Ein entscheidender Schritt bei der empirisch-qualitativen Analyse natürlicher Interaktionsereignisse ist neben der Datenerhebung die Datenaufbereitung. Grundsätzlich gilt: Sprache ist ein flüchtiges Medium, das gerade zu Zwecken der Mikroanalyse im Sinne der Konversationsanalyse verschriftet werden muss (vgl. Heath/Luff 2009; siehe auch Luckmann 1980:35). Transkripte sind dann graphische Repräsentationen und Dokumentationen kommunikativer Ereignisse (auf der Grundlage von Audio- und/oder Videoaufzeichnungen), die dazu dienen, die aufgenommenen flüchtigen Daten (dauerhaft) zu fixieren, zu speichern (vgl. Bohle 2013:992; siehe dazu auch Dittmar 2002:31) und für die wissenschaftliche Analyse zugänglich zu machen (vgl. Ayaß 2015:508).

Transkripte sind ein konstitutiver Bestandteil interaktionsanalytischer Forschung (vgl. Ayaß 2015:508, Birkner 2009:2) und stellen einen erheblichen Unterschied zu ethnographischen Feldnotizen dar: „In CA [...] transcription [...] entails a specific methodological attitude, that is, a non-reliance on researchers' memories and recollections along with the ambition to achieve a registering access to the social reality in question“ (Ayaß 2015:507; zur Konversationsanalyse siehe Abschnitt 10.2). Erst mithilfe von Transkripten wird es möglich, die Struktur kommunikativer Ereignisse in ihren Einzelheiten zu erfassen und für andere Forschende intersubjektiv sichtbar und nachvollziehbar zu machen (vgl. Hutchinson Guest 2005:7). Dabei sind Transkripte mitnichten objektive Repräsentationen tatsächlicher kommunikativer Ereignisse (vgl. Ayaß 2015:510; siehe auch Stukenbrock 2009:147). Vielmehr erfolgt durch die Transkribierenden

Elektronisches Zusatzmaterial Die elektronische Version dieses Kapitels enthält Zusatzmaterial, das berechtigten Benutzern zur Verfügung steht
https://doi.org/10.1007/978-3-658-33855-8_12.

bereits bei der Erstellung des Transkripts eine Auswahl, welche der beobachtbaren Phänomene im Transkript abgebildet werden und welche Aspekte (unter Berücksichtigung der jeweiligen Fragestellung) in den Hintergrund treten (vgl. Deppermann 2008:41). Jedes Transkript ist damit eine – auch analytisch notwendige – Reduktion der komplexen Interaktionsereignisse, auf deren Basis sie erstellt werden (vgl. Luckmann 2009:31, Stukenbrock 2009) und zugleich nicht nur Hilfsmittel der Analyse, sondern Teil der Analyse selbst (vgl. Ayaß 2015:508). Transkripte werden daher als eigene, durch die Transkribierenden generierte Datensorte (sog. Tertiärdaten) betrachtet, die in enger Verbindung zum eigentlichen Interaktionsereignis (Primärdaten) und den davon erhobenen Audio- und/oder Videodaten (Sekundärdaten) steht (vgl. Ayaß 2015:511). Dabei ändert sich die Materialität: „Ein Phänomen X in der Vorkommensart Y wird in einen anderen medialen und materiellen Zustand Z übertragen mit der theoretischen Prämisse, dass es als X‘ in einer methodisch kontrollierten Weise für empirische Analysezwecke erhalten bleibt“ (Stukenbrock 2009:145). Aus dem Wechsel der Materialitäten heraus entstehen große Herausforderungen für die Transkription von verbalen und vokalen Aspekten der Interaktion und erst recht für die Transkription visueller Aspekte, da nun die Simultaneität ausgeführter Handlungen nicht nur in ihrer eigenen Sequenzialität, sondern der Orchestrierung im linearen Medium der Schrift zweidimensional abgebildet werden muss (vgl. Stukenbrock 2009:145 f, Hutchinson Guest 2005:9, Imo/Lanwer 2019:122).

Darüber hinaus wird nicht nur die Forderung intersubjektiver Vermittlung, sondern auch sprachunabhängiger Lesbarkeit an das Transkript gestellt (ähnlich der Universalität und Sprachunabhängigkeit des Notensystems in der Musik; vgl. Sutton 1982:79). Diesem Anspruch sind alle in der Wissenschaft vorhandenen Transkriptionssysteme verpflichtet und versuchen ihn einzulösen, wenn auch auf unterschiedliche Art und Weise (vgl. Deppermann 2008: FN3). Mit GAT2 (Selting et al. 2009) hat sich im deutschsprachigen Raum ein Transkriptionssystem etabliert, das ein einheitliches und dabei an individuelle Anforderungen anpassbares Repertoire an Zeichen bereitstellt, um vor allem verbale und vokale Interaktionsaspekte möglichst präzise abzubilden. Für die Beteiligten an einer Interaktion ist jedoch gerade die Synchronisierung bzw. die Simultaneität von körperlichen und stimmlich-sprachlichen Ausdrucksressourcen ein essentieller und in der Hauptsache unreflektierter Vorgang. Die Interagierenden konstruieren den situativen Sinn über lokal hervorgebrachte Kontextualisierungshinweise und nutzen dabei alle ihnen zur Verfügung stehenden Ressourcen (siehe Kapitel 10). Es muss also das Ziel sein, genau diese Verknüpfung nicht nur in der Analyse zu berücksichtigen, sondern sie auch in den Transkripten abzubilden.

12.1 Transkription von Multimodalität

Im Zuge der Hinwendung zu visuellen Aspekten der Interaktion, der Nutzung neuer Datensorten in Form von Videos und der (multimodalen) Erweiterung sowohl der Konversationsanalyse als auch der Interaktionalen Linguistik bis hin zur Multimodalen Interaktionsanalyse (siehe Kapitel 10) wurde deutlich, dass das entwickelte Zeicheninventar der aktuell vorhandenen Transkriptionssysteme nur eine unzureichende Berücksichtigung der in Interaktionen erkennbaren visuellen Ressourcen erlaubt und so für die neuen Datensätze nicht mehr ausreichend ist (vgl. Mittelberg 2007:225, Bohle 2013:1000, Bressem 2013, Stukenbrock 2009:145). So wird der Ruf danach laut, alle von den Interagierenden in der Interaktion als relevant erkennbar eingesetzten Ausdrucksressourcen nicht nur in den Analysen zu beschreiben, sondern diese adäquat in den Transkripten abzubilden, um die Nachvollziehbarkeit der Ergebnisse zu erhöhen (vgl. Deppermann et al. 2016:13). Entsprechend ist es notwendig, ein für die Bedingungen und Bedürfnisse der Transkription körperlicher Bewegungen zugeschnittenes Zeicheninventar zusammenzustellen. Dass diese Anforderung der Nachvollziehbarkeit in Kombination mit der adäquaten Detailliertheit und der Lesbarkeit der Transkripte beträchtliche Herausforderungen mit sich bringt, zeigt schon der Umstand, dass trotz der stetig wachsenden Entwicklung der Multimodalitätsforschung nach wie vor kein einheitliches und standardisiertes System zur Transkription visueller Aspekte und Handlungen in der Interaktion existiert (vgl. Mittelberg 2007:227, Mondada 2013b:986, Bohle 2013:993, vom Lehn 2014:399, Luff/Heath 2015:367, Imo/Lanwer 2019:120). So fehlen Konventionalisierungen für die Darstellung körperlicher Ausdrucksressourcen wie Mimik, Gestik, Blick, Körperorientierung, Körperhaltung etc. sowie für den Gebrauch von Objekten und Artefakten in der Interaktion (vgl. Heath/Luff 2009:35). Eine Ausnahme bildet lediglich die Transkription des Blickverhaltens in der Interaktion. In vielen Arbeiten wird hier das von Rossano (2013) entwickelte System übernommen. Bei der Transkription aller anderen Ressourcen wird mit speziell für die jeweilige Analyse und das jeweilige Material passenden individuell gestalteten Konventionen gearbeitet, die z. T. nur schwer auch auf andere Daten übertragen werden können (vgl. Bohle 2013:993, vom Lehn 2014, Mondada 2016a:115). Das Fehlen eines einheitlichen Systems liegt auch daran, dass immer wieder die Frage gestellt wird, inwieweit eine Standardisierung im Bereich der Transkription von Multimodalität überhaupt produktiv ist (vgl. Birkner 2009:2). Mondada bezieht dazu eher kritisch Stellung. Sie betont, multimodale Transkripte

have to be flexible, relying neither on a canonical set of pregiven forms nor on their a priori hierarchization. They must be able to accommodate a variety of resources, including unique, ad hoc, and locally situated ones, besides more conventional ones. In other words, they must be able to represent the specific temporal trajectories of a diversity of multimodal details, including talk where this is relevant, but also silent embodied actions when talk is not the main resource or activity. (Mondada 2018a:88)

Andere sehen dagegen die Notwendigkeit eines „einheitliche[n] Notationssystem[s] für die multimodale Transkription in der Linguistik [...] aufbauend auf bestehenden verbalen Transkriptionskonventionen und im Rückgriff auf bereits usuell gewordene multimodale Transkriptionspraxen“ (Stukenbrock 2009:146), um den intersubjektiven Nachvollzug der zahlreichen Studien zur Multimodalität zu gewährleisten. Auch die vorliegende Arbeit folgt dieser Argumentation und sieht die Notwendigkeit für die Entwicklung eines anpassbaren und flexiblen, nichtsdestotrotz auf standardisierten Elementen und Konventionen aufbauenden Systems. Ziel soll es sein, ein Transkriptionssystem zu entwickeln, dass zum einen die verbale Dimension angemessen abbildet (wie z. B. GAT2 dies tut) und gleichzeitig auf der Grundlage einheitlicher Konventionen „die Spezifika der körperlichen Dimension sozialen Handelns“ (Mondada 2016a:118) erfasst.

Im Zuge der Entwicklung von GAT (vgl. Selting et al. 1998) und schließlich GAT2 (vgl. Selting et al. 2009) wurden allgemeine Kriterien für ein gutes Transkriptionssystem ausgearbeitet: Ausbaubarkeit bzw. Verfeinerbarkeit (Zwiebelprinzip), Lesbarkeit sowohl für mit dem System vertraute Forschende als auch für Nicht-Forschende, Ökonomie, Eindeutigkeit und Robustheit der verwendeten Zeichen, Ikonizität, die Erfassung relevanter Phänomene, formbezogene Parametrisierung und die Kompatibilität mit anderen Transkriptionssystemen (vgl. Selting et al. 1998:3; siehe dazu auch Deppermann 2008:41). Auch wenn diese Forderungen zunächst für die Transkription verbaler und vokaler Aspekte aufgestellt wurden, gelten sie in analoger Weise auch für die Transkription visueller Aspekte. So benennt Sager für ein gutes Transkript der Multimodalität: Praktikabilität, Relevanz, Adäquanz, Neutralität und Variabilität (vgl. Sager 2001:1069 f). Dabei wird immer wieder auf die Diskrepanz zwischen Detailliertheit und Lesbarkeit hingewiesen, denn die Fülle an abbildbaren Interaktionsressourcen führt in vielen Fällen zu einer „überkomplexen Darstellung“ (Selting et al. 1998:25) innerhalb der Transkripte. Entsprechend stark muss die selektive, analysebezogene Auswahl dessen, was im Transkript repräsentiert wird, sein.

Selting et al. schlagen vor, vor allem Phänomene auf den Ebenen der Proxemik (d. h. des „Territorial- und Distanzverhaltens“ – u. a. „Zu- und Abgänge von Personen sowie die Positionierung und Bewegung von Personen im Raum“ (Selting et al. 1998:26)), der Kinesik (d. h. „Körperbewegungen und körperliche

Aktivitäten; dazu gehören Haltungs-, Positions- und Orientierungsveränderungen“ (Selting et al. 1998:26)), der Gestik, der Mimik, der Blickrichtung und der Kopforientierung zu berücksichtigen (vgl. Selting et al. 1998:26). Folgende Konventionen und Layoutrichtlinien werden vorgestellt:¹

In der dem Transkript und/oder Transkriptausschnitt vorangestellten Beschreibung der sozialen Situation wird neben anderen Merkmalen auch die räumlich-körperliche Ausgangskonstellation des Gesprächs (Sitzkonfiguration, Haltung, Orientierung, Position der Teilnehmerinnen und Teilnehmer) erfasst. Ereignisse, d. h. Zustandsänderungen, werden dann zum Zeitpunkt ihres Auftretens parallel zu den verschrifteten lautlichen Äußerungen notiert. [...] Sichtbare Aktivitäten und Ereignisse werden in einer separaten Zeile unterhalb der gleichzeitig hörbaren lautlichen Aktivitäten (sofern gleichzeitig artikuliert wird) notiert. Im einzelnen sind dabei folgende Zeichen vorgesehen: die Dauer einer sichtbaren Aktivität wird durch Unterstriche gekennzeichnet; Anfangs- und Endpunkte werden durch senkrechte Striche („|“) angegeben; die Beschreibung der sichtbaren Aktivität wird in eine andere Type (z. B. Proportionalschrift) gesetzt, wodurch das Transkript übersichtlicher bleibt; Die Beschreibung der sichtbaren Aktivität ist verhaltensorientiert und möglichst wenig interpretierend. [...] Zeilen des sichtbaren Handelns werden nicht nummeriert, damit ein vorliegendes Transkript zu einem späteren Zeitpunkt um sichtbare Anteile ergänzt (oder auch gekürzt) werden kann, ohne dass sich dadurch die Zeilennummerierung ändert. (Selting et al. 1998:25–27)

Abbildung 12.1 zeigt, wie ein Transkript nach GAT aussehen kann, das sowohl verbale, vokale als auch visuell-leibliche Aspekte darstellt.

Die nach GAT und GAT2 etablierte Zeilentranskription nach Intonationsphrasen bleibt hier erhalten. Die Zeilen mit der Beschreibung visuell wahrnehmbarer Handlungen werden unterhalb der nummerierten Verbalzeile ergänzt. Zusätzlich sprechen sich die Autorinnen und Autoren für die Integration von Standbildern aus, um die Nachvollziehbarkeit zu erhöhen (vgl. Selting et al. 1998:28). Diese sind ein wesentlicher Bestandteil von allen hier vorgestellten Modellen der Transkription von Multimodalität. Sie dienen dabei nicht nur der Illustration, sondern sind ebenfalls Teil des Analyseprozesses, denn sie repräsentieren lediglich einen bestimmten Zeitpunkt innerhalb der Interaktion (vgl. Stukenbrock 2009:154, Mondada 2016a:131, Hausendorf/Schmitt 2016b:162 f) und geben somit „ein dynamisches Objekt durch eine starre Darstellung wieder“ (Mondada

¹Die hier wiedergegebenen Konventionen stammen aus dem ersten Entwurf des ‚Gesprächsanalytischen Transkriptionssystems‘ (GAT) (Selting et al. 1998). In der Überarbeitung und Erweiterung des Systems (GAT2; Selting et al. 2009) fehlen Konventionen zur Transkription visueller Interaktionsaspekte. Vielmehr verweisen die Autorinnen und Autoren darauf, dass bis zur Entwicklung entsprechender Standards die Konventionen des GAT genutzt werden sollen.

und versuchen bei der Entwicklung ihrer eigenen Konventionen diesem Kritikpunkt mithilfe unterschiedlicher Konventionen entgegenzuwirken, indem z. B. mathematische Zeichen als Code für bestimmte Bewegungen dienen. Jede dieser Varianten (Veränderung der Schriftart, Einfügen von Zeichen, Strichen, Punkten und Pfeilen usw.) steht jedoch weiterhin vor dem Problem, Detailliertheit, Nachvollziehbarkeit, die Abbildung aller relevanten verbalen, vokalen und visuellen Interaktionsaspekte, Übersichtlichkeit und Nachvollziehbarkeit bestmöglich zu berücksichtigen und zu gewährleisten. So setzt auch Mondada diese Aspekte für ihr eigenes System zentral:

These conventions are conceived to annotate all possibly relevant embodied actions, such as gesture, gaze, body posture, movements, etc. that happen simultaneously to talk or during moments of absence of talk. The convention is based on two principles: a) *Characterization of the temporal trajectory*. [...] b) *Characterization of the embodied action*: each embodied action is shortly described. The conventions are based on economy and consistency: by putting no more annotations than necessary, and [sic!] by choosing annotations that are interpretable univoquely. (Mondada 2016c:1; Hervorhebungen im Original)

Da Mondada sich dafür ausspricht, nicht mit ikonischen Zeichen zu arbeiten, ist es ihr Ziel, die Beschreibung der Bewegungen so kurz und intersubjektiv wie möglich zu halten und lediglich die ausgeführte Handlung zu bestimmen (z. B. in Abbildung 12.2, Z. 3: ‚points/zeigt‘), ggf. das Körperteil, mit dem die Handlung ausgeführt wird, und – sofern dies tatsächlich erkennbar ist – den Adressaten der Handlung (vgl. Mondada 2016c:6). Abbildung 12.2 zeigt, wie ein Transkript nach den von Mondada erstellten Konventionen aussehen kann.

Die Notation visueller Aspekte wird im System nach Mondada, ähnlich wie bei GAT, unterhalb des Gesprochenen angeführt und die eingefügte Transkriptzeile nicht nummeriert. Die Interagierenden erhalten dann ein eigenes Zeichen, das die ausgeführte Bewegung dieser Person zuordnet (z. B. in Abb. 12.2: BLA %). Zentral für die Notation, wie sie Mondada vorschlägt, ist die Markierung des Anfangspunktes und des Endpunktes einer ausgeführten Bewegung, d. h. die zeitliche Erstreckung. Dazu wird das der Person zugeordnete Symbol sowohl in der Verbalzeile als auch in der Visualzeile am Anfang und am Ende der Handlung vermerkt. Dadurch ist es möglich, körperlich ausgeführte Handlungen zeitlich mit dem Gesprochenen zu alignieren (vgl. Mondada 2016c:2). Auch der Verlauf z. B. einer Geste wird verzeichnet:

```

1  BLA                                     {yeah %i mean,
                                         %turns back->
2  (0.3) % (1.9)
   bla      -->%extends her arm backwards-->
3  BLA      anne rambeau# is {ehm the person who:
           -->%points back-->
           fig      #fig.1

```

Abb. 12.2 Transkription nach Mondada (2016c:8)

Their annotation is inspired by conventions used by Kendon for gesture and Goodwin for gaze: [...] small dots indicate that the embodied action is emerging, [...] commas indicate that the embodied action is withdrawing, retracting. The embodied action itself is described when it has reached its recognizable shape, which can also be maintained for some time. (Mondada 2016c:4)

Wie bei GAT ist es im System von Mondada möglich das Transkript an die jeweiligen Forschungsinteressen anzupassen und einen entsprechenden Grad der Detailliertheit und Feinheit des Transkripts zu wählen (vgl. Mondada 2016c:4). Die Nachvollziehbarkeit soll durch das Einfügen von Standbildern zusätzlich erhöht werden (vgl. Mondada 2016a:130). Dazu werden die Bilder aus dem Video punktuell in das Transkript eingefügt und zeitlich auf das verbale Transkript bezogen (vgl. Mondada 2016a:131; die Alignierung erfolgt durch das Zeichen # im Transkript).

Ähnliche Überlegungen zeigt die Darstellung nach Heath et al. Auch das Ziel dieser Transkriptionskonvention ist es, visuelle Interaktionsaspekte in Relation zu verbalen und vokalen Aspekten aufzuzeigen (vgl. Heath et al. 2010:71), dabei dem situativ eingebetteten, reflexiven, prozessualen und sequenziellen Charakter praktischer Handlungen in der Interaktion Rechnung zu tragen (vgl. Luff/Heath 2015:367) und das sequenzielle und simultane Zusammenspiel unterschiedlicher Ausdrucksressourcen zu dokumentieren (vgl. auch vom Lehn 2014:399).

unter verbale Elemente entstehen (vgl. Ayaß 2015:513). Heath et al. nehmen diese Überlegungen auf und schlagen vor, die Kennzeichnung visuell wahrnehmbarer Handlungen um das Verbaltranskript bzw. die fortschreitende Zeit herum anzuordnen (siehe Abb. 12.3). In der oben gezeigten Abbildung eines von Luff und Heath erstellten Transkripts zu den Abläufen einer Operation ist aufgrund der Abwesenheit gesprochener Sprache die zentrale Transkriptzeile die Notation der vergehenden Zeit (Zeile „time“). Die in dieser Zeit vollzogenen Handlungen des Arztes stehen oberhalb dieser Zeile, die der Operationsschwester darunter (vgl. Heath et al. 2010:71; siehe auch Luff/Heath 2015).² Striche, Punkte und Kommata markieren das körperliche Zu- und Abwenden von Personen und damit die Dauer der ausgeführten Bewegung. Zusätzlich dazu führt Heath kurze Beschreibungen („onset hand to surgeon“, Abb. 12.3) an. Auch Heath et al. ergänzen die Transkription aus Gründen der Eindeutigkeit und des besseren Nachvollzugs zudem mit Standbildern aus dem erhobenen Videomaterial.

Die optische Hierarchisierung hat jenseits theoretischer auch praktische Gründe, die zum einen mit der Zweidimensionalität des Blattes zusammenhängen und zum anderen mit der Notwendigkeit bei der Transkription mit einer Ausdrucksebene beginnen zu müssen. Dies bedingt, dass die Simultaneität und Parallelität von Sprache und Bewegung in einem System untereinander angeordnet wird. Transkripte bedürfen also immer noch einer ausführlichen Erklärung und Analyse, die das Dargestellte einordnen und aus Gründen der Praktikabilität getrennte Interaktionsebenen wieder zusammenführen.

Den schärfsten Kritikpunkt an den bisherigen Konventionen von Multimodaltranskripten sieht auch Ayaß in der Beschreibung und damit Verbalisierung von visuellen Aspekten (vgl. Ayaß 2015:514). Dies ist auch im Transkript nach Heath et al. zu erkennen. Die Autoren betonen jedoch:

We suggest, however, that in practice it is important to transcribe and map the fragments without adding vernacular descriptions, at least initially, since how you eventually describe or characterise the actions of the participants will depend upon analysis of the data. (Heath et al. 2010:71)

Wie schwer es jedoch ist, diesen Anspruch einzulösen, zeigen die drei vorgestellten Transkriptionskonventionen in ihrer praktischen Ausführung. Ein entscheidender Punkt ist an dieser Stelle die Frage nach den genutzten Transkriptionszeichen. Diese sollen sowohl praktikabel als auch adäquat und darüber hinaus nachvollziehbar sein, um die verbale Beschreibung von Bewegungen im Transkript

²Für genauere Informationen zu den einzelnen Konventionen siehe Heath et al. 2010:152 ff.

zu vermeiden. Eine Möglichkeit dafür ist der Gebrauch von gut mit gängigen Textverarbeitungsprogrammen nutzbaren Zeichen aus z. B. mathematischen Zusammenhängen, wie Plus-Zeichen, Punkte etc. Ayaß weist jedoch kritisch darauf hin, dass solche Symbole potenziell missverständlich, da bereits mit einer Bedeutung belegt, sind. Die Transkriptbedeutung dieser Zeichen muss dann immer wieder neu geklärt werden (vgl. Ayaß 2015:520). Während Ayaß sich also gegen den Gebrauch von Zeichen ausspricht, die in anderen Zusammenhängen genutzt werden und dort eigene Bedeutungen haben, werden etwa mathematische Zeichen oder andere Kodierungen von anderen Transkribierenden als praktikabel angesehen und häufig genutzt.

Die Diskussion ließe sich in andere Bahnen lenken, wenn man dazu übergehen würde, Zeichen zu nutzen, die speziell für die Verschriftlichung von Bewegungen entwickelt wurden. Stukenbrock bestimmt „eigens dafür entwickelte symbolische oder in begrenztem Umfang auch [...] ikonische Zeichen“ (Stukenbrock 2009:150) als Mittel der Wahl. Sie schlägt jedoch kein entsprechendes System vor. Diesen Schritt geht die vorliegende Arbeit und kombiniert die Konventionen des im deutschsprachigen Raum etablierten GAT2-Systems mit Elementen aus dem aus der Gebärdensprachforschung kommenden System des Sutton SignWriting, das speziell für die Notation und Transkription von Bewegungen entwickelt wurde.

Gerade für die Untersuchung visueller Kommunikation und Multimodalität bietet die Gebärdensprachlinguistik fruchtbare, innerhalb der Interaktionalen Linguistik jedoch bislang nicht ausreichend rezipierte Ansätze. Gerade in ihrer Konzentration auf und der Erfahrung mit der Arbeit an visuellen Daten bietet die Gebärdensprachforschung jedoch Potenziale, die für die Untersuchung der Multimodalität im Allgemeinen und die Transkription im Besonderen innerhalb der Multimodalen Interaktionsanalyse gewinnbringend sind. Bevor verdeutlicht wird, wie diese unterschiedlichen Konventionen für die hier vorliegende Arbeit adaptiert werden (Abschnitt 12.3), soll zunächst das für die Transkription der visuellen Aspekte genutzte Modell des Sutton SignWriting bzw. Sutton MovementWriting vorgestellt werden.

12.2 Sutton SignWriting

Stellt sich die Frage nach der ‚Verschriftlichung‘ von visuell-leiblichen Aspekten in der Interaktionsanalyse erst seit einigen Jahren im Zuge der verstärkten Arbeit an und mit Videodaten, so ist sie im Bereich der Gebärdensprachforschung und der Kunst, etwa im Bereich des Tanztheaters, ein seit Jahrzehnten

intensiv diskutiertes Thema und Anliegen. Das dahinterstehende Problem gilt aber für alle Disziplinen: Wie können Bewegungen bzw. Gebärden adäquat und nachvollziehbar verschriftlicht werden, und wie müssen Konventionen dafür aufgebaut sein? Um diese Fragen zu bearbeiten, entwickelte Valerie Sutton in den 1970er Jahren das Modell ‚Sutton MovementWriting‘ (SMW) als ein „allgemeines Bewegungsbeschreibungssystem“ (Wöhrmann 2005:7), das ein komplettes und komplexes Notationssystem ist, um sämtliche Körperbewegungen festzuhalten (vgl. Sutton 1973:1). Bewegungen werden dabei mithilfe ikonischer, indexikalischer und symbolischer Zeichen dargestellt und in einem speziellen Partitursystem abgebildet. Die Bandbreite der zu erfassenden Bewegungen führte zu einer Ausdifferenzierung des Sutton-Systems (vgl. Sutton 1982:79): ‚*dance writing*‘ (Verschriftlichung von Tanzchoreographien), ‚*sign writing*‘³ (Verschriftlichung von Gebärden(sprachen)), ‚*mime writing*‘ (Verschriftlichung von Mimik und Gestik allgemein), ‚*sports writing*‘ (Verschriftlichung von Bewegungsabläufen beim Sport, z. B. Eiskunstlauf, Turnübungen, Karate) und ‚*science writing*‘ (Verschriftlichung von Körpersprache im Allgemeinen sowie Bewegungen von Tieren und in Arbeitssettings wie Physiotherapie, Holzfällarbeiten etc.).⁴ Ursprünglich konzipiert, um Tanzchoreographien und damit Bewegungen und dynamische Handlungsabläufe mithilfe von speziellen Bewegungssymbolen (*movement symbols*) zu fixieren und zu verschriftlichen (vgl. Sutton 1973:2 und 1982:79, Hoffman-Dilloway 2011:346), ist es mittlerweile ein System, das vor allem im Bereich der Gebärdensprachforschung zu einer „spezielle[n] Notationsform für alle Gebärdensprachen der Welt“ (Wöhrmann 2005:7) ausgebaut wurde und u. a. für Bibelübersetzungen, Gedichte, Erzählungen und Zeitungsartikel genutzt wird (vgl. Hoffman-Dilloway 2011:345). Dabei ist zentral,

dass Frau Valerie Sutton [...] linguistischen Kategorien zunächst keine Beachtung geschenkt hat. Ihr Ziel war es eben nicht [...] Gebärdensprachen zu notieren. Ihr ging es [...] um die schriftliche Aufzeichnung von Bewegungen im Allgemeinen. Für sie spielt es von daher keine Rolle, ob es sich bei der beschriebenen Bewegung um eine physiotherapeutische Maßnahme, eine Folge von Tanzschritten, ein poetisches Ausdrucksmittel oder um die Bewegungsabfolge beim Karate handelt. Diese ursprüngliche Unabhängigkeit von vorangestellten sprachwissenschaftlichen Überlegungen ermöglichte es ihr, ein rein visuell orientiertes Schriftsystem zu entwickeln, was augenscheinlich ausgesprochen einfach und schnell vom Leser nachvollzogen werden kann. (Wöhrmann: <https://www.gebaerdenschrift.de/>; letzter Zugriff: 19.06.2019)

³Wird auch in der Form ‚SignWriting‘ (SW) genutzt. Übersetzt wird es häufig mit dem Terminus ‚Gebärdenschrift‘ bzw. in der Schreibung ‚Gebärdenschrift‘ (vgl. Wöhrmann 2005:7).

⁴Vgl. <https://dancewriting.org/movement/about/what/> [letzter Zugriff: 28.04.2021]

Gerade in der Möglichkeit, unterschiedlichste Bewegungen und Bewegungsabläufe abzubilden, sieht Sutton eine der Weiterentwicklungen und entscheidenden Vorteile des SMW gegenüber anderen Systemen der Bewegungsnotation (vgl. Sutton 1982:79).⁵ Darüber hinaus sei es leicht zu erlernen und leicht zu lesen (vgl. Sutton 1973:1f). Ähnlich wie andere Systeme arbeitet auch das SMW in seiner ursprünglichen Fassung mit Strichmännchen, die auf einer Art Notenzeile verortet werden (Sutton 1982:80). In diesem System können neben komplexen Bewegungen auch einzelne Teilbewegungen des Kopfes, der Schultern, des Oberkörpers sowie der Füße und Hände abgebildet werden. In etwas abstrakterer Form ist das u. a. noch bei den Zeichen des Sutton SignWriting (SSW) für die Kopfposition einer Person erkennbar. Ein waagerechter Strich steht dabei für die Schulterlinie und ein senkrechter Strich für die Kopfachse (Abb. 12.4).⁶

Zusammengefasst werden die Zeichen des SSW im Internationalen Gebärdenschrift-Alphabet (*International Sign Writing Alphabet*, ISWA).⁷ Die Bandbreite reicht von Zeichen für Fingerhaltungen, Berührungen und Berührungsqualitäten bis hin zu Augenbewegungen, Mundbildern und Körperorientierung.

SW also includes symbols that indicate direction and manner of movement, ‘different kinds of contact between hands and other body parts (for example, striking, holding, brushing, and rubbing), facial expressions, mouth movements, postural shifts, timing, and emphasis’. These resources can be recruited to represent signing with varying levels of detail [...]. (Hoffmann-Dilloway 2018:93)

Das ISWA 2010 umfasst 110 Basissymbole, die unterschiedliche Variationen haben (vgl. Sutton 2010). So ergeben sich auf der Ebene der Mimik u. a. 23 Symbole für unterschiedliche Augenbrauenstellungen, 20 Symbole für die Position der Augen, 16 Symbole für Blick, 33 für unterschiedliche Mundstellungen, 20 für Wangen, 4 für die Nase, 34 für Zungenpositionen, 8 für die Position der Zähne z. B. an den Lippen und 16 für Bewegungen des Kinns (siehe exemplarisch dafür Abbildung 12.5).

⁵Sutton selbst stellt ihr Transkriptionssystem in eine Reihe mit anderen Bewegungsnotationssystemen, etwa die Labanotation, die Choreologie oder die Eshkol-Wachmann Notation (vgl. Sutton 1982:79). Siehe für die Benesh-Notation bzw. Choreologie (Benesh/Benesh 1977) und für die Labanotation (Hutchinson Guest 2005 und Kennedy 2013; siehe auch die Übersicht über weitere Transkriptionssysteme in Frishberg et al. 2012 und Bressem 2013).

⁶<https://www.movementwriting.org/lessons/fullbody/fullbody07.html>). [letzter Zugriff: 28.04.2021]

⁷Abrufbar über die Homepage des Sutton SignWriting: www.signbank.org und <http://www.movementwriting.org/symbolbank/> [letzter Zugriff: 28.04.2021]

Abb. 12.4 Kopfpositionen im SSW

Abb. 12.5 Mimikzeichen nach Sutton (1982:82)

Zusätzlich zur Markierung statischer Positionen nutzt das ISWA Pfeile, um die Bewegungsrichtung und die Bewegungsdynamik einer visuell-leiblichen Handlung anzuzeigen (siehe Abbildung 12.6).

Abb. 12.6 Bewegungspfeile im ISWA (Parkhurst/Parkhurst 2008:24)

Die Gestalt der Pfeile ist abhängig von der Bewegungsrichtung und der Bewegungsebene. Zunächst wird zwischen zwei Ebenen unterschieden: die „Fensterputzebene“ (*wall plain*, d.h. Bewegungen, die parallel zu einer vorgestellten, vor dem Gebärdenden befindlichen Wand bzw. einer senkrechten Fläche ausgeführt werden) und die „Schreibtischebene“ (*floor plain*, d. h. Bewegungen, die parallel zum Boden ausgeführt werden; vgl. Wöhrmann 2005:48–61 und Parkhurst/Parkhurst2008, siehe Abbildung 12.6). Verändert ein Körper bzw. ein Körperteil seine Position nach oben, unten, rechts, links oder diagonal (Fensterputzebene), verwendet man einen *Doppelschaftpfeil* (vgl. Wöhrmann 2005:88). Werden die Bewegungen dagegen parallel zum Boden ausgeführt, verändert sich also die Lage vor oder zurück (Schreibtischebene), werden *Pfeile mit einfachem Schaft* verwendet (vgl. Wöhrmann 2005:111). Dabei ist zunächst unerheblich, ob die Bewegung durch eine Hand, einen Arm, den Kopf oder den gesamten Körper ausgeführt wird. Entscheidend ist bei der Wahl des Pfeilschaftes die Bewegungsebene. Eine genauere Differenzierung wird durch die Art der Pfeilspitze erzielt. Das SSW unterscheidet, ob Bewegungen mit der rechten Hand (*gefüllte Pfeilspitze*), mit der linken Hand (*nicht gefüllte Pfeilspitze*) oder mit beiden Händen (*offene Pfeilspitze*) ausgeführt werden. Pfeile markieren sowohl Bewegungen des gesamten Körpers als auch einzelner Körperteile, z. B. des Kopfes. So zeigen

Doppelschaftpfeile oberhalb des Kopfes ( ) , dass sich der Kopf nach oben und unten bewegt. Dies wird z. B. als Nicken interpretiert. Pfeile mit einfachem Schaft ( ) repräsentieren dann eine Bewegung des Kopfes nach vorn und nach hinten.

Abhängig von der Bewegungsebene und der Stellung z. B. der Hand zu der Person, die die Bewegung ausführt, werden unterschiedliche Positionen und Neigungswinkel dargestellt (siehe Abbildung 12.7).

Abb. 12.7 Handstellungen nach Sutton (Wöhrmann 2005:54 f)

So sind weiß ausgefüllte Zeichen die Markierung dafür, dass dem Produzenten die Handinnenfläche zugewandt ist, und die schwarz ausgefüllten Zeichen markieren, dass der Produzent den Handrücken vor sich hat. Entsprechend ist es sowohl in der Fensterputzebene als auch in der Schreibtiscebene. Neigungen der Hand werden mit halbausgefüllten Zeichen dargestellt. Entsprechend der ikonischen Darstellung der Finger kann zudem abgelesen werden, ob die rechte oder linke Hand repräsentiert wird.

Eine Frage, die sich speziell bei der Verschriftlichung von Bewegungen und anderen visuellen Aspekten der Interaktion stellt, ist die Perspektive, aus der heraus die Zeichen dargestellt werden (die ‚visuelle Origo‘ bzw. das ‚indexikale Zentrum‘ der Darstellung). Aufgrund der ursprünglichen Entwicklung als Verschriftlichung von Choreographien wurde das System des SMW und später des SW zunächst aus der *Rezipientenperspektive* (*receptive viewpoint*) notiert (vgl. Sutton 1982:79), d. h. „the embodied perspective of someone observing another person signing“ (Hoffmann-Dilloway 2018:88; vgl. auch Wöhrmann 2005:14, Parkhurst/Parkhurst 2008:2). Im Laufe der Entwicklung des Systems für die Bedürfnisse der Gebärdensprache wurde es jedoch zunehmend aus *Produzentenperspektive* (*expressive viewpoint*) notiert, d. h. „the embodied perspective of a signer“ (Hoffmann-Dilloway 2018:88, siehe dazu auch Parkhurst/Parkhurst 2008:2). Allerdings ist dieser Wechsel der Perspektive mit Herausforderungen verbunden und daher nicht unumstritten. Die Frage ist z. B., wie Mimik aus der Perspektive des Produzenten dargestellt werden soll bzw. kann, obwohl dieser seine eigene Mimik nicht sehen kann. Vor diesem Hintergrund wird gleichfalls gefragt, wie die Intersubjektivität des Zeicheninventars gewährleistet werden kann (vgl. Hoffmann-Dilloway 2018:90f). Fragen zur Perspektive stellen sich auch, wenn die audiovisuelle Aufnahme eines Gebärdenden in Gebärdenschrift transkribiert werden soll, der Betrachter dem Video gegenüber also eine Rezipientenperspektive einnimmt, die Gebärdenschrift jedoch aus Produzentenperspektive zu lesen und zu schreiben ist (vgl. Hoffmann-Dilloway 2018:95). Die Indexikalität der Zeichen wird dabei besonders deutlich.⁸ Als Begründung für den Wechsel der Darstellungsperspektive wurde u. a. die Entobjektivierung der Gebärdenden genannt, die als Produzierende auch in der Transkription wahrgenommen werden wollten (vgl. Hoffmann-Dilloway 2018:91 f).

Im Gegensatz zu den Ansprüchen von Gebärdenden in ihrem Alltag stehen die Anforderungen, die wissenschaftliche Arbeiten an eine Transkription stellen, wenn sie das Zeicheninventar des ISWA nutzen wollen, denn dabei liegt der Fokus auf der Beobachtung von Bewegungen und deren intersubjektiven Darstellung. Im wissenschaftlichen Kontext werden daher die möglichen Vorzüge der Rezipientenperspektive vor der Produzentenperspektive diskutiert:

Indeed, those using SW for research purposes sometimes frame receptive writing as more appropriate for objective research, due to a sense that it affords more critical

⁸In ihrem Artikel stellt Hoffmann-Dilloway jedoch auch dar, dass der Wechsel der Perspektive für häufig Gebärdenden i. d. R. kein Problem darstellt, die Rotation des Standpunkts vielmehr automatisch vollzogen wird – im Gegensatz zu Personen, die weniger geübt im Gebärden und im Lesen von Gebärdenschrifttexten sind (vgl. Hoffmann-Dilloway 2018:95).

distance. Further, a focus on the receptive visual modalities of signing can be seen as better aligned with the theoretical commitments of disciplines such as conversation analysis, in which 'only those dimensions of action that are publicly available can figure in the interactants' mutual sensemaking, thus, exclude(ing) the kinesthetic phenomena' encoded by expressive writing. (Hoffmann-Dilloway 2018:10)

Hoffmann-Dilloway nutzt die Gebärdenschrift als Analyseinstrument zunächst aus Rezipientenperspektive, plädiert aber im Verlauf ihrer Untersuchungen mehr und mehr dafür, die Produzentenperspektive einzunehmen, auch wenn das für Analysierende bedeutet, dass ein Perspektivwechsel bei der Transkription von Videodaten vorgenommen werden muss.

Stellt man die von Seiten der Multimodalen Interaktionsforschung geforderten Kriterien für Multimodaltranskripte neben das bereits existierende System des SSW, so wird erkennbar, dass ein großes Potenzial im Zeicheninventar des ISWA steckt, das die interaktionale Multimodalitätsforschung bisher nicht entdeckt hat. Durch die Stilisierung und Ikonisierung der tatsächlich ausgeführten Bewegungen umgeht das ISWA z. B. interpretierende Beschreibungen und löst damit eine grundlegende Forderung für die Transkription von visuellen Aspekten der Interaktion ein. Die Errungenschaften des Systems liegen also in der Entwicklung eines konventionalisierten Zeicheninventars, das Bewegungen abstrahiert abbildet, ohne auf verbale Beschreibungen zurückgreifen zu müssen, das dadurch relativ leicht erlernt und rezipiert werden kann und das im Sinne des Zwiebelprinzips unterschiedliche Feinheitsgrade annehmen kann. Die vorliegende Arbeit zeigt im nachfolgenden Kapitel und in den Analysen, wie produktiv die Integration und Kombination des ISWA mit bereits vorhandenen Konventionen zur Transkription verbaler und vokaler Interaktionsressourcen sein kann.

12.3 Transkription in dieser Arbeit

Bei den in der vorliegenden Arbeit zu untersuchenden christlichen Predigten handelt es sich durchgehend um Situationen zeitlicher und räumlicher Ko-Präsenz der beteiligten Personen. Prediger und Gemeinde sind gemeinsam in einem spezifischen Raum (Kirche, Gemeindehaus etc.; siehe Kapitel 14) anwesend und dadurch in der Lage, sich wechselseitig visuell und auditiv wahrzunehmen. Die Aspekte u. a. der körperlichen Ausrichtung sowie der Gestik und der Mimik sind somit relevante Phänomene, die es nicht nur in den Analysen zu beschreiben und auf ihre interaktive Funktion hin zu befragen, sondern auch in den Transkripten abzubilden gilt.

Wie die Vorstellung der derzeit gebrauchten Konventionen zeigt (siehe Abschnitt 12.1), gibt es dazu verschiedene Ansätze, die jedoch alle auf unterschiedlichen Ebenen Defizite aufweisen. Dies betrifft vor allem die Beschreibung von Bewegungen, die Nachvollziehbarkeit und die Lesbarkeit der angefertigten Transkripte. Aufgrund dessen entwirft die vorliegende Arbeit ein eigenes System, das auf bekannten Transkriptionssystemen aufsetzt, diese zusammenführt und zu einem komplexen, modalitätenübergreifenden System ausbaut. Konkret werden die Konventionen des GAT2 mit den Zeichen des internationalen Gebärdensprachalphabets (International SignWriting Alphabet, ISWA) kombiniert (vgl. Dix i.E.). Die Zeichen des ISWA sind z. B. über das Programm ‚SignMaker 2017‘⁹ oder über das Programm ‚SignWriter Studio‘ zugänglich, die wiederum über die Seite des Sutton SignWriting verfügbar sind.¹⁰ In der vorliegenden Arbeit wurden die Zeichen mit SignMaker 2017 einzeln oder in für die vorliegenden Daten angepassten Kombinationen als Bilddatei exportiert und in die Transkripte eingefügt. Zentral ist dabei, dass das hier vorgestellte Transkriptionssystem nicht nur für die in der vorliegenden Arbeit untersuchten Daten konzipiert und zugeschnitten ist, sondern dass es das Potenzial hat, für andere Daten und Forschungsfragen ebenso praktikabel zu sein.

Abbildung 12.8 zeigt, wie die multimodalen Transkripte aufgebaut sind.

Abb. 12.8 Transkriptauszug (eine Person)

Die erste Zeile des Transkripts erfasst die verbalen und vokalen Aspekte der Äußerung sowie Pausen und Stille. Entsprechend den Konventionen nach GAT2 (Selting et al. 2009; siehe elektronisches Zusatzmaterial für eine Übersicht der hier verwendeten Konventionen) wird nach Intonationsphasen transkribiert, die Zeile nummeriert und mit einer Sprechersigle versehen, um den Redebeitrag einer

⁹<https://slevinski.github.io/signmaker/#?ui=de> [letzter Zugriff: 28.04.2021]

¹⁰www.signwriting.org/downloads/ [letzter Zugriff: 28.04.2021]

Person zuordnen zu können. Phasen, in denen keine verbalen Äußerungen produziert werden, z. B. Pausen oder Stille, werden abweichend davon nicht nach den Konventionen des GAT2 transkribiert, sondern in Anlehnung an die Transkription nach Heath et al. (2010:78 f). Die Länge der Pause bzw. der Stille wird hier detailliert in Zehntelsekunden aufgeschlüsselt, die jeweils durch waagerechte Striche repräsentiert werden. Senkrechte Striche (|) grenzen dann die vollen Sekunden ab. Die Zahlen oberhalb der Striche verweisen auf die Dauer des Abschnitts (vgl. Heath et al. 2010:78).

Unterhalb dieser Verbalzeile befinden sich die einzelnen Zeilen für die visuellen Aspekte (Visualzeilen), die in einer Tabelle dargestellt werden. Diese werden nicht mit einer eigenen Zeilennummerierung versehen, aber mit einer eigenen Zeilensigle, die angibt, welche Ressourcenebene repräsentiert wird. Der Transkriptauszug zeigt eine Zeile für die Darstellung der körperlichen Orientierung (Zeilensigle O) und der Position des Kopfes (Zeilensigle K). Zum Zweck des besseren Nachvollzugs werden die einzelnen Modalitätsebenen getrennt dargestellt. Dabei handelt es sich um eine künstliche Trennung, die in der Rekonstruktion und Analyse wieder aufgehoben werden muss. Abhängig von den Forschungsinteressen und den Relevanzen des Materials können unterschiedliche Feinheitsgrade und Komplexitätsstufen dadurch erreicht werden, dass die jeweils notwendigen Visualebenen ergänzt werden. Somit wird das Transkript sowohl auf der Ebene der Verbalität als auch der Visualität dem ‚Zwiebelprinzip‘ gerecht. Folgende Ebenen wurden in der vorliegenden Arbeit verwendet (siehe dazu auch das elektronische Zusatzmaterial für eine Übersicht über die verwendeten Zeichen):

- **Orientierung (Sigle O):** z. B.

Grundsätzlich übernimmt das Transkript die im SSW etablierte Produzentenperspektive. Dies rückt nicht nur die handelnden Personen in das Zentrum der Transkription, sondern hebt methodisch auch unterschiedliche Videoperspektiven auf und sorgt so für eine Angleichung und Vereinheitlichung der Transkripte.

Die Abbildungen 12.9 und 12.10 zeigen jeweils Auszüge aus dem in der vorliegenden Arbeit untersuchten Material. Erkennbar wird, dass in Abbildung 12.9 die Kamera seitlich zum Prediger positioniert, in Abbildung 12.10 hingegen frontal auf ihn ausgerichtet ist. Würde das Visualtranskript nun aus der Perspektive des rezipierenden Forschenden angefertigt, würde sich die Darstellung der visuellen Aspekte mithilfe des ISWA jeweils verschieben und Lesende müssten sich immer neu auf die Leserichtung der Zeichen einstellen. Auch eine Transkription aus der Rezipientenperspektive der anwesenden Personen (hier: der Gemeinde), würde zu komplex, da, wie in Abbildung 12.9 erkennbar, sich der Ausrichtungswinkel und

Abb. 12.9 seitliche Kameraperspektive

Abb. 12.10 frontale Kameraperspektive

die Position der anwesenden Personen auf den Prediger innerhalb der Gemeinde verschiebt. Forschende müssten dann die Entscheidung treffen, welche Person sie exemplarisch auswählen, um deren Orientierung zum Prediger darzustellen. Daher wird für die Transkription visueller Aspekte die Produzentenperspektive eingenommen.

In Fällen, in denen die Ausrichtung der beteiligten Personen und die Veränderung der Interaktionsordnung relevant ist, wird mit achteckigen Symbolen gearbeitet (z. B.) . Der schwarze Balken repräsentiert dabei die Ausrichtung der Körpermitte der jeweiligen Person auf ein ‚Vorne‘, das für das jeweilige Setting bestimmt und festgelegt wird. Für die Ebene der (Körper)Orientierung ist also der (Interaktions)Raum der beteiligten Personen die Bezugsgröße (siehe Kapitel 14). Die visuelle Ebene der Orientierung ist stark vom jeweiligen Material abhängig. In den vorliegenden Daten bildet die Ausrichtung auf die Gemeinde das Vorne des Predigers.

- **Kopf** (Sigle **K**): z. B. und

Die Kopfposition der betreffenden Person kann durch zwei unterschiedliche Zeichenformen visualisiert werden. Zum einen besteht das Zeichen aus einem Kreis, der den Kopf repräsentiert und einem waagerechten Strich, der die Schulterlinie anzeigt sowie einem kleinen Strich, der die Position des Kopfes bezogen auf die Schulterlinie verdeutlicht – beides oberhalb des Kreises (z. B. nach unten gesenkter Kopf:). Zum anderen wird die Bewegung des Kopfes und die Positionsveränderung mit Pfeilen oberhalb des Kreises angezeigt (z. B. für die Bewegung des Kopfes nach rechts. Der Doppelschaftpfeil gibt dann an, dass es sich um eine horizontale Bewegung handelt (Fensterputzebene; siehe Abschnitt 12.2), die offene Pfeilspitze, dass der gesamte Kopf in Bewegung ist). In dem Transkript in Abbildung 12.8 ist erkennbar, dass der Prediger zunächst eine gerade Kopfposition hat ( ) , den Kopf dann nach unten führt ( ) , am Ende dieser Bewegung den Kopf gesenkt hat ( ) , bevor er ihn am Ende der Pause wieder nach oben anhebt ( ).

- **Bewegung** (Sigle **M**): z. B. und , oder

Dynamik und Richtung der Bewegung des gesamten Körpers, die mithilfe von Pfeilen angegeben wird. Entsprechend den Konventionen des ISWA verdeutlichen Pfeile mit einfachem Pfeilschaft Bewegungen nach vorn oder hinten (Schreib-tischebene) und Pfeile mit doppeltem Pfeilschaft Bewegungen nach oben oder unten (Fensterputzebene; siehe Abschnitt 12.2). Die Pfeile werden sowohl für die Bewegung des gesamten Körpers als auch einzelner Körperteile verwendet. Eine genauere Differenzierung erfolgt gemäß des ISWA über die Farbe der Pfeilspitzen (siehe Abschnitt 12.2). Als einzelne Partiturzeile repräsentieren die Pfeile die

Bewegungsrichtung der gesamten Person, z. B. beim Gehen durch den Raum. Handelt es sich um die Bewegung z. B. einer Hand, steht der Pfeil in der dafür angeführten Transkriptzeile.

- **Augenlider** (Sigue **L**): z. B.

Öffnungsgrad der Augen – offen, geschlossen, halb geschlossen etc. Repräsentiert werden die Augen durch Striche oder Halbkreise innerhalb eines Kreises, der wiederum für den Kopf steht. So verdeutlichen zwei nach unten gebogene Halbkreise geschlossene Augen. Ist nur ein Auge geschlossen, wird auch nur ein nach unten gebogener Halbkreis angeführt (z. B. markiert, dass die Person das rechte Auge geschlossen, das linke hingegen offen hat).

- **Arme und Hände** (Sigue **H**): z. B. und

Handpositionen, Kontakt der Hände miteinander oder am Körper etc. Helle Zeichen verdeutlichen, dass die Handfläche zu der Person zeigt, die die Bewegung ausführt, dunkle Zeichen markieren, dass der Handrücken zur ausführenden Person zeigt (siehe Abschnitt 12.2). Dies gilt sowohl für die vertikale als auch für die horizontale Ebene. Dazu gibt es unterschiedliche Kontaktsymbole, die die Art und die Dauer des Handkontakts darstellen. Ein Asterisk etwa markiert, dass sich z. B. zwei Hände berühren. Auch die Berührung der Hand z. B. am Oberkörper wird durch Kontaktsymbole gekennzeichnet.

- **Torso/Oberkörper und Schultern** (Sigue **T**): z. B.

Der Oberkörper wird durch eine schwarze waagerechte Linie repräsentiert, die die Schulterlinie verdeutlicht (siehe Abbildung 12.11).

Die Positionsveränderung der Schultern wird ebenfalls mit Pfeilen angezeigt. Verändert sich nicht nur die Ausrichtung der Schultern, sondern des gesamten Oberkörpers, wird dies durch einen kleinen Kreis mit einem Strich in der Ausrichtung der Bewegung angegeben (siehe Abbildung 12.12).

Abb. 12.11 ISWA-Zeichen für Schultern (Parkhurst/Parkhurst 2010:122)

Abb. 12.12 ISWA-Zeichen für Torso-Bewegung (Parkhurst/Parkhurst 2010:123)

- **Standbilder** (Sigle S)

In der letzten Zeile des Transkripts werden schließlich Standbilder besonders relevanter und markanter Passagen aus dem Video in das Transkript eingefügt. Diese zusätzliche Visualisierung dient zudem dazu, die getrennt dargestellten Ebenen bereits an dieser Stelle zusammenzuführen und deren Zusammenspiel zu einem bestimmten Zeitpunkt der Interaktion deutlich zu machen.

Die Alignierung bzw. Synchronisierung von Zeit und/oder Gesprochenem mit den Zeichen für die Bewegungen und Körperpositionen, erfolgt über die genaue Anordnung untereinander und zusätzlich über Markierungen im Text (fett gedruckte Silbe, auf die die Bewegung/Positionsveränderung erfolgt). Die Dauer

von Bewegungen wird mithilfe von Punkten in der entsprechenden Visualzeile nach dem Bewegungszeichen angegeben. Wird eine Position gehalten, wird kein zusätzliches Zeichen in Form von Punkten oder Strichen verwendet.

Das in dieser Arbeit genutzte System ist nicht nur dazu geeignet die (intrapersonelle) Koordinierung unterschiedlicher Ausdrucksressourcen nur einer Person zu dokumentieren und abzubilden, sondern auch die interpersonelle Koordinierung und Orchestrierung mehrere Personen sichtbar zu machen. Dazu werden die Visualebenen den beteiligten Personen zugeordnet, indem vor die Siglen der Ebenen die Sprechersigle gesetzt wird (siehe Abbildung 12.13). Dies ermöglicht es, die Koordinierung der Personen im Fortlauf der Interaktion zu visualisieren und nachvollziehbar zu machen.

227 mit_nem geDANKen in deinem HERzen;

P1	O	
	M	
	T	
	H	
P2	O	
	K	
S		

Abb. 12.13 Transkriptauszug (zwei Personen)

In dem gezeigten Auszug sind zwei Prediger an der Realisierung des Wortbeitrags beteiligt (Dialog-Predigt; siehe Abschnitt 17.4). Für den ersten Prediger (P1/Person links) wurden in diesem Transkript eine Reihe unterschiedlicher Visualebenen notiert: die körperliche Orientierung (O), die Bewegung im Raum (M), die Position des Oberkörpers (T) und die Haltung und Bewegung der Hände

(H). Dies bedeutet im Einzelnen: Als ‚Vorne‘ des Predigers wird die frontale Ausrichtung auf die Gemeinde festgelegt. P1 ist in einer körperlichen Orientierung also zunächst nach schräg links vorn ausgerichtet (, nicht mit im Transkript). Während P1 das Wort „mit“ (Z. 227) produziert, dreht er seinen Körper leicht nach rechts ( ) sodass er in eine gerade Position gelangt ( ). Gleichzeitig geht er zurück ( ). Dass dies eine Bewegung des ganzen Körpers ist, zeigt nicht nur die eigene Visualzeile an, sondern auch die offene Spitze der Bewegungspfeile. Dabei hält der Prediger die rechte Hand schräg in einer Zeigegeste, bei der der Zeigefinger nach vorn ausgestreckt ist und die Finger die Handflächen berühren ( ). In dieser Handhaltung beschreibt P1 eine Kreisbewegung zu seinem Körper hin ( ). Auf der betonten Silbe von „geDANke“ lehnt er seinen Oberkörper leicht nach vorn ( ) und berührt mit der geöffneten Handfläche der rechten Hand schließlich seinen Oberkörper ( ). Die Transkription besteht hier aus vier Elementen, die entsprechend den Anforderungen der Daten speziell aus dem Inventar des ISWA zusammengestellt wurden: der geöffneten, unausgefüllten Hand, die kennzeichnet, dass die Handfläche auf den Sprecher zeigt, der Anordnung der Finger, die ikonisch die rechte Hand repräsentieren, einem Asterisk, der markiert, dass die Hand den Oberkörper berührt, und einer stilisierten Person, bei der ein kleiner Kreis die Stelle anzeigt, an der die Hand den Körper berührt. Der zweite Prediger (P2/Person rechts) ist zunächst mit seinem Oberkörper nach schräg rechts vorn orientiert (, nicht im Transkript) und hat den Kopf leicht schräg gesenkt ( ). Auf die betonte Silbe von „geDANke“, während P1 seinen Oberkörper nach vorn lehnt, hebt P2 den Kopf an ( ) und bringt ihn in eine gerade Position ( ). Zudem verändert er seine körperliche Ausrichtung, indem er sich nach rechts dreht ( ) und schließlich auf P1 orientiert ist. Die zeitliche Erstreckung der jeweiligen Position bzw. der Bewegung wird durch Punkte in der jeweiligen Visualzeile markiert. Durch die hervorgehobenen Silben im Verbaltranskript können die ISWA-Zeichen noch besser aligniert und zugeordnet werden.

Vorteile aus der Verwendung des Zeicheninventars des ISWA und deren Verknüpfung mit den Konventionen der GAT2-Transkription entstehen vor allem auf folgenden Ebenen:

- Beibehaltung der TCU-Struktur nach GAT2

Die Transkription der verbalen und vokalen Elemente der Interaktion werden weiterhin in Zeilenschreibweise und unterteilt nach Intonationsphrasen notiert. Die

von Selting et al. (2009) entwickelten Konventionen werden übernommen. Auch bereits bestehende Transkripte können so schnell um eine Transkription visueller Aspekte ergänzt werden, ohne dass ihre Struktur oder die Zeilennummerierung aufgebrochen werden muss.

- Reduktion von sprachlichen Beschreibungen

Durch die Nutzung von ikonischen, indexikalischen und symbolischen Zeichen zur Darstellungen der Bewegungen entfällt eine verbale Beschreibung innerhalb der Transkripte. Dadurch erhöht sich der deskriptive Gehalt. Damit bietet das hier vorgestellte System eine innovative Alternative zu bisherigen Auseinandersetzungen mit diesem Problem bei der Transkription von Multimodalität.

- Gewährleistung von Übersichtlichkeit und Lesbarkeit

Trotz der zu erwartenden Komplexität ist das Transkript durch seine kompakte Darstellungsweise für die Lesenden nachvollziehbar und klar strukturiert. Zudem ist es das Ziel, dass auch mit dem System nicht vertraute Personen es nach einer kurzen Einarbeitungszeit verstehen und lesen können.

- Nachvollzug der Orchestrierung unterschiedlicher Modalitäten

Verbale, vokale und visuelle Aspekte der Interaktion werden in einer übersichtlichen Art und Weise eindeutig aufeinander beziehbar. Das Transkript macht auf den ersten Blick erkennbar, welche Ausdrucksressourcen gleichzeitig oder eng aufeinander bezogen von den Interagierenden genutzt werden. Dadurch wird die Synchronisierungsleistung einer Person sowie die Koordinierung und Orchestrierung unterschiedlicher Ausdrucksressourcen auch mehrerer Interagierender aufeinander sichtbar.

- Anpassbarkeit und Erweiterbarkeit

Das System ist variabel an die jeweiligen Daten und die spezifische Fragestellung einer Untersuchung anpassbar. Das Transkript selbst kann in seinem Umfang und seiner Detailliertheit sowohl ausgebaut als auch reduziert werden, da die Zeilen für die Darstellung visueller Aspekte individuell eingefügt werden können.

- Transkription schwer beschreibbarer Aspekte

Durch die *Nutzung ikonischer Zeichen*, die speziell für die Verschriftlichung von Bewegungen entwickelt wurden und bereits für die wichtigsten Ausdrucksressourcen in hoher Zahl und Variation vorliegen, werden auch schwer beschreibbare Bewegungen in den Transkripten darstellbar. Dies ermöglicht einen neuen und detaillierten Blick auf z. B. faziale Gesten (vgl. Dix i.E.).

Neben diesen zahlreichen Errungenschaften und Möglichkeiten, die das vorgestellte System für die Transkription von Multimodalität hat, sind jedoch zum jetzigen Zeitpunkt auch noch Probleme vorhanden, die in der weiteren kritischen Auseinandersetzung bearbeitet werden müssen. Schwierigkeiten entstehen bei der Verknüpfung von GAT2 und dem ISWA u. a. dadurch, dass mit einem Zeichenninventar gearbeitet wird, das für standardisierte Bewegungen (Tanz, Gebärden) entwickelt wurde. Entsprechend muss das Inventar, soll es für alle Datensätze praktikabel sein, ggf. erweitert werden. An welchen Stellen und in Bezug auf welche Ausdrucksressourcen dies tatsächlich nötig ist, müssen weiterführende Forschungen zeigen. Schon jetzt lässt sich aber absehen, dass vor allem bei der Darstellung des Umgangs mit Artefakten sowie den Bewegungen der unteren Körperhälfte eine Anpassung erfolgen muss. Ein weiteres Problem ist, dass es zum jetzigen Zeitpunkt keine etablierte Software gibt, die das System, wie es hier vorgeschlagen wird, erstellen kann. So wurden alle in der nachfolgenden Analyse abgebildeten Transkripte in Word erstellt, nachdem die verbalen und vokalen Anteile mit der Software f4 und Elan transkribiert wurden.

Die vorliegende Arbeit schlägt die hier entwickelte Transkriptionskonvention nicht nur als adäquates System für die vorgenommene Untersuchung vor, sondern als ein System, das dazu geeignet ist, das geforderte allgemeine Modell für die Transkription von Multimodalität zu liefern (vgl. Dix i.E.). An dieser Stelle bedarf es daher weiterer intensiver Auseinandersetzung.

Open Access Dieses Kapitel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Kapitel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.