaufnahmen ist es erforderlich, nicht nur einzelne Akteure herauszugreifen, sondern die gesamte interaktive Konstellation einschließlich der für das Handeln der Beteiligten relevanten räumlichen Umgebung zu erfassen. Oftmals müssen mehrere Videokameras und zusätzliche Mikrofone eingesetzt werden, um die Szenerie aus verschiedenen Blickwinkeln zu erfassen, eine gute Tonqualität aller Beiträge sicherzustellen und die wünschenswerte Konzentration auf das Handeln von Fokuspersonen (Schmitt & Deppermann 2007), die eine zentrale Rolle in der aufgenommenen Interaktion spielen, mit der Erfassung der Gesamtszene zu verbinden.

Um Videodaten, zumal im Kontext eines linguistischen Untersuchungsinteresses, auszuwerten, ist eine Audiotranskription nach üblichen Standards der gesprächsanalytischen Transkription notwendig (wie CA, Hepburn & Bolden 2017, oder wie in diesem Artikel – und im deutschen Sprachraum üblich – nach GAT 2, Selting et al. 2009). In die Transkripte (bzw. den Analysetext) werden Standbilder aus den Videodaten eingefügt. Ihre Auswahl ist von dem zu untersuchenden Phänomen und den Phasen seines Verlaufs, die für die Analyse von Belang sind, sowie ästhetischen Kriterien (Sichtbarkeit, Kontrast, Bildqualität etc.) abhängig (siehe dazu Stukenbrock 2009; Mondada 2016b; Schmitt 2016). Zur Frage der visuellen Transkription existieren unterschiedliche Positionen. Manche Forscher verzichten auf die visuelle Transkription und ziehen die detaillierte Beschreibung der für die Untersuchung relevanten visuellen Ereignisse im Analysetext vor. Der Vorteil liegt in der Möglichkeit, viele verschiedenartige, oft nur sehr kurz währende Aktivitätsmomente mit einer Genauigkeit und

3 Auch ein Video dokumentiert nicht vollständig die sinnliche Erfahrung in einer Interaktion. Olfaktorische, gustatorische und haptisch-taktile Erlebnisse werden nicht repräsentiert. Allerdings ist anzunehmen, dass olfaktorische und gustatorische Aspekte für viele Interaktionen nicht relevant sind. Haptisch-taktile Erlebnisse haben oft einen sichtbaren Aspekt, der im Video zu beobachten ist und so wenigstens teilweise zum Analysegegenstand werden kann (Mondada 2016a).

Ausführlichkeit zu beschreiben, die im Rahmen des knapp bemessenen Raumes einzelner Transkriptzeilen nicht zu leisten und außerdem schwer zu rezipieren ist. Ein anderes Motiv des Verzichts auf ein Transkript kann darin bestehen, Aktivitätsphasen zu beschreiben, während derer nicht gesprochen wird. Der Nachteil dieses Verfahrens besteht darin, dass der Verlauf des visuellen Handelns in Bezug auf das verbale und auf zeitliche Parameter ohne Transkript nur unzulänglich wiederzugeben ist. Außerdem besteht die Gefahr, dass nur die visuellen Ereignisse wiedergegeben werden, die der Analytiker für relevant hielt. Damit wird aber dem Leser (und auch dem Analytiker selbst!) die Möglichkeit genommen, durch eine kontinuierliche Darstellung visueller Ereignisse (z. B. der Blickorganisation oder einer Gestenfolge) den Verlauf des kinesisch-visuellen Handelns und das Zusammenspiel verschiedener modaler Ressourcen miteinander systematisch in ihrem Verlauf zu verfolgen. Bislang gibt es für die visuelle Transkription keinen Standard. Es ist auch nicht zu sehen, wie zahlreiche Aspekte der visuellen Transkription standardisiert werden sollten. Die wesentlichen Gründe dafür liegen darin, dass die visuelle Transkription im Unterschied zur auditiven nie vollständig, sondern immer nur selektiv sein kann und sich auf die Verhaltensereignisse und Umgebungsaspekte beschränkt, auf die sich die InteraktionsteilnehmerInnen in ihrem Handeln selbst erkennbar beziehen. Je nach Untersuchungsfragestellung sind außerdem andere Ereignisse in unterschiedlicher Feinkörnigkeit der Beschreibung relevant. Aus diesen Gründen hat das visuelle Transkript einen anderen epistemologischen und forschungspraktischen Status als das Audiotranskript: Es ist weniger die Grundlage einer Datenanalyse als vielmehr ihr Produkt; es entwickelt sich im Verlauf der Analyse und manifestiert, welche Ereignisse und welche Modalitäten des Handelns als grundlegend für die untersuchte Interaktion erkannt worden sind (Mondada 2018). In diesem Artikel wird die in der multimodalen Interaktionsanalyse am meisten verbreitete Konvention angewandt (Mondada 2014b, 2018). Ihre Besonderheit im Vergleich zu anderen Systemen besteht darin, dass in Anlehnung an Kendon (2004) der genaue Zeitverlauf des Einsatzes (*onset*) und der Beendigung (*offset*) einer Aktivität, ihre Vorbereitung (*preparation*), ihr Höhepunkt (*apex*/*stroke*), das Anhalten einer Aktivität (*post-stroke hold*/*freeze*) und ihr Rückzug (*retraction*) bzw. die Transition zwischen verschiedenen Aktivitäten wiedergegeben wird. Die präzise und kontinuierliche Repräsentation zeitlicher Verläufe ist unabdingbar um festzustellen, wie soziale Handlungszusammenhänge durch die simultane und sequenzielle Koordination verschiedener multimodaler Ressourcen entstehen.

5 Grundlegende Strukturen multimodaler Interaktion

Die Produktion und die Bedeutung sprachlicher Strukturen entfalten sich im Zusammenspiel mit anderen leiblichen Modalitäten des Kommunizierens und mit der räumlichen Umgebung . Um besser zu verstehen, wie sprachliche Praxis durch diese Konstitutionsbedingungen geprägt ist und umgekehrt selbst zu ihnen beiträgt, ist es zunächst notwendig, universale Strukturen multimodaler Interaktion in den Blick zu nehmen, die die Konstitution sprachlicher Praxis bestimmen und damit auch für ihre adäquate Analyse grundlegend sind.

5.1 Multimodale Ressourcen und multimodale Praktiken

Multimodale Interaktion unter Anwesenden ist leibliche Interaktion. Der Begriff der ‚Leiblichkeit' erscheint geeigneter als derjenige der ‚Multimodalität', um zu fassen, dass Handeln intentionale körperliche Aktivität ist, welche größtenteils nicht bewusst, nicht sprachlich und nicht reflektiert geschieht, dennoch aber sinnhaft orientiert und holistisch geformt ist (vgl. Merleau-Ponty [1945] 1966). ‚Multimodalität' hebt dagegen hervor, dass leibliches Handeln Ressourcen verschiedener Modalität benutzt. Diese Ressourcen sind Vokalität (einschließlich Sprache und Prosodie), Gestik, Blick, Mimik, die Einnahme von Körperposituren, die Bewegung im Raum und der Umgang mit Objekten. ‚Multimodalität' impliziert also einen analytischen Blick, der auf den spezifischen Einsatz, die besondere Leistung und die Koordination der einzelnen Ressourcen fokussiert. Die alltagsweltliche Akteursperspektive ist dagegen holistisch, sie versteht ganzheitliche Bedeutungen.

In der multimodalen Interaktion benutzen die Teilnehmer multimodale Praktiken. Diese bestehen in zeitlich organisierten Mustern der Koordination multimodaler Ressourcen (Wie) zum Vollzug bestimmter Handlungen (Wozu). Oftmals sind sie an bestimmte Kontexte (Wann) gebunden. Ein besonders gut untersuchtes Beispiel ist die Organisation der lokalen Referenz mit deiktischen Mitteln (C. Goodwin 2003; Fricke 2007; Stukenbrock 2015). Im Ausschnitt (1) aus der Rettungsübung antwortet AS1 auf die Frage von EL, wo sich die Ringerlösung befindet (Auszug aus (1), Transkript multimodal erweitert):[4]

4 Transkription der multimodalen Phänomene nach Mondada (2014); siehe Unterkapitel 9 für eine Auflistung der verwendeten Konventionen.

(1) FOLK_E_00138_SE_01_T_01, c842, 11:12–11:28: Rettungssanitäterübung

```
02 AS1:  +(0.5) d_%RINger %liegt #da*neben +*%da#+*%[unten.]
   as1:  +Blick->Ringer--------------------+.....+Blick->
         Strips-->
   as1:                                *.......*zeigt*,,,
03 EL:                                                [ah-   ]
   el:   gradeaus-%.......%Blick->links------%.....%Blick->
         Ringer---->
                                #Abb.3.2       #Abb.3.3
```

Abb. 3.2: AS1 bereitet Strip vor, OC sucht Ringerlösung.

Abb. 3.3: AS2 zeigt auf Ringerlösung mit linkem Daumen, OC schaut auf Ringerlösung.

Schon vor Beginn seiner Antwort schaut AS1 auf die Flasche mit der Ringerlösung, die rechts vor dem Fahrzeug liegt. Währenddessen blickt EL vor sich auf den Boden, wo er die Ringerlösung sucht. Als AS1 zu antworten beginnt, wendet EL seinen Blick nach links. Dorthin schaut er, als AS1 den ersten Lokalisierungsausdruck „daneben“ verwendet (Abb. 3.2). EL richtet seinen Blick aber auf einen Suchraum (C. Goodwin 2003), in dem sich die Flasche nicht befindet. Daraufhin präzisiert AS1 die Referenz durch eine Zeigegeste, deren Apex (Kendon 2004) mit der Produktion des deiktischen Ausdrucks „da“ koinzidiert (Abb. 3.3). EL folgt der Zeigegeste mit seinem Blick und macht mit der Partikel „ah–“ (Z. 03) deutlich, dass er nun das gewünschte Objekt identifiziert hat.

Diese Episode ist ein typisches Beispiel der multimodalen Praktik deiktischer Raumreferenz (Stukenbrock 2015). Sie beinhaltet seitens desjenigen, der referiert, die Koordination von Blick und Zeigegeste auf das Verweisziel sowie deiktischem Ausdruck; dabei gehen Blick und Ansatz der Zeigegeste der Nennung des Deiktikums oft voraus. Der Adressat folgt der Zeigehandlung mit seinem Blick (*gaze-following*) und verdeutlicht mit einer anschließenden Verste-

hensdokumentation, dass erfolgreich geteilte Aufmerksamkeit (*joint attention*) hergestellt und der Referent identifiziert wurde. Die multimodale Praktik besteht in einer systematischen Koordination unterschiedlicher multimodaler Ressourcen. Praktiken sind zwar durch bestimmte materiale Ressourcen und Verhältnisse zeitlicher Koordination ihrer Form nach bestimmt; sie sind aber flexibel und können an unterschiedlichste Situationen angepasst werden. So kann z. B. nicht nur mit dem Zeigefinger oder der Hand, sondern auch mit Objekten, Augen- oder Kopfbewegungen, den Lippen oder dem ganzen Körper gezeigt werden (Enfield 2009).

Koordination geschieht sowohl intra- als auch interpersonal (Deppermann & Schmitt 2007; Deppermann 2014). Intrapersonal koordiniert bspw. der Referierende das Timing von Blick, Zeigegeste und Strukturierung seines Redebeitrags. Interpersonal koordinieren beide Akteure ihre Aktivitäten. So reagiert der Referierende AS1 in Ausschnitt 1 darauf, dass der Adressat EL an der falschen Stelle sucht, indem er mit seiner Zeigegeste und dem Lokaldeiktikum „da“ (Z. 02) die erste unspezifische Raumreferenz „daneben“ (Z. 02) weiter präzisiert. Der Adressat seinerseits folgt mit dem Blick der Zeigegeste. Das Beispiel zeigt, wie die intrapersonale Koordination mit der interpersonalen verknüpft ist: Der Einsatz unterschiedlicher Ressourcen wird fortlaufend an die Reaktionen des Gegenübers angepasst.

Die Analyse des Beispiels zeigt, warum eine genaue multimodale Transkription nicht nur der sprachlichen Äußerungen, sondern auch der kinesisch-visuellen Aktivitäten wichtig ist. Die genauen Zeitverhältnisse des Einsatzes der einzelnen Ressourcen, ihre Ausgestaltung und Bewegungsrichtung sind erst mit Hilfe des multimodalen Transkripts präzise zu bestimmen. Das gilt zumal für das Verhältnis der Aktivitäten mehrerer Beteiligter. Die Transkripterstellung wirkt methodisch disziplinierend, da sie zum systematischen Verfolgen der Verwendung der einzelnen Ressourcen über die Episode hinweg zwingt und die Aufgabe stellt, eine genaue, stimmige und lückenlose Analyse der Produktion der Aktivitäten und ihrer Transitionsmomente zu liefern. Für den Leser wird durch das Transkript die Fundierung der Analyse in einer Weise transparent und prüfbar, wie es bei einer Beschreibung in Prosa nicht möglich ist.

In früheren Forschungsansätzen wurde zumindest begrifflich die Sprache als dominante Kommunikationsressource verstanden. Der Begriff der ‚nonverbalen Kommunikation‘ behandelt alle nicht-sprachlichen „Kommunikationskanäle“ als kontrastive Restkategorie; die soziolinguistische Theorie der *contextualization of language* (Auer & di Luzio 1992; Gumperz 1982) versteht Sprache als etwas, das im konkreten Gebrauch kontextualisiert wird. In der sozialen Interaktion ist aber Sprache keineswegs immer die primäre Ressource des Handelns – die Rettungsübung etwa zielt nicht darauf ab sich zu unterhal-

ten, sondern die medizinische Erstversorgung von Patienten einzuüben. Ebenso entspricht es nicht der Handlungsperspektive von Beteiligten, für abstrakte sprachliche Formen einen Kontext herzustellen, sondern mit Hilfe des koordinierten Einsatzes verschiedener körperlicher Ressourcen bestimmte Handlungen zu vollziehen. Sprache ist also ein integrales Element von multimodalen Praktiken. Daher ist es methodologisch und begrifflich nicht angemessen, von einem apriorischen Primat der Sprache in der sozialen Interaktion auszugehen (Schmitt 2015). Vielmehr ist es eine empirische Forschungsfrage, welche Rolle der Sprache in einem konkreten Interaktionskontext in Relation zu anderen Ressourcen zukommt (Mondada 2016b). Allerdings ist Vokalität, zumal wenn sie sprachlich artikuliert wird, die einzige Ressource, die auf Kommunikation spezialisiert ist. Sie steht daher von vornherein unter Semiotizitätsverdacht. Wenn Sprache benutzt wird, so können wir aber annehmen, dass sie (fast) immer auch eine kommunikative Verwendung hat.[5] Dies gilt für andere körperliche Ausdrucksformen nicht in vergleichbarer Weise, denn sie können immer auch nicht-kommunikativ eingesetzt werden. Die interaktionale Linguistik wendet sich natürlich solchen Typen multimodaler Interaktion zu, in denen Sprache eine tragende Rolle spielt. Dabei kann es sich um empraktische Interaktion handeln, bei der das gegenständliche Handeln den Handlungszweck bildet, während Sprache eine organisierende Rolle zukommt (wie im Erste-Hilfe-Training, in praktischen Fahrstunden oder beim Fußballspiel). Aus der Sicht einer mikrosoziologischen Praxeologie sind dagegen Formen der leiblichen sozialen Koordination (‚Interkorporealität') ebenso interessant, bei denen Sprache keine Rolle spielt (Schmitt 2015; Meyer, Streeck & Jordan 2017).

5.2 Zeitlichkeit: Sequenzialität und Simultaneität

Soziale Interaktionen sind Vollzugswirklichkeiten (Bergmann 1985). Es war die grundlegende Erkenntnis der Konversationsanalyse, dass diese durch die sequenzielle Organisation von Handlungen hergestellt werden (Schegloff 2007). Sowohl die Bedeutung als auch die Konstruktion (*composition*) eines Turns in der Interaktion ist durch seine Position (*position*) im sequenziellen Ablauf einer Interaktion bestimmt (vgl. Clift 2016). Jeder Turn ist auf einen vorangehenden Kontext zugeschnitten (*context-sensitive*); zugleich schafft er einen neuen Kontext (*context-renewing*; Heritage 1984: 242). Zur Schaffung neuer Kontexte gehört die Stiftung von Projektionen, d. h. die (normative) Erwartung an Turn-

5 Ausnahmen sind beispielsweise das Einüben von Sprechtexten in einer Theaterprobe (Schmidt 2014) oder das Stimmtraining eines Moderators vor Sendebeginn.

Fortsetzungen oder nächste Handlungen des Adressaten (Auer 2005). Für die Linguistik hat diese retrospektiv-prospektive Orientierung des Handelns (siehe auch Deppermann & Günthner 2015) die Konsequenz, dass grammatische Praktiken und die Herstellung von Bedeutung temporalisiert zu denken sind.

Die multimodale Perspektive erweitert den Blick auf Zeitlichkeit. Leibliches Handeln wird nicht nur sequenziell konstituiert; simultane Verhältnisse spielen eine ebenso fundamentale Rolle. Wir sahen dies bereits in unserem Beispiel aus Ausschnitt (1): Blick, Zeigegeste und Sprechen werden vom Referierenden simultan eingesetzt. Der Adressat reagiert bereits während des referierenden Turns mit *gaze-following* und zeigt noch vor seinem Ende an, dass die Suche erfolgreich war.

In der konversationsanalytischen Literatur kommt Simultaneität nur im Falle der Überlappung von Sprecherbeiträgen (*overlap*) ins Spiel (Sacks, Schegloff & Jefferson 1974; Jefferson 2004). Simultane Ereignisse sind aus dieser Sicht die Ausnahme und ein „Unfall", in dem das Prinzip des Rederechts (*floor*), des „one speaker at a time" (Sacks, Schegloff & Jefferson 1974: 700) verletzt wird. In der leiblichen Interaktion ist aber der simultane Einsatz mehrerer Ressourcen bei mehreren Beteiligten nicht die Ausnahme, sondern unvermeidlich. Das Konzept der Überlappung ist nicht brauchbar, um diese simultanen Verhältnisse zu beschreiben (Schmitt 2005).

Interaktivität beginnt nicht erst nach einem Sprecherwechsel. In *face-to-face*-Interaktion ist der Turn selbst eine mehr oder weniger interaktive Konstruktion.[6] SprecherInnen beobachten die Reaktionen ihrer GesprächspartnerInnen während ihres Turns (C. Goodwin 1979; M. H. Goodwin 1980) und passen die Fortsetzung der Produktion ihres eigenen Turns an diese Reaktionen an.

Die simultanen Reaktionen der Adressaten stellen eine mikrosequenzielle, interaktive Responsivität innerhalb einzelner Turns bzw. Handlungen dar. Handlungen werden nicht *en bloc* vollzogen und dann erst der Reaktion des anderen überantwortet.[7]

---

6 Tatsächlich existiert diese Form der Interaktivität in räumlich getrennter, computervermittelter Interaktion, etwa beim Chat, nicht (siehe Beißwenger 2007). Die simultane interaktive Rückkopplung ist auch bei Kopräsenz beeinträchtigt, wenn die Beteiligten bspw. eine *side-by-side*-Position einnehmen und daher nur ein eingeschränktes geteiltes Gesichtsfeld haben oder wenn kein Blickkontakt besteht (Oloff i. Dr.).

7 Dies ist allerdings bei den meisten Formen der computervermittelten Interaktion so – mit gravierenden Konsequenzen für Beitragskonstruktion und Verstehenssicherung (Beißwenger 2007).

5.3 Interaktiv Beteiligte statt Sprecher vs. Hörer

Für Linguistik und Kommunikationswissenschaft ist die Sprecher-Hörer-Beziehung die Grundeinheit des kommunikativen Austauschs. Während dieses Modell sicher für einen rein akustisch vermittelten Austausch (wie durchs Telefon) geeignet ist, reicht es für leibliches Interagieren unter der Bedingung von Kopräsenz nicht aus. Die Reduktion der Beteiligten auf Sprecher und Hörer ist irreführend, da mit ihr die konstitutive Rolle anderer Handlungsressourcen übersehen wird. Auch „verbal abstinente" InteraktionsteilnehmerInnen (Schmitt 2012) tragen durch Mimik, Blick, Nicken, Körperpositur etc. zur Konstitution des interaktiven Handelns bei (5.2). Hörer sind oftmals viel mehr als nur das, ebenso wie Sprecher meist Weiteres tun als nur zu sprechen. Als alternative Grundkategorie zu ‚Sprecher-Hörer' bietet sich die Kategorie der ‚Beteiligten' (*participants*) an, die bereits Goffman (1979) vorgeschlagen hatte. Während er hier an die Weisen dachte, in denen Menschen an der Produktion einer verbalen Äußerung beteiligt sein (z. B. als *animator*, *author* oder *sponsor*) und als Rezipienten an einem Kommunikationsereignis teilnehmen können (z. B. als *ratified* oder *non-ratified participant*, als *overhearer* oder *eavesdropper*), führt eine Dynamisierung des Konzepts der Beteiligung (siehe dazu Goodwin & Goodwin 2004) unweigerlich dazu, auch die leiblich-räumliche Beteiligung als grundlegende Dimensionen anzuerkennen (Deppermann & Schmitt 2007).

5.4 Raum

Hausendorf & Schmitt (in diesem Band) führen aus, dass Raum in wenigstens drei Hinsichten in Interaktionen unter Bedingungen von Kopräsenz ins Spiel kommt:

- als mit Bezug auf die Potenziale des Handelnden bestehende Vorgabe (‚Interaktionsarchitektur'), die bestimmte Möglichkeiten und Grenzen der Raumnutzung bereithält (wie z. B. Begehbarkeit, Sichtbarkeit, mögliche Sitzordnungen, vgl. den Begriff der *affordances*, Gibson 1979),
- als übliche bzw. sozial sanktionierte Formen der Raumnutzung (‚Sozialtopographie', z. B. welche Mitglieder welcher sozialen Kategorien welche Raumabschnitte für welche Handlungen benutzen dürfen) und
- als Ressource, die Interaktionsteilnehmer für den Vollzug ihrer Handlungen nutzen können (siehe auch Hausendorf, Mondada & Schmitt 2012).

Der für die Beteiligten *hic et nunc* relevante Interaktionsraum muss interaktiv immer wieder aufs Neue hergestellt werden, beispielsweise durch die Körperausrichtung zueinander und zu relevanten Objekten, Raumsegmenten und

Richtungen (Hindmarsh & Heath 2000; Mondada 2007a, 2009). Dies hat zur Konsequenz, dass die Bühler'sche ich-hier-jetzt-Origo (Bühler [1934] 1982) durch die raumzeitliche Positionierung des Handelnden allein für den erfolgreichen Handlungsvollzug noch nicht hinreichend bestimmt ist. Sie bedarf zusätzlich der räumlichen Ko-Orientierung (*joint attention*) mit dem Gesprächspartner, die reflexiv selbst durch Handlungen herzustellen ist.

5.5 Objekte

Objekte spielen in fast jeder Interaktion eine Rolle – selbst in verbal dominierten Tischgesprächen werden Tassen, Besteck, Smartphones usw. häufig interaktiv relevant (z. B. Hoey 2015). Umso wichtiger sind sie in empraktischen Interaktionen, wie in den *workplace studies*, die die multimodale Interaktion in Arbeitskontexten untersuchen (Heath & Luff 2000, 2013). Ähnlich wie dies in 5.4 für den Raum festgestellt wurde, kommen Objekte als physikalisch-materiale Bedingungen, als mit sozial bestimmten üblichen und erwartbaren Nutzungsweisen assoziierte Gegenstände und als in der konkreten Interaktion in bestimmter Weise genutzte Ressourcen ins Spiel. Die interaktive Nutzung entspricht dabei häufig nicht der konventionellen; z. B. werden Stifte zum Zeigen oder zur Beanspruchung des Rederechts benutzt (Mondada 2007b), das Absetzen einer Tasse signalisiert das Ende eines Gesprächsthemas (Mondada 2015) oder das Aufstellen eines Ordners, dass eine Besprechungspause zu beenden sei (Deppermann, Schmitt & Mondada 2010). Viele Objekte haben selbst semiotische Qualitäten. Über Objekte, die Kommunikationsmedien sind, wird die Interaktion unter den Beteiligten vor Ort mit mediatisierter Interaktion mit Nichtanwesenden verknüpft (Heath & Hindmarsh 2000).

Objektbezogene Praktiken finden häufig im Kontext professioneller Routinen mit funktionsrollenspezifischen Zuständigkeiten statt (Schmitt & Deppermann 2007). Sie erfordern professionelle Wissensbestände und Weisen des Wahrnehmens, eine *professional vision* (C. Goodwin 1994, 1997) für die schnelle, korrekte und relevante Eigenschaften erkennende Identifikation und Kategorisierung von Objekten. Die spezifische professionelle Perspektive manifestiert sich auch sprachlich, z. B. in Praktiken der Objektreferenz (Hindmarsh & Heath 2000), für die oft professionsspezifisches Vokabular benutzt wird. In unserem Eingangsbeispiel kürzen die Rettungssanitäter beispielsweise die Ringerlösung als „ringer" ab.

5.6 Multiaktivität

In vielen Interaktionen bearbeiten die Interaktionsteilnehmer nicht nur einen, sondern mehrere Aktivitätsstränge parallel. Dies wird als ‚Multiaktivität' (*multiactivity*, Haddington et al. 2014) bezeichnet. *Multi-* bezieht sich dabei auf verschiedene Handlungen, nicht auf den Einsatz verschiedener multimodaler Ressourcen für eine Handlung. So wäre z. B. die Koordination von Blick, Gestik und Sprache bei der Raumreferenz keine Multiaktivität, denn alle drei Ressourcen sind hier im Dienste der Handlung des Referierens. Wenn der Akteur aber gleichzeitig ein Auto steuert, haben wir es mit Multiaktivität zu tun. In unserem Eingangsbeispiel etwa findet die Suche nach der Ringerlösung statt, während der Einsatzleiter und sein Assistent einen Verband am Patienten anbringen. Multiaktivität kann verschieden organisiert sein: Aktivitäten können simultan erfolgen; für den einzelnen Beteiligten bedeutet dies, dass er den beiden Aktivitäten jeweils unterschiedliche Ressourcen widmen muss, z. B. spricht der Einsatzleiter über die Ringerlösung, hört AS1 zu und schaut auf dessen Gesten, während er manuell mit dem Verband beschäftigt ist. Längere Phasen simultaner Bearbeitungen sind selten und nur möglich, wenn eine der Aktivtäten routiniert durchzuführen ist und wenig Aufmerksamkeit beansprucht (Deppermann 2014). Häufig kommt es deshalb zum schnellen Wechsel zwischen Aktivitäten (Mondada 2014c). Dabei wird die eine Aktivität zugunsten der anderen oft in einer Weise suspendiert, die verdeutlicht, dass sie bald wieder aufgenommen wird, z. B. durch das Einfrieren von Gesten (Deppermann 2014) oder Körperdrehungen (*body torque*, Schegloff 1998), wobei die Ausrichtung des Rumpfs die längerfristige und weiterhin gültige, aber momentan unterbrochene Orientierung verkörpert, die Ausrichtung des Oberkörpers dagegen die kurzfristige Aufmerksamkeit für einen anderen Fokus anzeigt.

5.7 Bewegung

Noch recht neu ist die Erforschung der sozialen Interaktion im Kontext der Bewegung im Raum (Haddington, Mondada & Nevile 2013). Gemeint sind hier nicht-ortsstabile Interaktionen, bei denen einzelne oder alle Beteiligte in Bewegung sind. Interaktionen in Bewegung sind besonders komplex, da sich hier mit räumlichen Veränderungen permanent neue Aufgaben der räumlichen Koordination der Beteiligten ergeben, die auch mit Risiken verbunden sein können (z. B. im Verkehr). Da die Bewegung im Raum meist einen eigenen Handlungsstrang ausmacht, sind Interaktionen in Bewegung oft Multiaktivität. Die Art und Weise des Gehens erweist sich dabei etwa als kommunikative Ressource, mit der z. B. der Typ der Interaktion (aufgabenbezogen oder *small talk*,

Schmitt & Deppermann 2010), die Intensität und Dauer des Kontakts (kurzes Grüßen vs. längeres Gespräch) oder die für eine kommende Handlung vorgesehene Beteiligungsstruktur (private Antwort an ein Individuum vs. gruppenöffentliche Information, Mondada i. Dr.) projiziert wird. Die Kommunikation im Auto ist durch die besonderen räumlichen Sitz- und damit auch Seh- und Hörverhältnisse (nicht *face-to-face*, sondern *side-by-side* und *front-to-back*) beeinträchtigt sowie durch die Notwendigkeit, die Interaktion zwischen den Insassen mit den Anforderungen des Fahrens zu koordinieren, die oft durch unvorhersehbare Ereignisse und die Notwendigkeit zur raschen Reaktion geprägt sind (Haddington, Keisanen & Nevile 2012; Deppermann i. Dr.).

6 Sprachliche Praktiken in der multimodalen Interaktion

Schauen wir uns nun genauer an, wie sich die im vorigen Unterkapitel angesprochenen Eigenschaften multimodaler Interaktion auf die Verwendung von Sprache auswirken. Der Zusammenhang zwischen Sprache und Leiblichkeit ist wechselseitig: Leibliche Aktivitäten ermöglichen, erfordern, bedingen und beschränken sprachliche Praktiken – das Umgekehrte gilt ebenso. Der Schwerpunkt unserer Betrachtung wird auf der erstgenannten Wirkungsrichtung liegen. Der Einfluss des Leibes auf sprachliche Praktiken besteht in verschiedenen Hinsichten:

- Durch die Koordination mit anderen Ressourcen werden Bedeutung und Funktion sprachlicher Strukturen vereindeutigt (Unterkapitel 6.1);
- die Erweiterung sprachlicher Strukturen reagiert auf Rezipientenreaktionen, die simultan zur Turn-Produktion stattfinden (Unterkapitel 6.2);
- die Komplexität sprachlicher Strukturen hängt von der multimodalen Ko-Orientierung der Beteiligten ab (Unterkapitel 6.3).

6.1 Disambiguierung: Verschiedene *okay*s

Dass bestimmte Facetten von Bedeutung und Funktion sprachlicher Strukturen fast immer kontextabhängig sind, ist eine linguistische Binsenweisheit. Während der Einfluss kotextueller, kollokationaler oder konstruktionaler Faktoren eingehend untersucht wurde, weiß man weitaus weniger über die Rolle multimodaler Faktoren. Offensichtlich ist diese schon lange im Bereich von Deixis und referenzieller Bedeutung (z. B. Bühler [1934] 1982; Lyons 1983). Die lokale Bedeutung deiktischer Formen und definiter Nominalphrasen in der Redesitua-

tion hängt häufig von konkreten räumlichen Zeigzielen, die via Körperausrichtung, Geste und Blick verfügbar gemacht werden, ab (vgl. Unterkapitel 5.1). Wir wenden uns nun einem weniger offensichtlichen Fall zu, der Funktion von Gesprächspartikeln. Manche Gesprächspartikeln werden überaus vielfältig verwendet (Schwitalla 2002). Es ist bekannt, dass die Prosodie eine unterscheidende Rolle spielt (Ehlich 1986; Schmidt 2001). Doch auch gleiche prosodische Varianten können unterschiedliche Funktionen haben, je nach Koordination mit anderen multimodalen Ressourcen und abhängig von ihrer Position in einer Interaktionssequenz. Exemplarisch sei dies an zwei Verwendungen der Partikel *okay* gezeigt: zur Verstehensdokumentation vs. zum Abschluss einer Interaktionssequenz bzw. eines Themas.

Wie in anderen Sprachen wird *okay* im Deutschen nicht nur zur Signalisierung von Einverständnis, sondern für weitere interaktive Funktionen verwendet (z. B. Beach 1993). Eine davon ist die Verwendung von *okay* um anzuzeigen, dass eine eben erhaltene neue Information hinreichend für die gegenwärtigen Verstehensbedürfnisse des *okay*-Sprechers ist. Dies geschieht oft in dritter Position einer Frage-Antwort-Sequenz: A hatte B eine Frage gestellt, Bs Antwort wird von A mit *okay* als hinreichend ratifiziert. In Ausschnitt (2) beginnt AS ihre Selbstdarstellung in einem sogenannten „WG-Casting“, in dem eine Wohngemeinschaft BewerberInnen um ein freies Zimmer zum Kennenlernen einlädt. Bewerberin AS wird in Zeile 03 von Bewohnerin SL mit der Frage nach ihrem Namen unterbrochen:

(2) FOLK_E_00251_SE_01_T_01, c71, 00:54–01:04: WG-Casting

```
01 AS:   also ich bin äh zweiunzwanzig jahre alt;#
                                                  #Abb.3.4
02       °h (.) [ich         ]
03 SL:         *[du bis jetzt] +DIE?#
   sl:         *Blick->AS----------------->
   as:   >>Blick geradeaus-----+Blick->SL->
                                  #Abb.3.5
04 AS:   (.) AN§na.=*
   sl:   ---------->*Blick->Tasse->
   sl:         §nickt------------>
05 SL:   =<<t>o§K+#AY.>*
   sl:   ------------>*
   sl:   ----->§
   as:   ------->+Blick geradeaus--->>
                  #Abb.3.6
06       (0.6)
```

07 **AS:** **ich HAB *äh; (.)**
sl: ------->*Blick->AS--->>
08 **AS:** **ich studier grade bisher BIologie und gEographie auf lehr-amt,**

Abb. 3.4: AS schaut nach vorn. SL schaut vor sich hin.

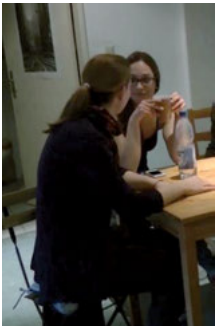

Abb. 3.5: SL hat AS nach ihrem Namen gefragt. SL und AS schauen einander an.

Abb. 3.6: SL sagt: „okay.“, schaut auf ihre Tasse. AS schaut wieder geradeaus.

SL unterbricht AS' Selbstdarstellung mit der Frage nach AS' Namen und schaut sie dabei an. AS wendet SL während der Frage auch den Blick zu (Z. 03, Abb. 3.5). Als AS mit ihrem Vornamen geantwortet hat (Z. 04), schaut SL wieder vor sich auf ihre Tasse und quittiert die Antwort mit einem schnell angeschlossenen „okay“ mit fallender Intonation (Z. 05, Abb. 3.6). Schon während der zweiten Silbe von AS' Namen („anna“, Z. 04) beginnt SL zu nicken. Sie nickt aufwärts. Die Abwärtsbewegung des Nickens endet auf der ersten Silbe von „okay“; gleichzeitig wendet sie ihren Blick von AS ab.

Die multimodale Gestalt (Mondada 2014a, 2016a) dieser Praktik der Dokumentation hinreichenden Verstehens besteht darin, dass *okay* mit fallender Intonation mit Aufwärtsnicken kombiniert wird, wobei nach dem *okay* der Blick vom Gesprächspartner abgewandt wird. Das *okay* fungiert hier als Abschluss (*sequence closing third*, Schegloff 2007: 186) einer Insertionssequenz (Jefferson 1972) und dient als *continuer* (Schegloff 1982), der das Rederecht weiter der Erzählerin zuweist.

Eine andere Praktik der Verwendung von *okay* mit fallender Intonation sehen wir dagegen in (3) aus einem Bewerbungstraining. Bewerber TB erzählte, dass er bei einer früheren Anstellung mit der Entlohnung unzufrieden war. Der Ausschnitt beginnt mit der Schilderung seiner daraus gezogenen Konsequenz, diesen Arbeitgeber zu verlassen (was dem Bewerbungstrainer TN bereits aus

TBs Unterlagen, die vor ihm auf dem Tisch liegen, bekannt war). TB projiziert, dass er die Erzählung dieser Etappe seiner Berufsbiographie zu Ende führt, mit einer Redewiedergabe „nee dann“, nach der er eine lange Pause (1.4 Sekunden) macht, um dann doch noch den Satz zu beenden.

(3) FOLK_E_00173_SE_01_T_01, c578, 13:49–13:59: Bewerbungstraining

```
01 TN:                  [°hh (.)ja;                          ]
02 TB:   und das war DA[mals der entschEIdende pUnkt weshalb i]ch
         gsagt hab;
03       NEE dann,
04       *   (0.8)  *(0.2)#+*(0.4)+
   tb:   *nickt tief*
   tn:                    +nickt-+
   tn:   >>Blick->TB--------*Blick->Unterlagen--->>
                         #Abb.3.7
05 TB:   +%geh ich+ den ANdern%& [weg.  ]
06 TN:                            [oKAY;>]#
   tn:   +nickt 2x+
   tn:    %...................%Oberkörper aufrecht--->>
   tn:                         &.......................
                                          #Abb.3.8
07 TN:   (.) mHM,&#
   tn:   ........&blättert in Unterlagen--->>
                 #Abb.3.9
08       (0.8)
09 TN:   °h gut;=
10       =dann sind se zu astellas geGANgen,
11       astellas
12 TB:   geNAU;
```

Abb. 3.7: TN vorgebeugt, schaut auf TB.

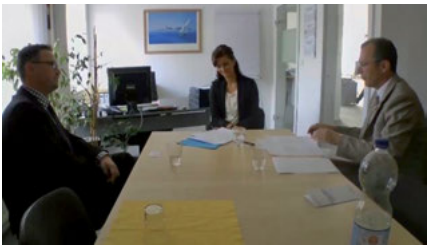

Abb. 3.8: TN richtet Oberkörper auf, schaut auf Unterlagen, sagt: „okay;“.

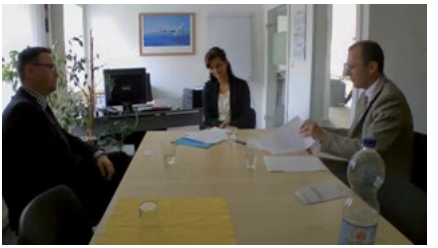

Abb. 3.9: TN aufrechte Körperpositur, blättert und schaut in Unterlagen, sagt: „mhm".

Als TB während seiner Turn-internen Pause tief nickt, antwortet TN seinerseits mit Nicken (Z. 04). Im Unterschied zu (2) beginnt es nicht mit der Aufwärts-, sondern der Abwärtsbewegung. Gleichzeitig wendet TN den Blick von TB ab, den er bis dahin während dessen Erzählung angeschaut hatte (vgl. Abb. 3.7), und schaut in TBs Unterlagen. Er richtet seinen bis dahin gegen TB vorgeneigten Oberkörper auf (Abb. 3.8). Mit Blickabwendung, Veränderung der Körperpositur und Zuwendung zu den Unterlagen wird von TN die Erzählung bereits ab Zeile 04 als verstanden und abgeschlossen behandelt, obwohl TB erst in Zeile 05 seinen abschließenden Turn vervollständigt. Dies quittiert TN mit einem „okay" mit fallender Intonation, wobei er sich weiterhin den Unterlagen zuwendet (Z. 06, Abb. 3.8). Anschließend produziert TN eine weitere abschließende Partikel „mhm." (Z. 07) und leitet mit „gut," (Z. 09) die Transition zur nächsten berufsbiographischen Etappe von TB ein. Währenddessen bleibt er aufrecht und schaut in TBs Unterlagen (Abb. 3.9).

Die multimodale Gestalt der Verwendung von *okay* zur Signalisierung eines thematischen bzw. handlungssequenzbezogenen Abschlusses unterscheidet sich von der Signalisierung hinreichenden Verstehens in einigen Punkten. Zwar wird auch hier *okay* mit fallender Intonation produziert, doch geht das Nicken als eigene Handlung (und nicht als gleichzeitige Aktivität) dem *okay* mit deutlichem Abstand voraus. *Okay* selbst wird nicht von Nicken begleitet. Die Veränderung der Körperpositur zeigt *disengagement* mit Bezug auf die noch laufende Erzählung von TB an, während der Blick in die Unterlagen projiziert, dass sich der *okay*-Sprecher einem neuen Aufmerksamkeitsfokus zuwendet.

Der Vergleich der *okay*-Verwendungen zeigt, dass auch prosodisch gleiche Formen von Gesprächspartikeln interaktiv unterschiedlich verwendet werden, wenn sie mit anderen leiblichen Aktivitäten koordiniert werden (unveränderte vs. disengagierte Körperpositur; unspezifische Blickabwendung vs. Blick auf Objekt, das einen nächsten Themenkomplex verkörpert) und wenn die gleichen Aktivitäten (hier: Nicken) in einem anderen zeitlichen Verhältnis zu *okay* produziert werden (zeitgleich mit *okay* vs. als eigene, dem *okay* vorangehende

Handlung). Außerdem scheint ein systematischer Unterschied zwischen Auf- und Abwärtsnicken zu bestehen: Bei der Dokumentation hinreichenden Verstehens einer neuen Information beginnt das Nicken mit der Aufwärtsrichtung (2); dagegen wird mit dem Abwärtsnicken das Verstehen einer vom Gesprächspartner nicht vollständig ausgeführten, aber schon bekannten Information quittiert (3). Während Aufwärtsnicken also den Gewinn neuer Information anzeigt, scheint Abwärtsnicken das Verstehen von schon Bekanntem bzw. Erwartetem anzuzeigen.

6.2 Syntaktische Komplexität als Reaktion auf (ausbleibende) Rezipientenreaktionen

Einer der wirkungsmächtigsten frühen Aufsätze zur multimodalen Interaktion trug den Titel *The interactive construction of a sentence in natural conversation* (C. Goodwin 1979). Charles Goodwin zeigte in ihm, wie ein Sprecher einen Satz immer weiter expandierte, während er nacheinander unterschiedliche Adressaten anschaute. Dies wurde notwendig, da ihn der erstgewählte nicht ansah und somit die Aufmerksamkeit des zunächst intendierten Adressaten nicht gesichert war. Mit dem Wechsel der Adressaten hatte der Sprecher aber unterschiedliche Vorwissensbestände in Rechnung zu stellen, was ihn veranlasste, weitere Informationen nachzutragen und so den Satz mit jedem Adressatenwechsel zu verlängern. Goodwin zeigte, wie syntaktische Strukturen auf Basis von simultan zur Turn-Produktion stattfindenden bzw. ausbleibenden Rezipientenreaktionen emergent entstehen. Beobachtung und Analyse der Reaktionen von Rezipienten sind eine wesentliche Quelle für Turn-Expansionen, im Deutschen zumal für Nachfeldrealisierungen jenseits eines möglichen (ersten) syntaktischen Abschlusspunktes (Auer 1996; Imo 2015; Proske 2015). Ein Beispiel dafür ist Ausschnitt (4) aus einer Deutschstunde. Der Lehrer stellt eine Frage zu einem Gedicht, ein Schüler nach dem anderen wendet während der Produktion der Frage den Blick vom Lehrer ab.

(4) FOLK_E_00124_SE_01_T_01, c617, 10:43–10:50: Deutschunterricht im Wirtschaftsgymnasium

```
01 LE:   #wer traut§ sich_s ZU§ #ne inhAlts$angabe $#zu machen;
   s1/2:  Blick->LE§..... §Blick->unten----------------------->>
   s3:    Blick->LE----------------------$.......$Blick->unten--->
   s4:    Blick->LE----------------------$.......$Blick->unten-->>
          #Abb.3.10             #Abb.3.11           #Abb.3.12
```

Abb. 3.10: Zu Beginn der Lehrerfrage: 5 Schüler schauen zum Lehrer.

Abb. 3.11: Lehrer sagt: „traut". S1 und S2 schauen nach unten.

```
02          (1.2)
03 LE:      $von diesem geDICHT;
   s3:      $Blick->LE--------->
04              (0.5) $&(0.6)$&#(1.9)
   s3:      ----------$......$Blick->unten--->>
   s5:      >>Blick->LE&......&Blick->S2----->>
                                  #Abb.3.13
```

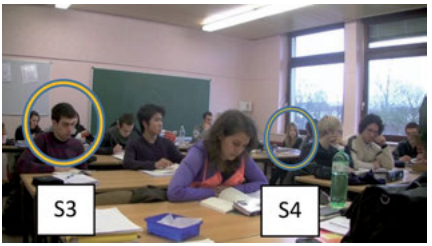

Abb. 3.12: Lehrer sagt: „inhaltsangabe", auch S3 und S4 schauen nach unten.

Abb. 3.13: Nach Turn-Ende des Lehrers: S5 schaut zur Seite; kein Schüler blickt zum Lehrer.

Unmittelbar vor Beginn des Turns des Lehrers schauen fünf der Schüler zum Lehrer (Abb. 3.10). Der Rest schaut auf den Text, einer zum Nachbarn. Als der Lehrer „wer traut sich_s zu" formuliert hat (Z. 01), wenden S1 und S2 den Blick vom Lehrer ab und schauen auf ihren Text (Abb. 3.11). An dieser Stelle ist klar, dass der Lehrer eine Frage stellt, die nicht leicht zu beantworten ist. Als der Lehrer die Aufgabe mit „ne inhaltsangabe" spezifiziert hat, wenden auch S3 und S4 ihren Blick vom Lehrer ab und schauen auf den Text (Abb. 3.12). Mit der Blickabwendung zeigen die Schüler, dass sie nicht für eine Antwort verfügbar sind, allerdings in einer schulisch legitimierten Weise, da sie demonstrieren, sich mit dem Text zu befassen. In der sich an seine Aufgabenstellung anschließende Pause von 1.2 Sekunden schaut der Lehrer in die Klasse. Niemand

meldet sich. Er ergänzt nun die Präpositionalphrase „von diesem gedicht“ (Z. 03) und erneuert damit die Aufforderung, ein Antwortangebot zu geben. Die syntaktische Expansion führt zu einer Nachfeldbesetzung („wer traut sich_s zu ne inhaltsangabe zu machen von diesem gedicht“). Diese Erweiterung der syntaktischen Struktur ist also interaktiv durch ausbleibende Antwortangebote und die Anzeige fast aller Schüler (bis auf einen), nicht zur Interaktion mit dem Lehrer zur Verfügung zu stehen, motiviert. Die Satzexpansion bringt den Satz zu einem nächsten möglichen syntaktischen Abschlusspunkt und damit zu einer *re-completion* des Turns und einer weiteren Redeübergabestelle (Auer 1996), wodurch den Schülern erneut die Aufgabe, Antwortangebote abzugeben, zugewiesen wird. Allerdings führt dies nur dazu, dass nach einer weiteren Pause S3, der zwischenzeitlich wieder zum Lehrer geschaut hatte (Z. 03), und auch noch der letzte verbliebene Schüler S5 den Blick vom Lehrer abwenden (Abb. 3.13; vgl. Schmitt 2004).

Syntaktische Expansionen sind also ein flexibles Instrument, um leibliche Rezipientenreaktionen zu behandeln, die parallel zur laufenden Turn-Produktion Unverständnis, Ablehnung, Überraschung, fehlende Aufmerksamkeit, Nicht-Bereitschaft den Turn zu übernehmen usw. anzeigen. Die Expansion kann Informationen nachliefern, die die Redeintention präzisieren, eine strittige Position abschwächen, eine Begründung liefern usw. und damit das Problem behandeln, welches die Rezipientenreaktion angedeutet hatte, ohne dass dieses zum expliziten Verhandlungsgegenstand werden muss. Die mikrosequenzielle Adaptation der Turn-Expansion ist gleichzeitig ökonomisch und problemvorbeugend. Sie wirkt möglichem Dissens, Un- oder Missverständnis und Disaffiliation entgegen, sobald deren Entstehung abzusehen ist.

6.3 Argumentrealisierung in Abhängigkeit von multimodaler Ko-Orientierung

Die Art und Weise, in der die Argumente eines Verbs realisiert werden, hängt von ihrer (vermutlichen) Zugänglichkeit (*accessibility*) für den Adressaten ab (Ariel 1990). Ob auf einen Referenten mit einer indefiniten oder einer definiten Nominalphrase, einer Periphrase oder einer Abkürzung, einem Pronomen oder ganz ohne overte Argumentrealisierung Bezug genommen wird, ist einerseits vom geteilten Wissen der Beteiligten, andererseits von der Vorerwähntheit des Referenten in der vorangegangenen Gesprächssequenz und seiner visuellen Verfügbarkeit abhängig („referential marking scale“, Ariel 2008: 44–52). Dabei wurde allerdings in der linguistischen Forschung, wie in Unterkapitel 5.4 angesprochen, vernachlässigt, dass die Interaktionsteilnehmer für erfolgreiche lo-

kale Referenz einen geteilten Interaktionsraum und geteilte Aufmerksamkeit durch Interaktion selbst herstellen müssen. In Ausschnitt (1) aus der Rettungsübung können wir sehen, wie die gemeinsame leibliche Ko-Orientierung auf intendierte Objekte es ermöglicht, sprachliche Referenzen minimal zu gestalten. Dazu erweitern wir das eingangs gezeigte Transkript um eine visuelle Transkription und Standbilder, die deutlich machen, wie die referierenden Handlungen leiblich organisiert sind.

(1) FOLK_E_00138_SE_01_T_01, c842, 11:12–11:28: Rettungssanitäterübung

```
01 EL:   %‡aber (.) wo_s jetz die #RINger-
   el:   %>>Blick rechts---------------->
   as1:   ‡>>packt Strip aus------------>
                                  #Abb.3.14
```

Abb. 3.14: EL fragt nach Ringerlösung und sucht sie.

In Zeile 01 (Abb. 3.14) sucht der Einsatzleiter (EL) nach der Ringerlösung. Die Abkürzung „ringer“ und der bestimmte Artikel verweisen auf das geteilte professionelle Wissen und die Vertrautheit der Rettungssanitäter mit dem Referenten. Die Ringerlösung ist aber weder im visuellen Fokus des Sprechers noch der Adressaten. EL hatte bereits knapp sechs Minuten zuvor den Assistenten AS2 aufgefordert, die Lösung vorzubereiten. Aus dem Aktivitätszusammenhang der Suche ist außerdem klar, dass „die ringer“ sich nicht auf die Lösung als solche, sondern auf die Flasche, in der sie sich befindet, bezieht. In seiner Antwort lokalisiert AS1 die Flasche mit der Ringerlösung mit einer Zeigegeste. Wie EL benutzt er die abkürzende definite Nominalphrase „d_ringer“ (vgl. Unterkapitel 5.1).

```
02 AS1:  +(0.5) d_%RINger %liegt #da*neben +*%da#+*%[unten.]
   as1:  +Blick->Ringer--------------------+.....+schaut auf
         Strips-->
   as1:                                *.......*zeigt*,,,
03 EL:                                                [ah-   ]
   el:   gradeaus-%.......%Blick->links------%.....%Blick->
         Ringer---->
                                  #Abb.3.2       #Abb.3.3
```

Während der Suche nach der Ringerlösung hatte AS1 einen Strip für den Verband vorbereitet, den EL am rechten Arm der Patientin anlegt (EL fixierte den Arm während seiner Suche nach der Ringerlösung). Er reicht EL nun den Strip mithilfe einer elliptischen Präsentationsformel:

```
04       (1.0)‡(0.4)
   as1:  ---->‡reicht EL Strip--->>
05 AS1:  dort JETZT,#
                    #Abb.3.15
```

Abb. 3.15: AS1 reicht EL den Strip mit der rechten Hand und sagt: „dort jetzt“.

AS1 und EL hatten bereits vor der Suche nach der Ringerlösung begonnen, gemeinsam am Verband für die Patientin zu arbeiten. Dieses *joint project* (vgl. Clark 1996: 191–220) ist noch nicht abgeschlossen und weiterhin für beide handlungsleitend; die Suche nach der Ringerlösung war nur zeitweilig als weitere Aktivität eingebettet worden (vgl. Mondada 2014c). Sowohl AS1 als auch EL zeigten ihre fortwährende Orientierung an der Aufgabe, den Patienten zu verbinden, durch ihr leibliches Handeln an: AS1 bereitete den Strip vor und EL fixierte den Arm des Patienten in der Verbandsposition (Abb. 3.14). Da dieses *joint project* für

beide weiterhin salient ist und AS1 den Strip in ELs Gesichtsfeld hält (Abb. 3.15), ist die unspezifische sprachliche Referenz auf den Strip („dort“) ausreichend. Für die Beteiligten ist deutlich, dass „dort jetzt“ sich nicht auf das Topik der vorangegangenen Turns, die Ringerlösung, bezieht, obwohl dies aus einer rein sprachlichen Kohärenzperspektive, die hier analeptische Topikkontinuität erwarten lassen würde (vgl. Helmer 2016), naheliegender wäre.

Im nächsten Turn dagegen bezieht sich EL wieder auf die Ringerlösung, während er nach ihr greift.

```
06       (2.6)
07 EL:   §also- (.) halt #mal,=
   el:   §.....................
                        #Abb.3.16
08 AS2:  =DRAN §machen oder nich?=
   el:   ......§greift Ringer und gibt sie AS2--->
09 EL:   =$und halt mal §#$HOCH und guck ob sie läuft.§
   el:   -------------->§.............................§nimmt
         Strip->>
   as2:   $..............$nimmt Ringer und hält sie hoch--------->>
                        #Abb.3.17
```

Abb. 3.16: EL greift nach Ringerlösung und sagt: „halt mal“.

Abb. 3.17: EL übergibt AS2 die Ringerlösung und sagt: „halt mal hoch“.

In den Zeilen 07–09 verweisen EL und AS2 drei Mal elliptisch auf die Flasche mit der Ringerlösung: „halt mal“ – „dran machen“ – „halt mal hoch“. Zusammen mit der vorangegangen Suche (Z. 01–03), die die Ringerlösung bereits als ein Objekt, das für eine nächste Handlung relevant ist, salient gemacht hatte, reichen die Greifgeste und der Blick ELs auf die Flasche aus (Abb. 3.16), um

AS2 mit einer elliptischen Äußerung, ohne Enkodierung des Objekts *Ringerlösung*, die Identifikation des gemeinten Referenten zu ermöglichen. In Zinken und Deppermann (2017) konnten wir feststellen, dass Imperative von (di)transitiven Verben wie *geben*, *nehmen*, *halten* ohne enkodiertes Objekt benutzt werden, wenn der Adressat bereits auf die auszuführende Handlung vororientiert ist (d. h. sie erwartet bzw. schon begonnen hat) und sich das Objekt in seinem Gesichtsfeld befindet. So auch hier: AS2 schaut auf die Flasche mit der Ringerlösung und er ist derjenige, der dafür zuständig ist, sie an der Vene an einem Arm der Patientin anzulegen. (AS2 hatte die Lösung bereits sechs Minuten zuvor auf Bitte von EL vorbereitet.) Die Aufforderung („halt mal" bzw. „halt mal hoch", Z. 09, Abb. 3.17) ist hier also erwartbar für AS2: Sie betrifft seine Verantwortlichkeit im *joint project* des Erste-Hilfe-Einsatzes. Aufgrund des geteilten professionellen Wissens über die Verwendung der Ringerlösung reicht es für AS2 in Zeile 08 aus, das örtliche Ziel der Anbringung, den Venenkatheter am Arm der Patientin, nur pronominal („dran") zu formulieren. Dieses Ziel ist aber im Moment nicht im gemeinsamen Aufmerksamkeitsfokus (und in der Tat wird später verhandelt, welcher Arm benutzt werden soll).

Auch in diesem Abschnitt bestimmt sich die konkrete, situativ relevante Semantik im Zusammenspiel von sprachlichen Äußerungen, Gesten, visuell verfügbaren Gegenständen und geteilten professionellen Routinen (vgl. C. Goodwin 2003). Während mit „ringer" in den Zeilen 01, 02, 08–09 die Flasche mit der Lösung gemeint ist, bezieht sich die zweite Referenz in Zeile 09 „guck ob se läuft" nicht auf die Flasche, sondern auf die Lösung selbst, die von der Flasche über den Schlauch in die Vene des Patienten zu laufen hat. Hierfür sind zwar die semantischen Valenzeigenschaften des Verbs *laufen* für das Subjektargument wichtig (vgl. Lesart 5 von *laufen*, „Flüssigkeit entweichen lassen", in E-VALBU, Kubczak 2009), doch liefern sie nicht mehr als eine mögliche Restriktion. Das sofortige reibungslose Verständnis ist nur aufgrund des professionellen Wissens über den Handlungsablauf des Anbringens und der Funktionsweise einer Ringerinfusion möglich.

Du Bois (2003) hat gezeigt, dass die Argumentrealisierung in der gesprochenen Sprache speziellen Restriktionen unterliegt, die für typologisch unterschiedliche Sprachen gelten. Auch für das Deutsche konnten die von Du Bois behaupteten Tendenzen, insbesondere die Präferenz für maximal ein neues Argument, generell bestätigt werden. Allerdings sind deutliche Unterschiede je nach Verb und Verwendungskontext festzustellen (Proske 2013; Deppermann, Proske & Zeschel 2017). Zu diesen gehören auch die besonderen Konstellationen des leiblichen Handelns und Wahrnehmens, der räumlichen Verfügbarkeit von Objekten und der aktuellen sprachlichen und leiblichen, intersubjektiv geteilten oder vom einzelnen Akteur unabhängig vom Interaktionspartner

verfolgten Handlungszusammenhänge sowie der Kenntnis geteilter, z. B. professioneller Handlungsroutinen (vgl. auch Hindmarsh & Heath 2000). Wenn die Beteiligten auf ein Objekt visuell ko-orientiert sind, dieses im aktuellen Handlungszusammenhangs salient ist und bereits eine mit dem Objekt zu vollziehende nächste Handlung erwartet wird, dann sind elliptische Referenzen ausreichend (Zinken & Deppermann 2017). Sprachliche Praktiken setzen also auf einer leiblichen Infrastruktur des geteilten Bezugs auf Gegenstände und der intersubjektiven Verankerung in Handlungssequenzen auf, welche für die von Bühler ([1934] 1982: 285) festgestellte, bloß „diakritische" Ökonomie der sprachlichen Praktiken die Voraussetzung bildet. Hier können wir anknüpfen an Ludwig Eichingers Feststellung in seinem Geleitwort zum Band der Jahrestagung des IDS 2015 zum Thema „Sprachliche und kommunikative Praktiken":

> Praktiken funktionieren ja nur, wenn sie als Muster, als wiederkehrende aber variable Konstellationen und Prozessformate erkennbar sind. Auf die strukturierende Wirkung solcher musterhafter Einheiten muss man sich verlassen können, wenn Interaktionen möglichst unaufwändig verlaufen sollen. Es ist offenkundig, dass die Verlässlichkeit, mit der wir sprachliche Äußerungen als konstituierende Elemente von gesellschaftlichen Übereinkünften zu Praktiken des Handelns lesen können, eben nicht nur an der Sprache festzumachen sind [...]. (Eichinger 2016: IX)

7 Schlussbemerkungen

In diesem Beitrag habe ich zu zeigen versucht, dass die Analyse von Interaktionen als multimodal organisierten, leiblich, räumlich und gegenständlich geprägten Handlungszusammenhängen nicht nur für das Verständnis von sozialer Interaktion als solcher grundlegend ist, sondern auch unser Verständnis sprachlicher Strukturen und Praktiken entscheidend vertiefen (und manchmal auch korrigieren) kann.

Sprachliche Praktiken werden systematisch mit anderen leiblichen Ressourcen koordiniert. Es entstehen so multimodale, zeitlich organisierte Gestalten, die wiedererkennbar sind und einen formalen, materialen Kern von Handlungspraktiken ausmachen, die ganzheitlich erkennbar und interpretierbar sind. Sprachliche Praktiken erweisen sich als zugeschnitten auf leibliche, räumliche und zeitliche Bedingungen des Interagierens, auf Hör- und Sichtbarkeit, auf Bewegungen und Ausrichtungen des Körpers. Relevanz haben die multimodalen Ressourcen und Konfigurationen für das Handeln und damit auch für die sprachliche Praxis oftmals nicht einfach per se, sondern nur im Zusammenhang ihrer Wahrnehmung und Produktion durch Akteure, die Wissen und frühere Erfahrungen mobilisieren und auf dieser Basis Erwartungen

bilden und ein lokal passendes Verständnis entwickeln können. Sprache und andere Modalitäten stehen in der multimodalen Interaktion in einem Wechselverhältnis: Sprachliche Praktiken setzen auf einer Infrastruktur leiblichen Handelns auf, die Voraussetzungen und Bedingungen schafft. Doch ist diese nicht einfach gegeben, sondern muss selbst erst in der Interaktion hergestellt werden – und dies geschieht oftmals auch nur mit Hilfe von Sprache.

In diesem Artikel haben wir vier klassische sprachwissenschaftliche Fragestellungen angesprochen, für die multimodale Koordinationen eine entscheidende Rolle spielen: die lokale Referenz, die Disambiguierung von Ausdrücken, die Konstruktion komplexer syntaktischer Einheiten und die Argumentrealisierung. Weitere Phänomenbereiche sind schon analysiert worden, wie z. B. modale Deixis (Streeck 1995, 2016; Stukenbrock 2010; Fricke 2012), Selbst-Reparaturen, Abbrüche und Verzögerungen (C. Goodwin 1980; M. H. Goodwin 1980; Streeck 1995; Mondada 2007a, 2009). Andere weitere werden sich mit großer Sicherheit als sensitiv gegenüber leiblichen Prozessen erweisen (z. B. Links- und Rechtsversetzungen, Responsivpartikeln, Modalpartikeln). Es werden in den kommenden Jahren viele weitere Entdeckungen zum Zusammenhang sprachlicher und leiblicher Praktiken zu machen sein. Sie werden uns zu einem vertieften Wissen verhelfen, welche Arten und Eigenschaften sprachlicher Strukturen in welcher Weise mit multimodalen Praktiken zusammenspielen und von ihnen abhängig sind.

Literatur

Argyle, Michael (1975): *Bodily communication*. London: Methuen.

Ariel, Mira (1990): *Accessing noun-phrase antecedents*. London: Routledge.

Ariel, Mira (2008): *Pragmatics and grammar*. Cambridge: Cambridge University Press.

Auer, Peter (1996): On the prosody and syntax of turn-continuations. In: Elizabeth Couper-Kuhlen & Margret Selting (Hrsg.), *Prosody in conversation*, 57–100. Cambridge: Cambridge University Press.

Auer, Peter (2005): Projection in interaction and projection in grammar. *Text* 25 (1), 7–36.

Auer, Peter & Aldo di Luzio (1992): *The contextualization of language*. Cambridge: Cambridge University Press.

Bachmann-Medick, Doris (2006): *Cultural Turns. Neuorientierungen in den Kulturwissenschaften*. Reinbek: Rowohlt.

Baldry, Anthony & Paul Thibaut (2005): *Multimodal transcription and text analysis*. London: Equinox.

Beach, Wayne (1993): Transitional regularities for 'casual' "Okay" usages. *Journal of Pragmatics* 19, 325–352.

Beißwenger, Michael (2007): *Sprachhandlungskoordination in der Chat-Kommunikation*. Berlin, New York: de Gruyter.

Bergmann, Jörg (1985): Flüchtigkeit und methodische Fixierung sozialer Wirklichkeit: Aufzeichnungen als Daten der interpretativen Soziologie. In Wolfgang Bonß & Heinz

Hartmann (Hrsg.), *Entzauberte Wissenschaft: Zur Relativität und Geltung soziologischer Forschung* (Sonderband 3 der Zeitschrift *Soziale Welt*), 299–320. Göttingen: Schwarz.
Birdwhistell, Ray L. (1970): *Kinesics and context: Essays on body-motion communication*. Philadelphia (PA): University of Pennsylvania Press.
Brigard, Emilie de (1995): The history of ethnographic film. In Paul Hockings (Hrsg.), *Principles of visual anthropology*, 13–44. Berlin, New York: de Gruyter.
Bühler, Karl ([1934] 1982): *Sprachtheorie*. Stuttgart: Fischer.
Clark, Herbert H. (1996): *Using language*. Cambridge: Cambridge University Press.
Clift, Rebecca (2016): *Conversation analysis*. Cambridge: Cambridge University Press.
Deppermann, Arnulf (2014): Multimodal participation in simultaneous joint projects: Interpersonal and intrapersonal coordination of paramedics in emergency drills. In Pentti Haddington, Tina Keisanen, Lorenza Mondada & Maurice Nevile (Hrsg.), *Multiactivity in social interaction: Beyond multitasking*, 247–281. Amsterdam: Benjamins.
Deppermann, Arnulf (Hrsg.) (i. Dr.): Special Issue „Instructions in driving lessons". *International Journal of Applied Linguistics*.
Deppermann, Arnulf & Reinhold Schmitt (2007): Koordination. Zur Begründung eines neuen Forschungsgegenstandes. In Reinhold Schmitt (Hrsg.), *Koordination. Analysen zur multimodalen Interaktion*, 15–54. Tübingen: Narr.
Deppermann, Arnulf, Reinhold Schmitt & Lorenza Mondada (2010): Agenda and emergence: Contingent and planned activities in a meeting. *Journal of Pragmatics* 42 (6), 1700–1718.
Deppermann, Arnulf & Susanne Günthner (2015): Introduction: Temporality in interaction. In Arnulf Deppermann & Susanne Günthner (Hrsg.), *Temporality in interaction*, 1–24. Amsterdam: Benjamins.
Deppermann, Arnulf, Helmuth Feilke & Angelika Linke (2016): Sprachliche und kommunikative Praktiken: Eine Annäherung aus linguistischer Sicht. In: Arnulf Deppermann, Helmuth Feilke & Angelika Linke (Hrsg.), *Sprachliche und kommunikative Praktiken*, 1–23. Berlin: de Gruyter.
Deppermann, Arnulf, Nadine Proske & Arne Zeschel (Hrsg.) (2017): *Verben im interaktiven Kontext. Bewegungsverben und mentale Verben im gesprochenen Deutsch*. Tübingen: Narr.
Du Bois, John W. (2003): Discourse and grammar. In Michael Tomasello (Hrsg.), *The new psychology of language*, 47–87, Vol. 2. London: Erlbaum.
Ehlich, Konrad (1986): *Interjektionen*. Tübingen: Niemeyer.
Eichinger, Ludwig M. (2016): Praktiken: etwas Gewissheit im Geflecht der alltäglichen Welt. In Arnulf Deppermann, Helmuth Feilke & Angelika Linke (Hrsg.), *Sprachliche und kommunikative Praktiken*, VII–XIII. Berlin, Boston: de Gruyter.
Ekman, Paul, Wallace Friesen & Phoebe Ellsworth (1972): *Emotion in the human face*. London: Penguin.
Enfield, Nick J. (2009): *The anatomy of meaning*. Cambridge: Cambridge University Press.
Fricke, Ellen (2007): *Origo, Geste und Raum– Lokaldeixis im Deutschen*. Berlin, New York: de Gruyter.
Fricke, Ellen (2012): *Grammatik multimodal: Wie Wörter und Gesten zusammenwirken*. Berlin: de Gruyter.
Garz, Detlef & Klaus Kraimer (Hrsg.) (1994): *Die Welt als Text. Theorie, Kritik und Praxis der objektiven Hermeneutik*. Frankfurt a. M.: Suhrkamp.

Gibson, James J. (1979): *The ecological approach to visual perception*. Boston: Houghton Mifflin.
Goffman, Erving (1967): *Interaction ritual: Essays in face-to-face behavior*. London: Aldine.
Goffman, Erving (1979): Footing. *Semiotica* 25 (1–2), 1–29.
Goodwin, Charles (1979): The interactive construction of a sentence in natural conversation. In George Psathas (Hrsg.), *Everyday language: Studies in ethnomethodology*, 97–121. New York: Irvington.
Goodwin, Charles (1980): Restarts, pauses, and the achievement of mutual gaze at turn-beginning. *Sociological Inquiry* 50 (3–4), 272–302.
Goodwin, Charles (1981): *Conversational organization. Interaction between speakers and hearers*. New York: Academic.
Goodwin, Charles & Marjorie H. Goodwin (1986): Gesture and coparticipation in the activity of searching for a word. *Semiotica* 62 (1–2), 51–75.
Goodwin, Charles (1994): Professional vision. *American Anthropologist* 96 (3), 606–633.
Goodwin, Charles (1997): The blackness of black: Color categories as situated practice. In Lauren Resnick, Roger Säljö, Clotilde Pontecorvo & Barbara Burge (Hrsg.), *Discourse, tools and reasoning: Essays on situated cognition*, 111–140. New York: Springer.
Goodwin, Charles (2003): Pointing as situated practice. In Sotaro Kita (Hrsg.), *Pointing: Where language, culture and cognition meet*, 217–241. Mahwah (NJ): Lawrence Erlbaum.
Goodwin, Marjorie H. (1980): Processes of mutual monitoring Implicated in the production of description sequences. *Sociological Inquiry* 50 (3–4), 303–317.
Goodwin, Charles & Marjorie H. Goodwin (2004): Participation. In Alessandro Duranti (Hrsg.), *A companion to linguistic anthropology*, 222–244. Malden (MA): Blackwell.
Grifoni, Patrizia (Hrsg.) (2009): *Multimodal human computer interaction and pervasive services*. Hershey (PA): Information Science Reference.
Gumperz, John J. (1982): *Discourse strategies*. Cambridge: Cambridge University Press.
Haddington, Pentti, Tiina Keisanen, Lorenza Mondada & Maurice Nevile (Hrsg.) (2014), *Multiactivity in social interaction: Beyond multitasking*. Amsterdam: Benjamins.
Haddington, Pentti, Tiina Keisanen & Maurice Nevile (Hrsg.) (2012): Meaning in motion: Sharing the car, sharing the drive. *Semiotica*, 191 (1/4).
Haddington, Pentti, Lorenza Mondada & Maurice Nevile (Hrsg.) (2013): *Interaction and mobility. Language and the body in motion*. Berlin: de Gruyter.
Hausendorf, Heiko, Lorenza Mondada & Reinhold Schmitt (2012): Raumals interaktive Ressource: Eine Explikation. In Heiko Hausendorf, Lorenza Mondada & Reinhold Schmitt (Hrsg.), *Raumals interaktive Ressource*, 7–36. Tübingen: Narr.
Heath, Christian (1986): *Body movement and speech in medical interaction*. Cambridge: Cambridge University Press.
Heath, Christian, Jon Hindmarsh & Paul Luff (2010): *Video in qualitative research*. London: Sage.
Heath, Christian & Paul Luff (2000): *Technology in action*. Cambridge: Cambridge University Press.
Heath, Christian & Paul Luff (2013): Embodied action and organizational activity. In Jack Sidnell & Tanya Stivers (Hrsg.), *The handbook of conversation analysis*, 283–307. Oxford: Wiley-Blackwell.
Heath, Christian & Jon Hindmarsh (2000): Configuring action in objects: From mutual space to media space. *Mind, culture and activity* 7 (1–2), 81–104.
Hepburn, Alexa & Galina Bolden (2017): *Transcribing for social research*. London: Sage.

Helmer, Henrike (2016): *Analepsen in der Interaktion. Semantische und sequenzielle Eigenschaften von Topik-Drop im gesprochenen Deutsch*. Heidelberg: Winter.
Heritage, John (1984): *Garfinkel and ethnomethodology*. Oxford: Polity.
Hindmarsh, Jon & Christian Heath (2000): Embodied reference: A study of deixis in workplace interaction. *Journal of Pragmatics* 32 (12), 1855–1878.
Hoey, Elliott (2015): Lapses: How people arrive at, and deal with, discontinuities in talk. *Research on Language and Social Interaction* 48 (4), 430–453.
Imo, Wolfgang (2015): Nachträge im Spannungsfeld von Medialität, Situation und interaktionaler Funktion. In Hélène Vinckel-Roisin (Hrsg.), *Das Nachfeld im Deutschen: Theorie und Empirie*, 231–253. Berlin: de Gruyter.
Jefferson, Gail (1972): Side sequences. In David N. Sudnow (Hrsg.), *Studies in social interaction*, 294–333. New York (NY): Free Press.
Jefferson, Gail (2004): A sketch of some orderly aspects of overlap in conversation. In Gene H. Lerner (Hrsg.), *Conversation analysis. Studies from the first generation*, 43–59. Amsterdam: Benjamins.
Kendon, Adam (1972): Some relationships between body motion and speech. In Aaron Siegman & Benjamin Pope (Hrsg.), *Studies in dyadic communication*, 177–216. Elmsford: Pergamon.
Kendon, Adam (2004): *Gesture*. Cambridge: Cambridge University Press.
Kubczak, Jacqueline (2009): Eintrag *laufen*. In Jacqueline Kubczak, *E-VALBU – Das elektronische Valenzwörterbuch deutscher Verben*. Mannheim: Institut für Deutsche Sprache. http://hypermedia.ids-mannheim.de/evalbu/index.html (letzter Zugriff 17. 6. 2017).
Leeds-Hurwitz, Wendy (Hrsg.) (2010): *The social history of language and social interaction research: People, places, ideas*. Cresskill (NJ): Hampton Press.
Lyons, John (1983): *Semantik*. Bd. 2. München: C. H. Beck.
Marx, Konstanze & Georg Weidacher (2014): *Internetlinguistik. Ein Lehr- und Arbeitsbuch*. Tübingen: Narr.
Merleau-Ponty, Maurice ([1945] 1966): *Phänomenologie der Wahrnehmung*. Berlin: de Gruyter.
Meyer, Christian, Jürgen Streeck & Scott Jordan (Hrsg.) (2017): *Intercorporeality. Emerging socialities in interaction*. Oxford: Oxford University Press.
Mondada, Lorenza (2007a): Interaktionsraum und Koordinierung. In Reinhold Schmitt (Hrsg.), *Koordination. Analysen zur multimodalen Interaktion*, 55–94. Tübingen: Narr.
Mondada, Lorenza (2007b): Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies* 9 (2), 195–226.
Mondada, Lorenza (2009): Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of Pragmatics* 41, 1977–1997.
Mondada, Lorenza (2013): The conversation analytic approach to data collection. In Jack Sidnell & Tanya Stivers (Hrsg.), *Handbook of conversation analysis*, 32–56. New York: Blackwell-Wiley.
Mondada, Lorenza (2014a): The local constitution of multimodal resources for social interaction. *Journal of Pragmatics* 65, 137–156.
Mondada, Lorenza (2014b): Conventions for multimodal transcription. Basel: Universität Basel. https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/mondada_multimodal_conventions.pdf (letzter Zugriff 27. 4. 2018).

Mondada, Lorenza (2014c): The temporal orders of multiactivity: operating and demonstrating in the surgical theatre. In Pentti Haddington, Tiina Keisanen, Lorenza Mondada & Maurice Nevile (Hrsg.), *Multiactivity in social interaction: Beyond multitasking*, 33–75. Amsterdam: Benjamins.

Mondada, Lorenza (2015): Multimodal completions. In Arnulf Deppermann & Susanne Günthner (Hrsg.), *Temporality in interaction*, 267–307. Amsterdam: Benjamins.

Mondada, Lorenza (2016a): Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics* 20 (2), 2–32.

Mondada, Lorenza (2016b): Zwischen Text und Bild: Multimodale Transkription. In Heiko Hausendorf, Reinhold Schmitt & Wolfgang Kesselheim (Hrsg.), *Interaktionsarchitektur, Sozialtopograhie und Interaktionsraum*, 111–160. Tübingen: Narr.

Mondada, Lorenza (2018): Multiple temporalities and language and body in interaction. Challenges for transcribing multimodality. *Research on Language and Social Interaction* 51 (1), 85–106.

Mondada, Lorenza (i. Dr.): Questions on the move. The ecology of question-answer sequences in mobile settings. In Arnulf Deppermann & Jürgen Streeck (Hrsg.), *Modalities and temporalities: Convergences and divergences of bodily resources in interaction*. Amsterdam: Benjamins.

Muybridge, Eadweard (1901): *The human figure in motion*. London: Chapman and Hall.

Oloff, Florence (i. Dr.): Revisiting delayed completions: The retrospective management of co-participant action. In Arnulf Deppermann & Jürgen Streeck (Hrsg.), *Modalities and temporalities: Convergences and divergences of bodily resources in interaction*. Amsterdam: Benjamins.

Proske, Nadine (2013): *Informationsmanagement im gesprochenen Deutsch. Eine diskurspragmatische Untersuchung syntaktischer Strukturen in Alltagsgesprächen*. Heidelberg: Winter.

Proske, Nadine (2015): Die Rolle komplexer Nachfeldbesetzungen bei der Einheitenbildung im gesprochenen Deutsch. In Hélène Vinckel-Roisin (Hrsg.), *Das Nachfeld im Deutschen. Theorie und Empirie*, 279–297. Berlin: de Gruyter.

Sacks, Harvey (1963): Sociological description. *Berkeley Journal of Sociology* 8, 1–16.

Sacks, Harvey ([1964–1972] 1992): *Lectures on conversation*. 2 Bde. Oxford: Blackwell.

Sacks, Harvey, Emanuel A. Schegloff & Gail Jefferson (1974): A simplest systematics for the organization of turn-taking for conversation. *Language* 50 (4), 696–735.

Sacks, Harvey & Emanuel A. Schegloff ([1974] 2002): Home position. *Gesture* 2, 133–146.

Scheflen, Albert E. (1972): *Body language and social order: Communication as behavioral control*. Englewood Cliffs (NJ): Prentice-Hall.

Schegloff, Emanuel A. (1982): Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In Deborah Tannen (Hrsg.), *Analyzing discourse: Text and talk*, 71–93. Washington DC: Georgetown University Press.

Schegloff, Emanuel A. (1984): On some gestures' relation to talk. In John M. Atkinson & John Heritage (Hrsg.), *Structures of social action*, 266–298. Cambridge: Cambridge University Press.

Schegloff, Emanuel A. (1998): Body torque. *Social Research* 65 (3), 535–596.

Schegloff, Emanuel A. (2007): *Sequence organization*. Cambridge: Cambridge University Press.

Schmidt, Axel (2014): *Spiel oder nicht Spiel? Zur interaktiven Organisation von Übergängen zwischen Spielwelt und Realwelt in Theaterproben*. Mannheim: Verlag für

Gesprächsforschung. http://www.verlag-gespraechsforschung.de/2014/pdf/theaterproben.pdf (letzter Zugriff 17. 6. 2017).

Schmidt, Jürgen Erich (2001): Bausteine der Intonation? *Germanistische Linguistik* 157–158, 9–32.

Schmitt, Reinhold (2004): Die Gesprächspause: Verbale „Auszeiten" aus multimodaler Perspektive. *Deutsche Sprache* 32 (1), 56–84.

Schmitt, Reinhold (2005): Zur multimodalen Struktur von turn-taking. *Gesprächsforschung* 6, 17–61. http://www.gespraechsforschung-ozs.de/fileadmin/dateien/heft2005/ga-schmitt.pdf (letzter Zugriff 17. 6. 2017).

Schmitt, Reinhold (2012): Körperlich-räumliche Grundlagen interaktiver Beteiligung am Filmset: Das Konzept ‚Interaktionsensemble'. In Heiko Hausendorf, Lorenza Mondada & Reinhold Schmitt (Hrsg.), *Raumals interaktive Ressource*, 37–87. Tübingen: Narr.

Schmitt, Reinhold (2015): Positionspapier: Multimodale Interaktionsanalyse. In Ulrich Dausendschön-Gay, Elisabeth Gülich & Ulrich Krafft (Hrsg.), *Ko-Konstruktionen in der Interaktion. Die gemeinsame Arbeit an Äußerungen und anderen sozialen Ereignissen*, 43–51. Bielefeld: transcript.

Schmitt, Reinhold (2016): Der „Frame-Comic" als Dokument multimodaler Interaktionsanalysen. In Heiko Hausendorf, Reinhold Schmitt & Wolfgang Kesselheim (Hrsg.), *Interaktionsarchitektur, Sozialtopographie und Interaktionsraum*, 189–224. Tübingen: Narr.

Schmitt, Reinhold & Arnulf Deppermann (2007): Monitoring und Koordination als Voraussetzungen der multimodalen Konstitution von Interaktionsräumen. In Reinhold Schmitt (Hrsg.), *Koordination. Analysen zur multimodalen Interaktion*, 95–128. Tübingen: Narr.

Schmitt, Reinhold & Arnulf Deppermann (2010): Die Transition von Interaktionsräumen als Eröffnung einer neuen Situation. In Lorenza Mondada & Reinhold Schmitt (Hrsg.), *Situationseröffnungen. Zur multimodalen Herstellung fokussierter Interaktion*, 335–386. Tübingen: Narr.

Schneider, Jan G. & Georg Albert (2013): Medialität und Standardsprache – oder: Warum die Rede von einem gesprochenen Gebrauchsstandard sinnvoll ist. In Jörg Hagemann, Wolf Peter Klein & Sven Staffeldt (Hrsg.), *Pragmatischer Standard*, 49–60. Tübingen: Stauffenburg.

Schwitalla, Johannes (2002): Kleine Wörter. Partikeln im Gespräch. In Jürgen Dittmann & Claudia Schmidt (Hrsg.), *Über Wörter*, 259–282. Freiburg: Rombach.

Selting, Margret et al. (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT2). *Gesprächsforschung* 10, 353–402. http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf (letzter Zugriff 17. 6. 2017).

Streeck, Jürgen (1995): On projection. In Edward Goody (Hrsg.), *Interaction and social intelligence*, 84–110. Cambridge: Cambridge University Press.

Streeck, Jürgen (2016): Gestische Praxis und sprachliche Form. In Arnulf Deppermann, Helmuth Feilke & Angelika Linke (Hrsg.), *Sprachliche und kommunikative Praktiken*, 57–80. Berlin: de Gruyter.

Stukenbrock, Anja (2009): Herausforderungen der multimodalen Transkription. In Karin Birkner & Anja Stukenbrock (Hrsg.), *Die Arbeit mit Transkripten in Fortbildung, Lehre und Forschung, Mannheim: Verlag für Gesprächsforschung*, 144–170.

Stukenbrock, Anja (2010): Überlegungen zu einem multimodalen Verständnis der gesprochenen Sprache am Beispiel deiktischer Verwendungsweisen des Ausdrucks

„so“. In Norbert Dittmar & Nils Bahlo (Hrsg.), *Beschreibungen für gesprochenes Deutsch auf dem Prüfstand*, 165–193. Frankfurt a. M.: Peter Lang.
Stukenbrock, Anja (2015): *Deixis in der face-to-face Interaktion*. Berlin: de Gruyter.
Zinken, Jörg & Arnulf Deppermann (2017): A cline of visible commitment in the situated design of imperative turns. Evidence from German and Polish. In Elizabeth Couper-Kuhlen, Liisa Raevaara & Marja-Leena Sorjonen (Hrsg.), *Imperative turns at talk*, 27–63. Amsterdam: Benjamins.

Anhang: Konventionen für die Transkription visueller Phänomene

Die Transkriptionen entsprechen den Konventionen von Mondada (2014).
Die Siglen der Interaktionsbeteiligten werden klein geschrieben, wenn visuelle Phänomene notiert werden:

abc	Siglen der Interaktionsbeteiligten
	Die zeitliche Erstreckung visueller Phänomene, d. h. kinesischer, non-verbaler Aktivitäten wird durch ihre Verankerung in Sprecherzeilen bzw. in Pausenzeilen angezeigt. Folgende Zeichen werden zur Markierung des Start- und Endpunkts einer Aktivität benutzt:
+‡*§$%&	Markierung der Extension einer visuellen Aktivität
	Die einzelnen Phasen einer Aktivität werden gemäß der von Kendon (2004) eingeführten Konvention wiedergegeben:
. . .	Präparation einer Aktivität
---	Andauern einer Aktivität
, , ,	Retraktion einer Aktivität