

Zur Rolle der Gestik beim gemeinsamen Imaginieren

Interaktive Mittel der Gestik zur Herstellung geteilter mentaler Räume

Maya Marx

ayamxram@googlemail.com

Cognitive Linguistics posits that to ensure understanding during conversations, interlocutors need to establish a temporarily shared semantic framework. This implies that individual mental concepts need to be aligned in interaction to the degree that understanding is possible. One case in point where the collaborative process of meaning construction becomes visible are co-constructions, i.e. the construal of one syntactic unit by more than one speaker. While syntactic co-constructions have received considerable attention in interactional linguistics, collaborative interaction at the multimodal and especially the gestural level has received little to no attention in previous research. Therefore, this paper examines dyadic interaction and focuses on co-constructed gestures and how they are used to collaboratively construct meaning. The forms of these co-constructed gestures observed in the data range from exact gestural mirroring to contexts in which speakers adopt and paraphrase previous gestures and meanings. The analysis of the functions of these gestural co-constructions shows that they mainly serve to ground mental spaces in the current speech situation by making them visually perceptible and therefore quotable by other speakers.

Keywords:

mentale Räume, gemeinsames Imaginieren, Blending, Gestik, grounding, shared cognition

1. Einleitung

„Kommunikation [...] funktioniert an der Schnittstelle geteilter Konzeptualisierungen.“ (Zima 2021: 123). Diese der Kognitiven Linguistik zuzurechnende Position impliziert, dass Bedeutungen nicht primär oder ausschließlich als inhärenter Teil sprachlicher Strukturen betrachtet wird, sondern Bedeutungen vielmehr als mentale Vorstellungen konzeptualisiert werden, die aktiv im Gesprächskontext konstruiert und geteilt werden (vgl. Barcelona & Valenzuela 2011: 18). Mit seiner *Theorie der mentalen Räume* bietet Gilles Fauconnier ([1985] 1994; [1997] 1999) ein Modell, das die Grundzüge lokaler Bedeutungskonstruktion während des Denkens und Sprechens erfasst, und das die kognitiv-räumliche Repräsentation sprachlichen Inputs veranschaulicht. Darauf aufbauende Publikationen, wie die von Oliver Ehmer (2011) zeigen, dass die körperliche Ebene von Sprecher*innen eine reliable Ressource zur Etablierung physisch räumlicher Repräsentationen von Vorstellung ist (ebd.: 318), und dass die Theorie der mentalen Räume dem theoretischen Modell für die Erfassung von im Gespräch konstituierter Vorstellung entspricht (ebd.: 14). Zahlreiche weitere Arbeiten (unter anderem Fricke 2007; Liddell 1995; Liddell & Metzger 1998; Stukenbrock 2014) demonstrieren, wie Gesten in Kombination mit sprachlichen Elementen in der Face-to-Face-Interaktion dazu dienen können, imaginierten Objekten eine visuell erfahrbare Komponente zu verschaffen, die der geteilten Konzeptualisierung dient.

Auf diesen Studien aufbauend unternimmt die vorliegende Arbeit den Versuch, in qualitativen Analysen einzelner Sequenzen einer Videoaufnahme herauszuarbeiten, wie Gestik eingesetzt werden kann, um individuelle Vorstellungen zu geteilten zu machen und gemeinsam in der Interaktion Bedeutung zu konstruieren. Zentral für die vorliegende Arbeit ist dabei die Auffassung gemeinsamer Kognition (*shared cognition*) als gleichzeitig aktivierte mentale Repräsentation einer temporär aufgebauten semantischen Struktur (Ehmer 2011: 47f.). Wie Deppermann (2008: 231) argumentiert, bleiben diese mentalen Strukturen nicht grundsätzlich in der mentalen Blackbox der Interaktionsteilnehmer*innen, sondern geteilte Vorstellungen spiegeln sich im koordinierten Handeln von Interaktionspartner*innen wider und können deshalb durch Methoden der Konversationsanalyse herausgearbeitet werden. Für die Auseinandersetzung mit diesem Phänomenbereich bearbeite ich konkret die folgenden Fragen:

- Welche Arten mentaler Räume sind in der Interaktion zu finden und wie verhalten sie sich zum Wahrnehmungsraum der Teilnehmer*innen?
- Wie werden Gesten eingesetzt, um geteilte Vorstellung zu entwickeln?
- Wie kann Verstehen mithilfe interpersonal koordinierter Gestik dokumentiert werden?
- Welche Rolle spielen Modi der Expressivität auf verbaler und nonverbaler Ebene?

Die Datengrundlage der Arbeit ist die Aufnahme einer dyadischen Interaktion, in der die Teilnehmer*innen die Aufgabe bearbeiten, ein Bühnenbild für die märchenhafte Kinderoper *Hänsel und Gretel* des Komponisten Engelbert Humperdinck zu entwerfen. Der Entwurf des Bühnenbilds wurde zunächst im Gespräch geplant und ausgehandelt. Im Verlauf der Aufnahme fertigten die Teilnehmer*innen Skizzen als physischen Abgleich der mentalen Vorstellungsräume an, die stellvertretend als Bühnenbildelemente im Wahrnehmungsraum verteilt wurden. Ort der Aufnahme war ein Raum mit Bühne, dieser bietet gleichzeitig auch die räumliche Ressource für den Entwurf des Bühnenbilds.

Zur Beantwortung der Forschungsfragen, werde ich zunächst darstellen, wie sich der theoretisch-methodische Ausgangspunkt der Arbeit konstituiert (Kap. 2.1) und daraus resultierend erläutern, wie gemeinsame Imagination als Aufbau mentaler Räume in der Interaktion gefasst werden kann (Kap. 2.2). Der theoretische Überblick schließt mit der Darstellung gestischer Dimensionen und der Relevanz der Perspektivität im Gespräch (Kap. 2.3 & 2.4). Da Art und Bedingung der Datenaufnahme speziell und hochrelevant für die Analyse sind, wird das anschließende Kapitel zu Daten und Methoden (3) zusätzlich mit analytischen Aspekten aufbereitet. Der empirische Teil der Arbeit ist in drei Kapitel gegliedert, in denen verschiedene Aspekte multireferenzieller Gestik (4.1), körperlich-gestischer Spiegelung (4.2) und gestischer Versetzung in den Vorstellungsräume (4.3) diskutiert werden.

2. Theoretische Grundlagen

2.1 Imagination und Interaktion

Die vorliegende Arbeit fasst *Imagination* nach Ehmer als „gemeinsam vollzogener Prozess des Aufbaus mentaler Räume“ (2011: 1) im Gespräch. Zentral für diesen Ansatz ist die Auffassung von *Imaginieren* als mentale Versetzung in eine Vorstellung, die nicht als individueller Prozess stattfindet, sondern interaktional modelliert wird und sich als Veränderung der Wahrnehmung auf konversationaler Ebene widerspiegelt (ebd.: 29). Methodisch interaktional ausgerichtet, verbindet diese Arbeit also die Prämisse der Konversationsanalyse, nämlich dass sich Interaktionsteilnehmer*innen fortwährend ihr aktuelles Verständnis der Interaktion durch beobachtbares Handeln signalisieren (Deppermann 2008: 231), mit der Annahme der gebrauchsbasierten Kognitiven Linguistik, die von einer *on-line Bedeutungskonstitution* von Sprache ausgeht (Barcelona & Valenzuela 2011: 4). Dieses Konzept fasst ein Verständnis von Sprache, das stark vom Kontext einer Äußerung abhängig ist und in lokalen Prozessen der Sprachproduktion, -verarbeitung und -interpretation verankert ist. Diesem Ansatz folgend ist

Bedeutung sprachlichen Strukturen nicht inhärent, sondern sprachliche Strukturen gelten als Anreize, Teile semantischer Netzwerke zu aktivieren (*prompts to construct meaning*, Turner 1991: 26). Dieser Prozess der semantischen Interpretation wird aber – zumeist – nicht versprachlicht und der *backstage cognition* zugeordnet (Zima 2021: 103). In der Verbindung dieser Ansätze bedeutet das, dass situative, mentale Prozesse in der Sprache an sich nicht explizit erschließbar sind, in der Interaktion aber fortlaufend Prozesse der wechselseitigen Bezugnahme sichtbar werden, die mit Methoden der Konversationsanalyse beschrieben werden können (Ehmer 2011: 51). Der Aufbau einer temporär gemeinsamen semantischen Struktur (*shared cognition*) wird durch die situative, wechselseitige Aktualisierung von Hintergrundwissen bzw. Annahmen analysierbar. Für übereinstimmend aufgebaute mentale Räume müssen die Interaktionsteilnehmer*innen diese Annahmen aber individuell und *backstage*, auf der Grundlage neuen situativen Inputs, aktualisieren. Im Folgenden erläutere ich die für diese Arbeit zentralen theoretischen Aspekte der Imagination im Gespräch bzw. ihrer kognitiv-linguistischen Entsprechung der lokalen mentalen Räume (dazu auch ebd.: 13). Da in der empirischen Analyse der Fokus auf der Gestik liegt, schließe ich ein Kapitel zur Betrachtung einiger Phänomene und Kategorien gestischer Mittel als multimodaler Beitrag des Aufbaus gemeinsamer Vorstellungen an.

2.2 Mentale Räume als situatives gemeinsames Imaginieren

Die Theorie der mentalen Räume (*mental space theory*) ist ein semantischer Ansatz aus der kognitiven Linguistik und geht auf die Arbeiten von Gilles Fauconnier ([1985], 1994; [1997], 1999) zurück. Begriffspakete repräsentierend, die wir beim Denken und Sprechen konstruieren (Fauconnier & Turner 2002: 241), entstehen mentale Räume prinzipiell lokal, hierarchisch aufeinander aufbauend und on-line in der Interaktion (vgl. Zima 2021: 104). In seiner Arbeit zum gemeinsamen Imaginieren im Gespräch als konversationelle Herstellung mentaler Räume, spricht Ehmer auch von „szenische[n] Vorstellungen“, die „ad hoc im Gespräch“ aufgebaut werden (2011: 32). Dabei wird auf konzeptuelle Erfahrungs- und Wissensstrukturen zurückgegriffen, sog. *Frames* (Fillmore 1985). Sie strukturieren mentale Räume und dienen ihnen als „statische Ressource“ (Ehmer 2011: 33) des Wissens. Im Gesprächsverlauf wird dieses Hintergrundwissen nach Bedarf aktiviert und einzelne Elemente temporär evoziert. Im fortschreitenden Diskurs können so situativ Aspekte aktivierter Frames in mentalen Räumen ausgebaut und als kohärenzschaffendes Mittel von Raum zu Raum aufrechterhalten und erweitert werden (vgl. Zima 2021: 104).

In der Analyse von Gesprächsausschnitten bieten bestimmte sprachliche Konstruktionen Anhaltspunkte, um in der Interaktion entstehende mentale Räume zu identifizieren und zu kategorisieren. Diese *space builder* (Fauconnier [1985] 1994: 17) eröffnen spezifische mentale Räume bzw. sie verschieben die Aufmerksamkeit in den mentalen *Fokusraum* mit der temporär relevantesten Information. Typische *space builder*-Konstruktionen sind z.B. Adverbien. Je nach Tempus und Modus evozieren sie unterschiedliche Kategorien mentaler Räume. Temporale Adverbien, wie *damals* oder *demnächst*, können Vergangene-Event-Räume (*past event spaces*) bzw. Zukünftige-Event-Räume (*future event spaces*) evozieren. Konjunktiv-Konstruktionen dagegen situieren Vorstellungen typischerweise in einem hypothetischen oder konditionalen Raum (*hypothetical space*), der die Bedeutung des sprachlichen Inputs als ‚nicht real‘, bedingungsabhängig oder fiktional deklariert (vgl. zu Kategorien und Aufbau mentaler Räume auch Zima 2021: 108–114). In Erzählungen und mit Bezug auf die aktuelle Realität der Gesprächssituation können mentale Räume so in der Vergangenheit, Zukunft oder auch alternativen Dimensionen verankert werden (vgl. Ehmer 2011: 56f.). Die Kategorien mentaler Räume sind vielfältig, häufig fließend und können hier kaum erschöpfend erläutert werden. Mit Bezug auf die Datengrundlage der Arbeit werde ich hier aber auf zwei Arten von Räumen näher eingehen, die auch in der *mental space theory* einen besonderen Status einnehmen (vgl. Zima 2021: 108). Das ist zunächst der *reality space*, d.h. die geteilte Kommunikationssituation, mit ihren physisch-räumlichen Gegebenheiten ebenso wie mit den Umständen der Interaktion. In der in dieser Arbeit analysierten Gesprächsaufnahme ist dieser *reality space* der Bühnenraum mit seiner mobiliaren Ausstattung sowie dem geteilten Wissen darüber, dass der Sinn und Zweck der Interaktion das gemeinsame Entwerfen eines Bühnenbilds für die Kinderoper *Hänsel und Gretel* ist. In der Aufnahme ist der *reality space* die wichtigste Bezugsgröße für alle auf ihm aufbauenden mentalen Räume. In der Theorie kann ausgehend vom *reality space* ein *Basisraum* (*base space*) initiiert und entwickelt werden. Er dient in der Regel als Ausgangspunkt für hierarchisch auf ihm aufbauende, kohärent vernetzte mentale Räume (vgl. Ehmer 2011: 35). Für die Gesamtsituation der Aufnahme erscheint als Basisraum die situative Vorstellung der Teilnehmer*innen von sich als Bühnenbildbauer*innen, die einen ‚echten‘ Entwurf des Bühnenbilds erarbeiten. Statische Wissensressource und ständig mindestens latent aktiver Frame bildet die Oper *Hänsel und Gretel* bzw. die Vorstellung eines Bühnenbilds der Oper, die beide Teilnehmer*innen schon in mindestens einer Inszenierung gesehen haben.¹

¹ Die Teilnehmer*innen studieren klassischen Gesang an derselben Universität und sind sowohl mit universitätsinternen als auch -externen Inszenierungen der Oper *Hänsel und Gretel* vertraut. Im Verlauf der Aufnahme unterhalten sich die Teilnehmer*innen mehrfach darüber, welche Inszenierungen der Oper sie bereits gesehen haben und wie jeweils das Bühnenbild interpretiert war.

In der Aufnahme werden einmal evozierte Vorstellungsräume häufig miteinander verflochten und Konzepte aus verschiedenen mentalen Räumen neu zusammengefügt. Für die Analyse komplexer Vermischungen von Vorstellungsräumen entwickelten Gilles Fauconnier und Marc Turner die *Theorie des konzeptuellen Blendings* (*blending theory*). In einem Blend (*blended space*) sind Bedeutungen aus zwei oder mehreren *input-spaces* enthalten, die erst durch ihre Komprimierung oder *Überblendung*² entstehen (Fauconnier & Turner 2002). An der Schnittstelle der *Theorie der mentalen Räume* und der *konzeptuellen Metaphertheorie* (Lakoff & Johnson 1980) geht die Theorie des konzeptuellen Blendings davon aus, dass durch den Prozess der Integration unterschiedlicher konzeptueller Strukturen emergente Bedeutung entsteht. Anhand des Beispiels *Demokratatur*³, eine Amalgamierung aus *Demokratie* und *Diktatur*, kann erklärt werden, wie so eine Fusion von Bedeutungskomponenten auf Text- bzw. Wortebene aufgebaut sein kann. Die Verschmelzung von *Demokr-* (aus dem Lexem Demokratie) + *-at* (Fusionsschnittstelle) + *-ur* (aus dem Lexem Diktatur) führt zu einer emergenten, neuen Form mit neuer Bedeutung, die nicht in den Ursprungslexemen kodiert ist. Beide Lexeme verweisen auf den Bereich *Staatsform*, referieren jedoch auf grundlegend unterschiedliche politische Systeme. Im Fall der Diktatur auf eine Herrschaft durch eine nicht frei gewählte Einzelperson; im Fall der Demokratie auf eine Volksherrschaft. Diese beiden Strukturen repräsentieren die Inhalte der *input-spaces*. Bei der Übertragung der beiden Ausgangsstrukturen in den Blend werden die Elemente *Volk* und *Staatsoberhaupt* so in eine neue Struktur integriert, dass die neue Bedeutung *Scheindemokratie* oder *Diktatur, die sich als Demokratie tarnt* entsteht.

² Den Begriff *Überblendung* übernehme ich hier nach Zima (2021: 125), wo er vor allem als *Vermischung* oder *Komprimierung* von mindestens zwei *Input-Spaces* zu einem neuen mentalen Raum eingeführt wird. Ich verwende *Überblendung* im Folgenden und im Sinne meiner empirischen Beobachtungen teilweise aber wörtlicher eher als *Überlagerung* und *Doppelbesetzung*.

³ Das Beispiel ist in verkürzter Version übernommen von Graumann 2013: https://www.degruyter.com/database/WSK/entry/wsk_id80ef7232-ced1-49ad-8c1d-9cce2f0e227b/html.

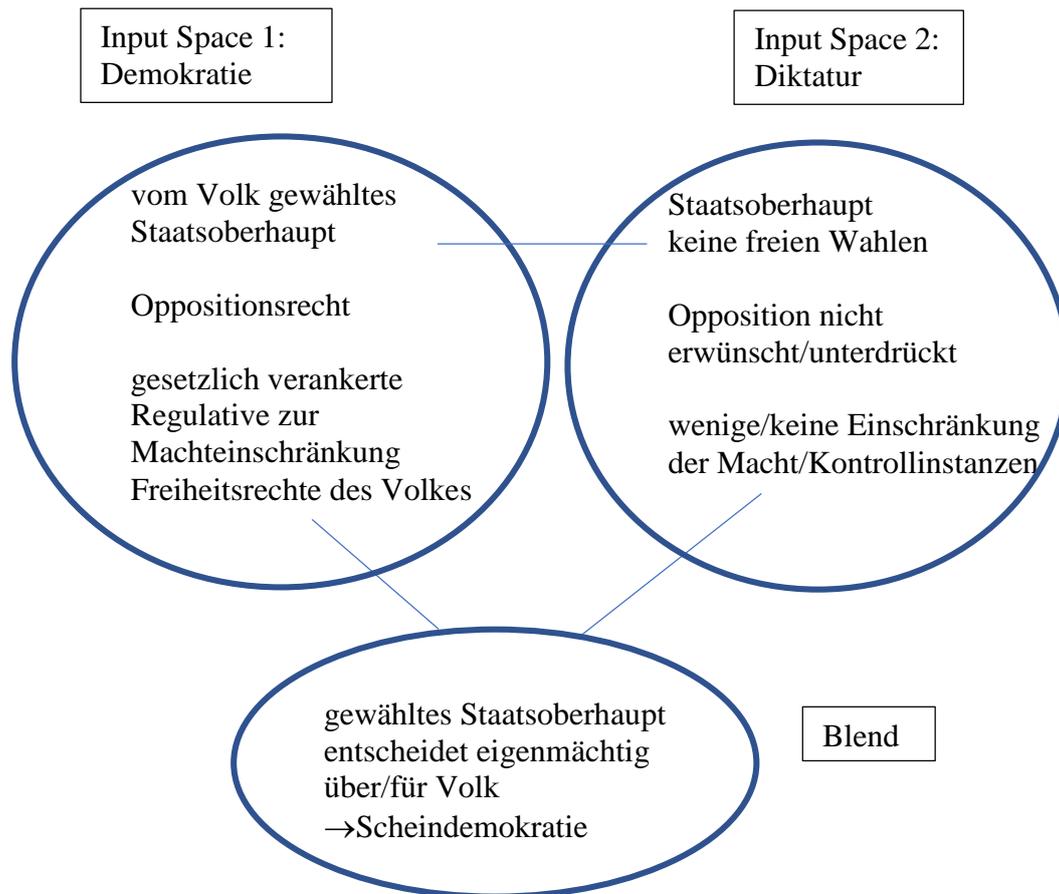


Abbildung 01: Darstellung des Demokratur-Blendingmodells (angelehnt an Graumann 2013; Modelldarstellung adaptiert nach Fauconnier & Turner 2002)

Während Blends in der Regel besonders zahlreich im kreativen Sprachgebrauch (humoristische Anspielungen, Metaphern, Phraseologismen usw.) zu finden sind (Zima 2021: 125), spielt die ihnen zugrundeliegende Theorie für diese Arbeit insofern eine Rolle, als dass sie unter anderem auch die Ko-Aktivierung der Informationen von *reality space* und lokalen mentalen Räumen erklären kann (dazu ausführlicher in Kap. 3).

Neben den oben besprochenen verbalen Mitteln zur Evozierung mentaler Räume, können zur Kontextualisierung individueller kognitiver Strukturen auch körperliche Mittel eingesetzt werden. Die Theorie der *grounded mental spaces* (Liddell 1995; Liddell & Metzger 1998) und der daran anknüpfenden *grounded blends* (Liddell 1998) wurde entwickelt, um unterschiedliche Arten gestisch-körperlicher Repräsentationen von Vorstellungsräumen zu erfassen. Dabei handelt es sich um eine Theorie, die grundlegend multimodale und vor allem gestische Aspekte mit der Repräsentation mentaler Räume in der Interaktion verbinden will. Die Repräsentation funktioniert in zwei Richtungen: Der Theorie folgend gleicht jeder Vorstellungsräum zunächst seiner real-physischen Entsprechung. Mit dem Beispiel eines Stiftes, der vor einer Person liegt und der dank der mentalen Repräsentation des Wahrnehmungsraumes auch noch im Dunkeln

gefunden werden kann, wollen Liddell und Metzger dieses Konzept verdeutlichen (1998: 662). Dem Beispiel folgend kann die Person im plötzlich dunklen Raum und ihres Sehsinns beraubt, den Stift deshalb problemlos finden, weil sie ein mentales Abbild des Raumes abrufen kann, das den genauen Standort des Stiftes in diesem Raum beinhaltet. Des Weiteren kann einem mentalen Raum durch gestische Ausarbeitung bzw. Repräsentation seiner Aspekte eine physische, sinnlich wahrnehmbare und damit im ‚realen‘ Wahrnehmungsraum verankerte Komponente zukommen. So können Sprecher*innen durch Bewegungen ihres eigenen Körpers, Handlungen von Figuren oder Szenen demonstrieren und diese für die anderen Interaktionsteilnehmer*innen visuell erfahrbar machen. Durch das *grounding* erhalten mentale Räume „wirklichkeitsähnliche[n] Charakter“ (Ehmer 2011: 92) und Interaktionsteilnehmende referieren bspw. mit Zeigegesten auf mentale Entitäten, als wären sie physisch präsent (Liddell 1995: 27). Mithilfe von Gesten können mentale Räume in den physischen *reality space* ‚gegeben‘ bzw. in diesen projiziert werden (vgl. Ehmer 2011: 91). Kognitive Elemente können so in einem symbolischen Raum (*token space*⁴) – häufig durch gestikulierende Hände unmittelbar vor dem Körper – oder in einem stellvertretenden und körperlich lebensgroß ausdifferenzierten Raum (*surrogate space*) gezeigt und im *reality space* verankert werden (Liddell 1995). Dabei wird die Bedeutung des *reality space* konzeptuell überblendet, sodass die Evozierung eines *surrogate* oder *token space* in der Regel auch einen Blendingprozess einleitet (vgl. ebd.: 29). In Fällen, in denen Elemente aus der unmittelbaren physischen Umgebung mit Elementen aus einem mentalen Raum miteinander überblendet werden, spricht Liddell auch von einem *grounded blend* (1998: 291). Mentale Repräsentationen von Elementen, Abläufen und Figurenhandlungen können in einem *grounded blend* unmittelbar wahrnehmbar und teilbar gemacht werden und das durch Elemente der ‚realen‘ Wahrnehmungssituation.

Ebenfalls an der Schnittstelle multimodaler Konzepte und der Erfassung konversationeller Herstellung mentaler Räume bewegt sich das Konzept der Demonstration, das auf einen Aufsatz von Clark und Gerrig (1990) zurückgeht. Sie fassen *Demonstrationen* als Darstellung von Ereignissen, Gegenständen, Zuständen und Prozessen. Dabei können diese demonstrativen Handlungen sowohl auf rein verbaler Ebene (bspw. Zitate), häufiger aber auf körperlich-gestischer Ebene stattfinden und funktional die Kommunikation von schwer Beschreiblichem übernehmen (Clark & Gerrig 1990: 786). Aus ihrem Konzept arbeitet Ehmer zentral heraus, dass eine *Demonstration* eine direkt sinnlich wahrnehmbare Verkörperung einer Figur oder

⁴ Im Sinne der Einheitlichkeit habe ich mich dafür entschieden in dieser Arbeit alle englischen Begriffe in Minuskeln und grundsätzlich kursiv zu schreiben. Daher ist an dieser Stelle anzumerken, dass die Begriffe *Token Space* und *Surrogate Space* bei Liddell (1995) mit Majuskeln versehen sind. (Bei Begriffen wie *Frame* oder *Face-to-Face-Interaktion*, deren Status als Entlehnung in den deutschen Wortschatz fortgeschrittener ist, übernehme ich auch die Regeln der deutschen Rechtschreibung und verzichte auf durchgängige Kursivierung.)

Rolle ist. Sprecher*innen können so aus Perspektive einer anderen Person handeln bzw. eine Figurenhandlung inszenieren (2011: 75). Auf verbaler Ebene definiert er das Demonstrieren einer Sprechhandlung auch als *animierte Rede* (vgl. ebd.: 62). Dadurch, dass demonstrierte Handlungen – verbal und körperlich – die gedankliche Versetzung in einen Vorstellungsraum erfordern, erfasst Ehmer *Demonstrieren* gleichbedeutend mit der Eröffnung und Elaboration eines mentalen Raums (ebd.). Funktional ähnlich, aber perspektivische Distanz zu einer Figur wahrend, definiert er den *symbolisierenden* Sprechmodus. Symbolisiert ein*e Sprecher*in, so *beschreibt* er*sie eine Handlung, anstatt sich in die (Figuren-)Handlung hineinzusetzen und ‚als die Figur‘ eine Aktion auszuführen. Durch die Aktivierung semantischer Frames können auch im symbolisierenden Sprechen mentale Räume aufgebaut und geteilt werden (ebd.: 99). In seinen Analysen zu gemeinsamem Fantasieren in Face-to-Face-Interaktionen nutzt Ehmer diese beiden Kategorien vor allem, um zu erfassen, welche Perspektive Sprecher*innen in der Herstellung eines mentalen Raums einnehmen und wie der Wechsel der Modi vom symbolisierenden Beschreiben in die Animation einer Figur im sequenziellen Verlauf eines Gesprächs stattfinden kann.

Wie zuvor dargestellt, befasst sich die Theorie der mentalen Räume also damit, wie und welche kognitiv-semantischen Räume durch sprachliche Strukturen hervorgerufen werden, um Bedeutung des sprachlichen Inputs zu konstruieren. Neben den on-line im Gespräch erfassbaren sprachlichen Mittel zur Konstituierung *von* und Versetzung *in* Vorstellungsräume kann aber, wie oben angemerkt, auch die körperliche Ebene bei der Kontextualisierung mentaler Räume helfen. Um zu verstehen, welchen Stellenwert diese Ebene für den Aufbau geteilter semantischer Strukturen haben kann, erläutere ich im Folgenden einige zentrale Konzepte der Gestenforschung sowie einiger deiktischer Aspekte der Gestik.

2.3 Zur Rolle der Gestik

Die Gestenforschung ist ein seit langem etabliertes Forschungsfeld an der Schnittstelle von Psycholinguistik, Semiotik und Konversationsanalyse (umfangreiche Überblicke bieten hier Ehlich 1979, Fricke 2007 oder Kendon 1980 & 2004). Auch die kognitiv linguistische Forschung widmet sich seit einigen Jahren verstärkt Aspekten der Multimodalität und insbesondere auch der Perspektivvermittlung durch Gestik (u.a. Hoffmann 2017; McNeill 1992 & 2007; Mondada 2014; Stukenbrock 2014; Zima & Bergs 2017). So werden körperlich-gestischen Elementen eine zentrale Rolle zugeschrieben, wenn in Kommunikationssituationen eine individuelle Wahrnehmung geteilt und konkrete Aspekte eines Wahrnehmungsraumes in den Fokus der gemeinsamen Aufmerksamkeit rücken sollen (vgl. Hausendorf 2003: 268).

Zu körperlichen Konstituierung mentaler Räume werden neben der Gestik mit Gliedmaßen aber auch Mimik, Körpertonus, -haltung und -bewegung sowie Blickbewegung (Ehmer 2011: 84) genutzt. Da im empirischen Teil dieser Arbeit der Fokus auf der Gestik liegt, setze ich jedoch in diesem definitorischen Überblick einen Schwerpunkt auf gestische Aspekte und betrachte andere non-verbale Mittel darüber hinaus punktuell und in ihren jeweiligen Sequenzkontexten. An dieser Stelle soll aber vermerkt sein, dass die aufgezählten multimodalen Ausdrucksressourcen in der Regel ko-expressiv gebraucht werden und die folgenden Dimensionen der Gestik aus Sicht der Interaktionsanalyse kaum isoliert betrachtet werden können. Auch die strikte Klassifizierung einzelner Gesten wird in der Forschung häufig als wenig sinnvoll angesehen (dazu Ehmer 2011; Kendon 2004; McNeill 2007), da in der Interaktion nicht selten verschiedene Formen und Aspekte von Gesten kombiniert werden. Die folgende Kategorisierung bestimmter, d.h. für die nachfolgende Studie relevanter Gestentypen dient also vor allem Beschreibungszwecken.

Ein zentrales Mittel für die Darstellung imaginierter Objekte ist die *ikonische Geste* (vgl. Ehmer 2011: 89). Sie hat vor allem Abbildfunktion, d.h. sie dient dazu, konkrete Gegenstände oder Ereignisse gestisch darzustellen. Für die Interpretation einer ikonischen Geste ist vor allem die Analogiebildung zwischen körperlicher Präsentation und abgebildeter Größe des darzustellenden Gegenstands relevant (Stukenbrock 2015: 21). Mit ihr artverwandt, aber grundsätzlich abstraktere Konzepte abbildend, ist die *metaphorische Geste* (ebd.: 22). Entkoppeln sich ikonische – mitunter auch metaphorische – Gesten auf der Skala der redebegleitenden Gesten von ihrem verbalen Aspekt, wird teilweise von *Pantomime* gesprochen, einer nicht konventionalisierten Kategorie, die narrative Eigenschaften auf gestischer Ebene besitzt und obligatorisch ohne Sprache auskommt (McNeill 2007: 5). Dagegen werden *Taktstockgesten* fast immer redebegleitend eingesetzt. Sie stellen rhythmische Auf-und-Ab-Bewegungen dar, die in der Regel prosodisch mit Akzentsilben korrelieren und so betonende oder unterstreichende Wirkung erzielen (Zima 2021: 264).

Da die vorliegende Arbeit auch die interpersonelle Koordinierung körperlich-visueller und verbaler Ausdrucksressourcen im Raum zum Gegenstand hat, ist vor allem die *Zeigegeste* als zentrales Referenzmittel in der Aufnahme zu finden. Zeigegesten sind *bezugnehmende Gesten*, die auf Objekte, Personen und im Wahrnehmungsraum verweisen können und häufig auch als *deiktische Gesten* bezeichnet werden (ebd.: 263). Für die Analyse von Zeigegesten ist die *Perspektive* der Sprecher*innen zentral (Ehmer 2011: 16). Mithilfe bestimmter sprachlicher Ausdrücke (*Deiktika*) können perspektivische Aspekte in einer Sprechsituation kontextualisiert werden (Stukenbrock 2015: 2). Die Dimensionen und Funktionen von Zeigegesten,

Sprecher*innenperspektive und Raumrelevanz lassen sich aufschlussreich durch die *Deixistheorie* nach Karl Bühler (1934) erläutern.

2.4 Deiktische Aspekte der Gestik

Bis heute für die Deixisforschung wegweisend ist Karl Bühler und seine 1934 erschienene *Sprachtheorie*. In ihr argumentiert Bühler für eine enge Verflechtung von Deixis und Gestik, denn das Verständnis deiktischer Ausdrücke sei notwendigerweise an „sinnliche Zeighilfen“ gebunden (Bühler [1934] 1999: 93). Er klassifiziert drei Modi des Zeigens und unterscheidet zunächst den sinnlich im unmittelbaren Zeigfeld, d.h. Wahrnehmungsraum, zugänglichen Modus der *demonstratio ad oculus* vom *anaphorischen Zeigen*. Anaphorische Deixis erweitert das Zeigfeld im diskursreferenziellen Gebrauch über das direkt Demonstrierbare hinaus und kann damit auf Elemente Bezug nehmen, die im Kommunikationskontext verankert sind. Des Weiteren spricht er von der *Deixis am Phantasma*, einer deiktischen Praktik, die auf Phänomene verweist, die der sinnlichen Wahrnehmung nicht zugänglich sind, auf eine Vorstellung oder Imagination Bezug nimmt und diese so – zumindest vermeintlich – teilbar macht (ebd.: 121–140). Das impliziert, dass auch deiktische Gestik nicht auf Referenz im *reality space* beschränkt ist, sondern in Kombination mit der Verbalebene dazu dienen kann, Zugang zu mentalen Räumen und imaginierten Objekten zu schaffen (dazu auch Stukenbrock 2014: 88). Den zentralen Bezugspunkt aller Modi des Zeigens bildet dabei das deiktische Zentrum oder die *Origo*. Die *Origo* – bei Bühler auch das *hier-jetzt-ich-System der subjektiven Orientierung* – ist dabei nicht im aktuellen Wahrnehmungsraum des*der Sprecher*in fixiert, sondern bspw. auch in den Vorstellungsraum verschiebbar (Bühler [1934] 1999: 102).

Da in den in dieser Arbeit analysierten Daten das deiktisch-gestische Navigieren zwischen Wahrnehmungsraum und Vorstellungsraum ständig ausgehandelt und aktualisiert werden muss, spielt auch die origorelative *Lokaldeixis* nach der Definition von Ellen Fricke (2007) eine wichtige Rolle. In Form deiktischer Ausdrücke und/oder Zeigegesten dient die Lokaldeixis dazu, die Aufmerksamkeit des*der Adressat*in in Abhängigkeit von der jeweiligen *Origo* auf den zu lokalisierenden Gegenstand zu lenken und das so, dass Sprecher*in und Adressat*in eine möglichst übereinstimmende mentale Repräsentation des Referenzobjekts mit ähnlicher räumlicher Orientierung aufbauen (Fricke 2007: 86). Wichtig ist auch hier, dass, wie im Fall der *Deixis am Phantasma*, die *Origo* sich nicht nur im ‚realen‘ Wahrnehmungsraum, sondern auch im Vorstellungsraum befinden kann. Das heißt auch, dass wenn durch Zeigegesten ein mentaler Referent explizit, also sinnlich wahrnehmbar gemacht wird, *Deixis am Phantasma* auch *ad oculus* agieren kann. Dies ist immer dann der Fall, wenn ein mentaler Raum auf den

Wahrnehmungsraum projiziert wird (vgl. dazu auch Liddell 1995: 285). Nach Stukenbrock besteht diese perzeptorische Funktion einer deiktischen Zeigehandlung vor allem darin, ein „im umgebenden Raum vorhandenes Phänomen in eine gemeinsame und als gemeinsam gewusste Wahrnehmungserfahrung zu verwandeln“ (2015: 85).

Voraussetzung für das Gelingen einer Zeigehandlung ist eine *fokussierte Interaktion* (Goffmann 1963), das den Zustand absichtsvoller, wechselseitiger Orientierung von Interaktionsteilnehmenden bezeichnet. In einer fokussierten Interaktion kann aus der Perspektive der*s Zeigenden mittels Blick- und Körperausrichtung ein *Verweisraum* eröffnet werden, der das *Zeigziel* beinhaltet. Aus Perspektive der Adressat*innen ist das der *Suchraum*, auf den sich Interaktionsteilnehmer*innen unter Zuhilfenahme der verweisenden verbalen und visuellen Orientierungshilfen der*des Zeigenden, ausrichten (Stukenbrock 2015: 56).

Wie schon dieser kurze Überblick zeigt, ist die Forschung zu Gestentypen und auch der Rolle der Gestik für die Deixis ein hinreichend untersuchtes Phänomen. Kaum empirische Studien beschäftigen sich dagegen bisher mit *gemeinsamer, interaktiver, körperlich-gestischer* Handlungen, im Sinne der gemeinsamen Konstruktion von Bedeutung in der Interaktion. Alle hier diskutierten empirischen Arbeiten beziehen sich auf Beobachtungen und Erkenntnisse der Trennung bzw. wechselseitigen Beziehung von Produktion und Wahrnehmung. Ein Phänomen aus dem Bereich der *On-line-Syntax*⁵ (Auer 2000) weist aber Ähnlichkeiten zu dem hier fokussierten Phänomenen auf: die *Ko-Konstruktion*. Ko-Konstruktionen bezeichnen die *kollaborative* Produktion *einer* syntaktischen Gestalt oder Redeeinheit von mindestens zwei Sprecher*innen (Brenning 2015:1). Alle Sequenzen, die in dieser Arbeit untersucht werden, stimmen darin überein, dass diese Form kollaborativen sprachlichen Handelns (auch) auf gestischer Ebene stattfinden. In den Ausschnitten steigen die Teilnehmer*innen in die beobachtbare gestische Ausarbeitung der mentalen Räume mit ein, wiederholen, paraphrasieren und erweitern sie. Ein weiteres Ziel der Arbeit ist es zu beschreiben, wie diese *gestischen Ko-Konstruktionen* im Einzelnen aussehen und welchem Zweck sie beim gemeinsamen Imaginieren dienen können.

⁵ Auch hier ist die Bezeichnung *on-line* zentral auf ihre echtzeitliche, kontextsensitive Beschaffenheit ausgelegt. Im speziellen Fall der Syntax bezeichnet der Begriff *On-line-Syntax* nach Peter Auer (2000) die Integration verschiedener Aspekte der Zeitlichkeit gesprochener Sprache. Zentral ist hier das Verständnis von mündlicher Syntax in ihrer Emergenz von flüchtigen, irreversiblen und synchronisierten sprachlichen Strukturen, deren kognitive und interaktive Bearbeitung in Echtzeit des Gesprächs stattfindet.

3. Daten und Methode

Die Datengrundlage der Arbeit bildet eine Aufnahme, die auf eine gemeinsame Idee von Barbara Laner und mir zurückgeht und im Rahmen von Forschung zu Blickverhalten in der Interaktion des MobEyeLab der Universität Freiburg erhoben wurde. Bei der Aufnahme handelt es sich um eine 62 Minuten und 45 Sekunden lange Audio- und Videoaufnahme dyadischer, aufgabenbasierter Interaktion, die in einem Veranstaltungssaal mit Bühne erhoben wurde. Die Aufgabe der Teilnehmenden⁶ – und damit der zentrale Aspekt der Aufnahmen – bestand darin, gemeinsam ein Bühnenbild für die Kinderoper *Hänsel und Gretel* von Engelbert Humperdinck⁷ für den zur Verfügung stehenden Bühnenraum zu entwerfen. Der Interaktionsrahmen ist also nicht unabhängig von der Beschaffenheit und Ausstattung der gegebenen Räumlichkeit und Aufgabenstellung. Daher werde ich diese im Folgenden – und nach der Vorstellung der technischen Details der Datenerhebung und Analysemethoden – ebenfalls kurz erläutern und im Hinblick auf ihre speziellen Anforderungen analysieren.

Die Interaktion der Teilnehmenden wurde mit einer externen, frontal vor der Bühne platzierten Kamera aufgenommen. Zudem trugen die Versuchspersonen während der gesamten Aufnahme mobile Eyetracking-Brillen des Herstellers Tobii⁸, die eine freie Bewegung im Raum ermöglichen, während sie präzise aufzeichnen, worüber gesprochen und wohin geschaut wird. Mit zwei verschiedenen Kameras ausgestattet, zeichnet die Brille mit Infrarotlicht die Pupillenbewegung beider Augen auf, während die Szenenkamera die Umgebung aufnimmt, die in etwa der Perspektive der*des Trägerin*s entspricht. Nach der Erhebung wurden aus den Daten aller Aufnahmegeräte synchronisierte Split-Screen-Videos erstellt, die sowohl die frontale Perspektive auf die Bühne zeigen als auch Aufzeichnungen des jeweiligen visuellen Fokus der Interaktionsteilnehmenden mittels farbiger Kreise (Cursor) im Video.

⁶ Die Namen der Interaktionsteilnehmenden wurden anonymisiert. In dieser Arbeit nenne ich sie Tim (männlicher Teilnehmer) und Ana (weibliche Teilnehmerin). Schriftliche Einverständniserklärungen der Teilnehmenden zur Analyse und Veröffentlichung von Transkripten und Standbildern liegt der Versuchsleitung vor, wurden dieser Arbeit aus Gründen des Datenschutzes aber nicht beigelegt.

⁷ *Hänsel und Gretel* ist eine spätromantische Oper in drei Akten. Die Musik stammt von Engelbert Humperdinck, das Libretto (den Text der Oper) schrieb seine Schwester Adelheid Wette inspiriert durch das gleichnamige Märchen der Brüder Grimm. Für Einzelheiten zur Oper verweise ich auf Kloiber/ Konold/ Maschka (2016): *Das Handbuch der Oper*. S. 328–331. Das Werk stand auch den Versuchspersonen während der Aufnahme zur Verfügung.

⁸ Tobii Pro Glasses 2: <https://www.tobii.com/de/produkte/tobii-pro-glasses-2/>



Abbildung 02: Standbild der synchronisierten Eyetracking- und Frontkamera-Aufnahme im Splitscreen. Links oben im Bild die Eyetracking-Aufnahme von Tim mit Gaze-Punkt im Gesicht von Ana (grüner Cursor). Rechts oben im Bild die Eyetracking-Aufnahme von Ana mit Gaze-Punkt im Gesicht von Tim (roter Cursor). Mittig unten die Aufnahme der Frontkamera im Weitwinkel.

Anschließend wurden die Aufnahmen im Videoannotationstool ELAN (Wittenburg et al. 2006) nach den GAT2-Transkriptionskonventionen (Selting et al. 2009) annotiert.

3.1 Transkription und Standbildanalyse

Die Analyse der gewählten Sequenzen erfolgt auf Grundlage der Verbaltranskripte. Zur Erfassung der Gestik werde ich außerdem genaue Beschreibungen der durch die Videoaufnahme festgehaltenen Gesten vornehmen und meine Beobachtungen punktuell mit Standbildern belegen. Wenn nicht explizit anders begründet, legitimieren sich die einzelnen Standbilder aus den jeweiligen Gestenhöhepunkten der Gestenphrasen (dazu auch Stukenbrock 2009: 156). Auf genaue, durchgehende Gestentranskription, wie sie bspw. Lorenza Mondada (2014) vorschlägt, werde ich im Hinblick auf den Verlust an Übersichtlichkeit der Transkripte verzichten. Wo ich aber einen Mehrwert in der genaueren und verbaltranskriptalignierten symbolhaften Darstellung der Gestik sehe, werde ich auch vereinzelt multimodale Transkriptionen vornehmen. So kann der sequenzielle Ablauf von Zeigegesten mithilfe von Symboltranskription funktional veranschaulicht werden. Für ikonische Gesten oder Handlungsdemonstrationen ist eine symbolische Transkription dagegen mitunter weniger aufschlussreich als das Einfügen von Standbildern (dazu auch Stukenbrock 2009: 167). Das unten aufgeführte Gestentranskriptionssystem ist eine verkürzte Kombination der Gestentranskriptionssysteme nach Mondada (2014: 122) und Stukenbrock (2009: 163).

Die Analyse stellt keinen Anspruch auf eine erschöpfende Beschreibung aller (redebegleitenden) Gesten, sondern konzentriert sich auf den Einsatz körperlicher Ausdrucksressourcen, die als relevant für die interaktionale Herstellung geteilter Vorstellung auszumachen sind. Mir ist bewusst, dass ich dadurch eine interpretative Vorauswahl treffe, abhängig von Gegenstand und Forschungsinteresse der vorliegenden Arbeit.

Abbildung 03: Zusammenstellung Gestentranskriptionssystem

- Halten/Statik/Einfrieren einer Aktivität
- Bewegungsphase einer Geste/Aktivität
- ▲ Apex/Gipfelpunkt einer Geste/Aktivität
- ▼ Umkehrpunkt einer Geste/Aktivität
- >> Geste/Aktion beginnt vor Ausschnittbeginn
- >> Geste/Aktion verläuft über Ausschnitt hinaus
- # zeigt die Position des Gestenhöhepunkts im Verbaltranskript an

3.2 Aufgabe

Die Aufgabe der Teilnehmenden bestand darin, gemeinsam ein Bühnenbild für die drei Akte der ihnen gut bekannten Oper *Hänsel und Gretel* zu entwerfen. Als Hilfsmittel zur Bearbeitung der Aufgabe stand ihnen *Das Handbuch der Oper* (2016), ein Opernführer mit Details zur Oper sowie eine Zusammenfassung zu den zentralen Aspekten der jeweiligen Akte zur Verfügung. DIN A4 und DIN A2 formatiges Papier sowie Filzstifte durften für die Aufzeichnung von Bühnenbildelementen verwendet werden. Durch ihre Anordnung und Verteilung auf der Bühne, sollen die 2D Zeichnungen den visuellen Abgleich des ausgedachten Bühnenbilds im gegebenen Raum abbilden.



Abbildung 04: Abbildung des fertigen Bühnenbilds. Die skizzierten Bühnenbildelemente sind intentional im Raum angebracht worden.

Die Untersuchungsleiterinnen waren während der Aufnahme nicht im Raum und haben, abgesehen von der Aufforderung ein Bühnenbild für die Oper *Hänsel und Gretel* zu erstellen und ggf. auf die Hilfsmittel zurückzugreifen, keinen Einfluss auf die Art oder Interpretation des Bühnenbilds, den Ablauf oder die Länge der Interaktion oder die Gesprächsthemen genommen. Die Interaktionsteilnehmenden sollten nach eigenem Ermessen der Fertigstellung des Bühnenbilds die Interaktions- bzw. Aufnahmesituation beenden.

3.2 Relevanz und Analyse der Räumlichkeit

Das Bühnenbild sollte für die gegebene Räumlichkeit entworfen werden, die mit einer leicht vom restlichen Zuschauerraum erhöhte, ca. fünf Meter breite und drei Meter lange Bühne ausgestattet ist und nach hinten von einem Vorhang begrenzt wird. Die Bühne kann außerdem auf der rechten Seite durch einen separaten Bühnenaufgang in einen Garderobenraum verlassen werden. Auf der Bühne befanden sich ein Klavier sowie eine Bank und zwei Stehlampen, die von den Teilnehmenden als Teil des Bühnenbildes eingebunden werden konnten. Weiterhin standen ihnen Stühle, Notenständer und Bekleidungsstücke aus der Garderobe als Requisiten zur Verfügung.

Die aufgabenbezogene Aufnahmesituation erfordert also die Vorstellung eines Bühnenbilds, das in der Interaktion ausgehandelt bzw. kommuniziert und abgestimmt werden muss. Das impliziert, dass die Versuchspersonen ihre Imaginationen einander zugänglich machen mussten. Die gewählten Beispielsequenzen werden zeigen, dass die Teilnehmenden ihre Vorstellungen nicht nur verbal beschreiben, sondern einzelne Handlungen und Vorgänge physisch verkörpern und sich durch das Ausführen tatsächlicher, physischer Handlungen im mentalen Raum der Oper positionieren. Dabei müssen sie ständig die Koexistenz verschiedener lokaler Vorstellungsräume sowie des ‚realen‘ Wahrnehmungsraums kognitiv präsent halten und aufeinander abstimmen, damit sie für Aushandlungssituationen, Ein- und Rückbezüge zur Verfügung stehen und damit ‚sinnvolle‘ Verflechtungen der Räume möglich sind. Unter ‚lokale Vorstellungsräume‘ fasse ich zunächst alle gegenständlichen oder szenischen Elemente, die in der Interaktion aufgebaut werden. Wie in Kapitel 2.2 erwähnt, ist der gesamtsituative Basisraum eine 1:1 Anpassung an die physisch-räumliche Dimension des realen Bühnenraums und wird diesem im Aufbau einer gemeinsamen Imagination eines Bühnenbilds ‚übergestülpt‘. Konzeptionell mental, erhält dieser Basisraum durch die gegebenen Umstände der Aufnahmesituation einen modalen Sonderstatus: Den Teilnehmenden*innen ist bewusst, dass ihr Entwurf des Bühnenbilds nie realisiert werden wird, trotzdem ist die Interaktion zu (fast) jedem Zeitpunkt der Aufnahme so angelegt, als würden sie tatsächlich davon ausgehen, dass ihr Entwurf einer realen Aufführung dienen muss. Der Basisraum bekommt also eine grundlegend hypothetische Dimension, die im Gespräch aber frequent als zukünftig evoziert wird. Für die Beispielsequenzen bedeutet das vor allem, dass lokale Vorstellungsräume in der Regel als *future event spaces* bzw. *hypothetical spaces* aufgebaut werden.

Im Verlauf der Aufnahme erhält der rein in der Vorstellung der Interaktionsteilnehmer*innen existierende Basisraum eine konstante, physische Komponente. Elemente aus dem Basisraum werden schematisch auf Papier aufgezeichnet und mithilfe von

Klebeband visuell und als physischer Abgleich der mentalen Räume im *reality space* der Bühne verortet. Dabei werden die Requisiten oder räumlichen Umstände in die ausgehandelten Bühnenbildrequisiten übersetzt, sodass bspw. ein Stuhl *aus* und *im reality space* zu einem dreibeinigen Hocker im Basisraum wird (siehe Abb. 04). Der mentale Basisraum erhält dadurch eine – zwar abstrakte und auf mentale Kontextualisierung angewiesene – ‚reale‘ Verankerung. Die Wahrnehmung dieses *Skizzenraums*⁹ tritt an die Stelle der Wahrnehmung des *reality space* und wird so zum *grounded blend* der Interaktionssituation. Zeitlich überdauernd *überblendet* der mentale Input aus dem Basisraum die räumlich-physische Konstitutionen des *reality space* und verleiht ihm ein wirklichkeitsähnliches Format (vgl. Liddell 1998: 310). Auch lässt sich vermuten, dass die im Raum verteilten Skizzen als kognitive Stützen, für die im Interaktionsverlauf stetig weiterentwickelnde Komplexität des Basisraums fungieren. In der Aufnahme lässt sich beobachten, dass die Teilnehmenden auf bereits ‚besetzte‘ Orte des *reality space* Bezug nehmen und mithilfe der Skizzen auf mentale Räume verweisen (dazu mehr in Kap. 4.3)

Nach diesen theoretischen Vorannahmen zur gegebenen Kommunikationssituation, werde ich in der folgenden empirischen Analyse darauf eingehen, wie die räumlichen Verflechtungen in der Interaktion angezeigt, ausgestaltet und verdichtet werden. Die Gesprächssequenzen stammen aus unterschiedlichen Zeitpunkten der Aufnahme, sodass verschiedene Stadien der Planung und Ausarbeitung betrachtet werden können. Im Analysekapitel sind Ausschnitte nicht nach sequenziellem Ablauf der Aufnahme, sondern nach Gegenstandsschwerpunkten und zunehmender Komplexität der Vorstellungsräume sortiert.

4. Empirische Analyse

4.1 Die mentale Überschreibung des *reality space* mithilfe der Gestik

Die erste Sequenz stammt aus dem letzten Drittel der Aufnahme, in der die Teilnehmenden das Bühnenbild für den finalen Akt der Oper entwerfen. Ich stelle sie an den Anfang des folgenden Analyseteils, da sich anhand der Interaktion anschaulich und exemplarisch nachvollziehen lässt, wie sich der grundsätzliche Charakter der Aufnahmesituation konstituiert: Es werden lokale mentale Räume aufgebaut, die sich im Wahrnehmungsraum gründen, mit diesem sogar räumlich identisch sind, aber imaginierte Entitäten aufweisen, die die Objekte des *reality space* überschreiben. Zusätzlich werden in dieser Sequenz *reality space*, Basisvorstellungsraum und

⁹ Gemeint ist hier der Bühnenraum, in dem die einzelnen Skizzen intentional angeordnet sind.

der *grounded blend* der Skizzen mehr oder weniger explizit thematisiert. Dadurch lässt sich nachvollziehen, wie mühelos diese drei stets latent vorhandenen Räume miteinander verflochten werden und wie wörtlich die Überschreibung des *reality space* in der Aufnahme evoziert wird.

Fokusobjekt der Interaktion ist der Vorhang, der als Gegenstand des *reality spaces* und somit in der Aufnahme konstant zu sehen ist. Im Laufe der Aufnahme haben sich die Interaktionspartner*innen mehrfach darauf geeinigt, dass dieser Teil des Wahrnehmungsraums im Bühnenbildraum den Wald abbilden soll. Zu diesem Zeitpunkt existiert aber noch keine Skizze, die als physischer Abgleich der geteilten Imagination im Raum bzw. am Vorhang angebracht worden wäre. Das Transkript setzt ein, als Ana diesen Umstand, als Frage formuliert, aufbringt.

(1) Vorhang ((00:46:25 - 00:46:42 min))

→01 Ana: sollen wir da überhaupt BÄUme hInschreiben?
 02 (0.8)
 03 Auf den,
 04 Tim: (0.6) °h ich GLAUbe dass das
 05 hh°
 06 Ana: weil (.) [so ein GRO]ßer ofen den sieht man auf jEden fall,
 07 Tim: [<<p>ja;>]
 08 zumindest so baumUMrisse;=
 →09 =ich glaub hIEr vorne sieht man dann nich so viel aber so im
 HINter=
 →10 =das man so_n [bIsschen(--)] SIEHT das] es:-
 →11 Ana: [ja auf den auf den VORhang mein ich-]
 12 Tim: [das is gro]ßer wALd da;
 13 Ana: [vor BAUM;]
 14 ja oder WALD einfach drauf schreiben [ja;]
 15 Tim: [ich]

Die Interaktion beginnt mit der Frage von Ana, ob es überhaupt notwendig sei, diese mentale Besetzung des Wahrnehmungsobjekts *Vorhang* auch visuell zu verankern (Z. 01). Redebegleitend hebt sie den Arm und deutet zunächst (bis inklusive da überhaupt (Z. 01)) *ad oculus* mit der ganzen Hand auf den Vorhang im Hintergrund der Bühne, um mit BÄUme hInschreiben? (Z. 01) in den Modus des verweisenden Zeigens zu wechseln und die gesamte Breite des Vorhangs als Zeigeziel zu bestimmen.

Verbal nicht explizit eingeführt, wird der Vorhang erst durch die deiktische Zeigehandlung Teil des Kontextes¹⁰ und als Zeigeziel zum Fokus im Wahrnehmungsraum (vgl. Stukenbrock 2015: 39). Zeile 01 ist als Frage formuliert, sodass auch auf Turn-Taking-Ebene eine

¹⁰Auch in den Kommunikationssequenzen, die diesem Ausschnitt unmittelbar vorgelagert sind, war der Vorhang nicht Thema. Die deiktische (Zeige-)Handlung ist also nicht anaphorisch interpretierbar, sondern kontextualisiert sich allein durch die Zeigegeste.

Validierung durch den Interaktionspartner konditionell relevant wird (vgl. Auer 2013: 231). Eine entsprechende Reaktionshandlung bleibt aber zunächst aus. Mehr noch versäumt es Tim, durch seinen nach unten gerichteten Blick, sich perzeptorisch auf den eröffneten Verweisraum von Ana auszurichten. Nach einer mittellang gemessenen Pause setzt Ana daher erneut an, um verbal explizit zu machen, was in ihrer Zeigegeste nicht wahrgenommen oder zumindest nicht als Wahrnehmung bzw. Verstehen dokumentiert wurde. Ohne Anzeichen, dass Tim als Adressat, den Referent identifiziert hat, muss Ana davon ausgehen, dass ihre Zeigehandlung unerfolgreich war (Stukenbrock 2015: 92). Während sie auf verbaler Ebene nach dem passenden Referenten sucht (Z. 03), wiederholt sie ihre Gestenphrase¹¹ des am Vorhang hin und her fahrenden Zeigefingers. Erst mit Übernahme des Rederechts durch Tim in Zeile 04 beendet Ana die Zeigegeste und orientiert sich körperlich-perzeptorisch in Richtung des Sprechers. Syntaktisch wirkt Anas Äußerung in Zeile 03 wie ein Angebot zur Vervollständigung durch ihr Gegenüber (steigende Intonation, verzögerte Produktion, vgl. Brenning 2015: 213f.). In der mittellangen Pause lässt Tim diese sequenzielle Projektion aber verstreichen und setzt auf verbaler Ebene zur Eröffnung eines hypothetischen (Glaubens-)Raums an ((0.6) °h ich GLAUbe dass das (Z. 04)). Diesen beginnt er zeitgleich mit vor dem Körper ausgestreckten Unterarmen im Wahrnehmungsraum zu verankern (*grounding*) und im *reality space* durch seine Körper- und Blickausrichtung auf der rechten Bühnenseite zu verorten. Er bricht seine glaubensraumevozierende Äußerung aber ab und leicht versetzt, mit einer hörbaren Ausatmung in Zeile 05, stellt er auch die *grounding*konstitutive Geste ein. Stattdessen richtet er seinen Blick, noch im Gestenabbruch, auf den Vorhang aus und dokumentiert damit erstens, die Wahrnehmung des *reality space*-Referenten und bestätigt zweitens – schon in Überlappung mit Anas, zu ihrer ursprünglichen Frage, kohäsiv eingeleiteten neuen Intonationsphrase – durch seine leise, in mittel fallender Intonation realisierten Zustimmungspartikel <<p>ja;> (Z. 07), dass er nicht beabsichtigt die *belief space* evozierende Sequenz zu vervollständigen. Stattdessen bedient Tim sich mit seiner nächsten Äußerung der gleichen Form des ko-expressiven, aber (mental) räumlich abweichenden Verweises wie er bereits bei Ana sequenzeinleitend zu beobachten war. Er referiert mit einer Zeigegeste (sein Zeigefinger fährt in wellenförmigen Bewegungen den Vorhang entlang und zeichnet dabei dessen Umriss nach) auf den Vorhang im *reality space*. Auf verbaler Ebene aktualisiert und spezifiziert er aber dessen *mental space*-Entsprechung als BÄUme (Z. 01) bzw.

¹¹ Mit *Gestenphrase* beziehe ich mich auf ein Konzept von Einbettungsbeziehungen deiktischer Zeigehandlungen, das Anja Stukenbrock in ihrem Werk zur Deixis in der Face-to-Face-Interaktion näher ausgearbeitet hat. Die *Gestenphrase* bezeichnet die Gesamtheit der einzelnen *Phasen* (Bewegungsabläufe) einer Zeigegeste mit einem Zeigziel (Stukenbrock 2015: 37f.).

als baumUMrisse (Z. 08). Die deiktische Geste verweist dabei auf einen Ort des *reality space*, macht aber gleichzeitig Aspekte des Phantasmas sinnlich zugänglich. In schnellem Anschluss gibt Tim zu bedenken, dass Teile des *reality space* bereits durch mentale Entitäten besetzt sind und mitbedacht werden müssen (=ich glaub hIER vorne sieht man dann nich so viel aber so im HINter= (Z. 09)). Hier zeigt sich auch die Komplexität der Anforderung, die räumliche Beschaffenheit dieser lokalen mentalen Räume in der mentalen Repräsentation aufrecht zu erhalten. Eingebettet in einen *belief space*, bezieht sich das dann auf den hypothetischen *future event space* des fertigen Bühnenbilds, der bereits in Kapitel 3.3 erläutert wurde und in dem der lebensgroße Ofen (siehe Z. 06) einen Großteil des Vordergrundes einnehmen soll. Im Wahrnehmungsraum ist dieser Ofen aber nur als kleine Skizze vertreten.

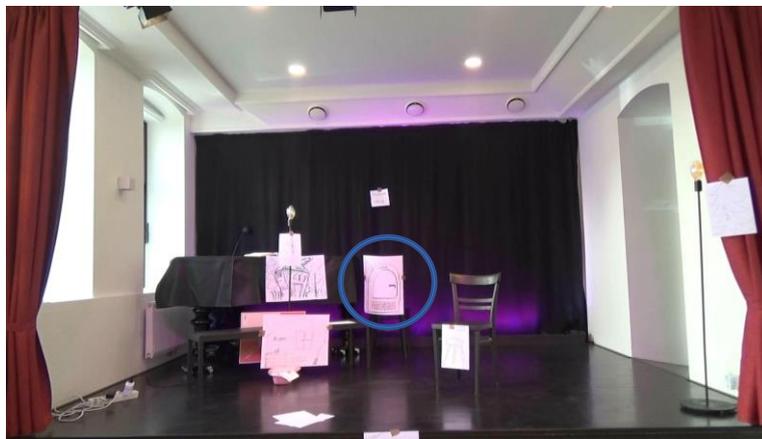
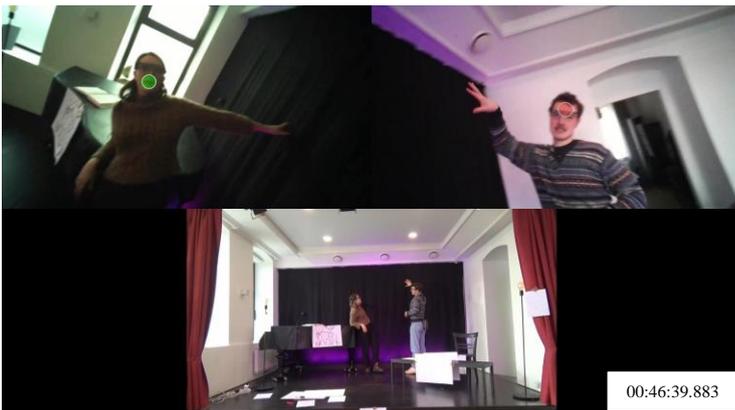


Abbildung 05: Skizze des Ofens als Teil des Bühnenbilds. Im mentalen Basisraum besitzt der Ofen lebensgroßes Format.

Mit =das man so_n [bIsschen (--)] SIEHT das] es:- (Z. 10) retrahiert Tim zu seiner Äußerung in Zeile 08 und nimmt gleichzeitig seine Geste der wellenförmigen, Umriss nachzeichnenden Hand am Vorhang wieder auf. In vollständiger Überlappung mit dieser Äußerung macht Ana an dieser Stelle (Z. 11) verbal explizit, was sie bereits in Zeile 01 unter Zuhilfenahme ihrer Zeigegeste zum deiktischen Fokusobjekt des Vorstellungsrums gemacht hat: den Vorhang. Aligniert mit der wiederholten Präposition zeigt sie dabei erneut in einer bogenförmigen Verweigeste über die Länge des Vorhangs. In der Überlappung der Äußerungen überschneiden sich auch die Zeigegesten, die aus der jeweiligen Origo der Zeigenden *ad oculus* auf den Vorhang im Wahrnehmungsraum deuten. In beiden Zeigehandlungen ist der Vorhang Zeigeziel und Referenzobjekt. Tims Geste verweist aber gleichzeitig am Phantasma und macht durch symbolische Handbewegungen die Baumumrisse im Vorstellungsrums visuell zugänglich. Weitere Verflechtungen der Räume entstehen durch

die unterschiedlichen verbalen Bezugsebenen. Tim fasst sein Zeigeziel ko-expressiv¹² als *baumUMrisse* (Z. 08) oder *wAld* (Z. 12). Ana kontextualisiert dasselbe Zeigeziel in seiner *reality space*-Größe als *Vorhang* (Z. 01, 03, 11). Hier zeigt sich, wie durch interpersonell multireferenzielle Verweise ein mentaler Raum im Wahrnehmungsraum verankert sein kann. Außerdem lässt sich nachvollziehen, wie temporär und medial verschieden mentale Räume in der Interaktion eröffnet werden können.

Standbild 01



10 Tim: =das man so_n [bIsschen (--) SIEH#T das] es:-
11 Ana: [ja auf den auf den VOR#hang mein ich-]
12 Tim: [das is gro]ßer wAld da;

Exemplarisch lässt sich anhand dieser Interaktion zeigen, inwiefern durch die im Wahrnehmungsraum verteilten Skizzen ein *grounded blend* entsteht. Die *input spaces* ergeben sich aus der *räumlichen Beschaffenheit der Bühne* und den in der Interaktion ausgehandelten *Vorstellungsräumen*. Die Vorstellungsräume werden in Form von Skizzen im Wahrnehmungsraum verteilt und dadurch dauerhaft *gegrounded*. Somit können die einzelnen Skizzen der mentalen Räume auch als eine Form der *grounded mental spaces* verstanden werden, in denen die temporären mentalen Räume kodifiziert sind. Unter Beachtung unterschiedlicher Dimensionen der Flüchtigkeit lässt sich das Konzept, das ursprünglich die physische Verkörperung mentaler Abläufe oder Objekte fasst (Liddell 1995: 23), ebenso auf andere Arten sinnlich wahrnehmbarer Repräsentation mentaler Räume übertragen. Der daraus resultierende beispielbare Bühnenraum ließe sich demnach als *grounded blend* fassen. Gerahmt werden diese Räume durch den Frame der Märchenoper *Hänsel und Gretel*, der elementare Abläufe und Rollen der Oper als Hintergrundwissen strukturiert. Abbildung 06 zeigt, wie die

¹² Mit *Ko-Expressivität* beziehe ich mich auf ein Gestenkonzept nach Kendon (1980), das Ehmer (2011: 85–88) Bezug nimmt und spezifiziert. Für dieses Konzept zentral ist, dass Gesten Bedeutung tragen und in Verbindung mit der verbalsprachlichen Ebene semantisch-pragmatisch und zeitlich abgestimmt auftreten. Dabei „drücken Gesten und verbale Äußerung meist unterschiedliche Aspekte einer zugrundeliegenden ‚Idee‘ aus.“ (ebd.: 85).

Modelldarstellung der Beziehung von Frames und mentalen Räumen nach Fauconnier ([1997] 1999: 12) auf die Aufnahmesituation übertragen werden kann. Die Grafik veranschaulicht die Konzeptidee des Blends als räumliche Übereinstimmung von Wahrnehmungsraum und Vorstellungsraum.

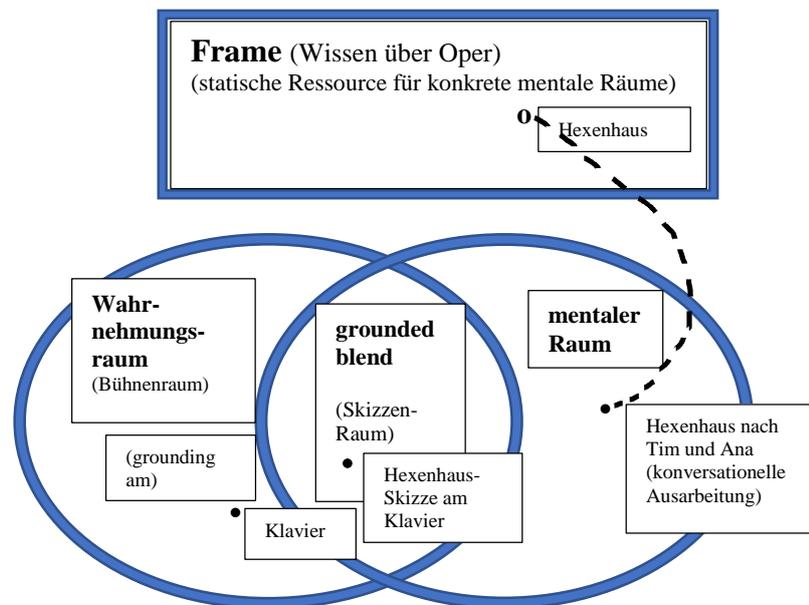


Abbildung 06: Modell zur Veranschaulichung des *grounded blends* durch Überschreibung von Wahrnehmungsraum mit mentalen Räumen (anhand des elementaren Bühnenbildelements Hexenhaus).

Im folgenden Ausschnitt geht es wie zuvor um die Anpassung eines lokalen mentalen Raums an den Wahrnehmungsraum mithilfe deiktischer Gesten. Im Unterschied zur vorherigen Sequenz überschreibt der hier eröffnete mentale Raum aber kein konkretes Objekt des *reality space* mit *blend*ähnlichem Ergebnis. Stattdessen wird deutlich, wie der stets latent aktive Frame der Oper die Interaktion rahmt und wie durch sprunghaftes zurückgreifen auf diesen, eröffnete mentale Räume beschränkt werden (dazu Fauconnier [1997] 1999: 39). In dem abgebildeten Gesprächsausschnitt geht es um die Frage, was dem bisher entworfenen Bühnenbild ggf. noch fehlt. Ihre Annahmen dazu können die Interaktionsteilnehmenden aus ihrem Hintergrundwissen über die Oper und die klassische Bühnenbildinterpretation abrufen. Gleichzeitig müssen neue, als relevant erachtete Elemente an das bisher rein mental projizierte Bühnenbild – die Skizzierung der mentalen Räume hat in der folgenden Sequenz noch nicht begonnen – angepasst werden.

(2) Zaun ((00:19:42 – 00:19:59 min))

01 Tim: was gä=
02 =was is was is NOCH so:-

→03 Ana: ein ZAUN sollte es geben;
 04 aber [ich MEIne den]
 05 Tim [ZAUN,]
 06 aber das HAT man ja
 07 das das kAnn [man ja auch durchaus:] ähm in DIEsem bereich
 [haben;]=

→08 Ana: [das geht bis da HIN ja;]
 09 [ja;]

→10 Tim: =weil dann hast du hier hälfte [ZAUN hÄlfte]
 11 Ana: [STÖRT auch nicht;]
 12 Tim: (0.7)
 13 <<gedehnt>HÄlfte:->
 14 (2.3)
 15 <<stöhnend>ah> (.) was wollt ich sAgen,
 16 Ana: nur wa Aber WARte mal ganz kurz;

Die Sequenz beginnt –im Konjunktiv angesetzt – mit der Eröffnung eines neuen hypothetischen mentalen Raums (Z. 01). Tim bricht seine Äußerung jedoch ab, retrahiert zur Fragepartikel und verwirft in seiner Folgeäußerung den explizit hypothetischen Status des Vorstellungsraums (Z. 02). Die latente Verfügbarkeit¹³ des einleitenden *was gä* (Z. 01) – und obwohl abgebrochen – wird durch Ana mit ihrem Einstieg in den neuen mentalen Raum auf der Konjunktiv-Ebene *ein ZAUN sollte es geben;* (Z. 03) betont. Diese syntaktische Latenz des Konjunktivs, die die Hypothesizität des Raums sichtbar macht, scheint so präsent, dass Tim sie, nach seiner Bestätigung des ‚Zaun-Raums‘ in Zeile 05 (*ZAUN,*) syntaktisch wiederaufnimmt. In einer selbsteingeleiteten Selbstreparatur (zu syntaktischen Reparaturen siehe Auer 2013: 241–246) repariert er die Indikativform *HAT* (Z. 06) zu der fakultativeren Formulierung *kAnn* (Z. 07) und verfestigt dadurch den Eindruck eines hypothetischen Raums. Unterstrichen wird dieser Eindruck noch durch die Verwendung des generalisierenden Indefinitpronomens *man* (Z. 06 & 07). Trotzdem ist der Zaun offenbar als obligatorisches Element für das Bühnenbild aus dem Hintergrundwissen beider abrufbar und benötigt keine weitere Relevanzdiskussion, sodass sich die folgende Interaktion unmittelbar auf dessen Verortung im Wahrnehmungsraum im Abgleich mit den bisher evozierten lokalen mentalen Räumen ausrichtet. Parallel zu seiner abgebrochenen Äußerung in Zeile 06 eröffnet Tim mit seinem zur Seite/hinten öffnenden Körper einen Verweisraum, unmittelbar gefolgt von einer bereichsverweisenden *demonstratio ad oculus* mit geöffneter Hand entlang des unteren Teils des Vorhangs im Hintergrund der Bühne. Diese Geste erhält er, trotz ko-verbaler Einschübe von Ana, bis Zeile 10 aufrecht. Zunächst aber scheint Tims Phrasenabbruch in Zeile 06, seine mit Verzögerungssignal eingeleitete Selbstreparatur sowie seine exponierte Gestik Ana zur syntaktischen Hilfestellung

¹³ Einmal im Gespräch verwendete syntaktische Strukturen können über mehrere Äußerungen hinweg kognitiv präsent bleiben und von den Gesprächsteilnehmenden über einen gewissen Zeitraum hinweg durch Wiederaufnahme sprachlich aktualisiert werden (Auer 2007).

zu veranlassen. In Überlappung mit Tims (...) man ja auch durchaus: (Z. 07) retrahiert Ana zu seinem phraseneinleitenden *das* und bietet in ihrer Vervollständigung *das geht bis da HIN ja* (Z. 08) verbal ihre Wahrnehmung als Bestätigung an. Verstärkend steigt sie in die bereichsabsteckende Geste von Tim ein und verweist von links nach rechts einmal auf den Bereich unmittelbar vor dem rückseitigen Vorhang. Parallel zu Anas phrasenbeendenden Zustimmungspartikel (*ja*; (Z. 08)), verweisen beide Interaktionsteilnehmenden mit dem ausgestreckten Arm auf den Bereich des Vorhangs, während der Blick jeweils auf das Gesicht des Gegenübers gerichtet ist. Die Versicherung der Wahrnehmungswahrnehmung – der Kontrollblick, der sicherstellen soll, dass die eigene Handlung wahrgenommen wird (Stukenbrock 2015: 87) – wird hier von beiden gleichzeitig ausgeführt. In der Kombination mit Anas körperlich-gestischen Ausrichtung wird die Wahrnehmungsbekundung so zur visuellen Spiegelung des Gegenübers (Standbild 02).

Standbild 02



Anas Einstieg, in den *ad oculus* verorteten mentalen Raum, bildet also auf verbaler Ebene die ko-konstruierte Aushilfe bei einer Wortsuche, die ein gestisches Deiktikum betrifft, dessen körperlich-visueller Part bereits vom Sprecher eingelöst wurde. Wie das Standbild 02 und die Häitationen in seiner Äußerung während der Überlappung mit Anas verbalem Vervollständigungsveruch zeigt, nimmt Tim ihren Einwurf wahr. Statt explizit auf Anas Äußerung Bezug zu nehmen, verbleibt er aber in seiner eigenen syntaktischen Struktur, die er so weit verzögert, dass er erst nach Anas Angebot einen Abschlusspunkt erreicht. Parallel zur un abgeschlossenen Intonationsphrase hält er seine Geste, deren kommunikative Relevanz bis zu seinem Phrasenabschluss noch unerfüllt ist, in exponierter Position.¹⁴ Mit der lexiko-semanticen Einlösung des gestischen Deiktikons (in DIEsem bereich (Z. 07)) und Anas

¹⁴ Jürgen Streeck hält in seinem Aufsatz zu Prozessualität und zeitlichen Aspekten von Gesten fest, dass „das Halten einer Geste unerfüllte konditionale Relevanz bedeuten“ kann (Streeck 2007: 168).

erneuter Zustimmungsausprägung (Z. 08), führt Tim in schnellem Anschluss weiter aus, warum er den soeben abgesteckten Bereich des Wahrnehmungsraums für den mentalen Zaun-Raum passend hält (=weil dann hast du hier hälft e [ZAUN hälft e](Z.10). Visuell schafft Tim an dieser Stelle einen beinahe nahtlosen Übergang seiner Hin-und-Her-Geste zu einer taktstockartigeren Hier-und-Da-Geste (die zwei Betonungsmarkierungen der *beats* (McNeill 1992) jeweils auf der ersten Silbe von hälft e), die die Konkretisierung des Zaun-Bereichs aus Zeile 07 und 08 kontextualisiert. Verbal befindet er sich nach der deixiskonstitutiven Absteckung des Raums wieder in einem Wortfindungsproblem, das er – mit weiterhin exponierter Gestik – nach einer mittleren Pause und der gedehnten Wiederholung seines zuvor auslautenden hälft e (Z. 10), nach einer weiteren längeren Pause explizit verbalisiert (<<stöhnend>ah> (.) was wollt ich sAgen, (Z.15)). Die Sequenz – und damit auch die Evozierung des Zaun-Raums – findet ihr Ende in Anas Einwurf in Zeile 16, der weder explizit auf die Wortsuche von Tim eingeht noch den Zaun thematisiert, sondern einen neuen mentalen Raum eröffnet.

Zusammenfassung

Die obigen Beispiele konnten exemplarisch zeigen, wie in der Aufnahme lokale mentale Räume hergestellt und gemeinsam ausgearbeitet werden. Dabei tritt der gestische Aspekt der Raumevozierung insofern hervor, als dass deiktische Gestik der Kontextualisierung einzelner Vorstellungsräume dient und ihnen eine visuelle Repräsentation im Wahrnehmungsraum (*grounding*) ermöglicht. Temporäre mentale Räume können so mit der*dem Gegenüber geteilt und in der Übernahme, Wiederholung und Weiterentwicklung einer Gestenphrase prozesshaft und gemeinsam elaboriert werden. Es hat sich gezeigt, dass Zeigegesten *ad oculus* verweisen und gleichzeitig auf mentale Entitäten referieren können. Deiktische Gestik muss weder mit ihrer körperlich-visuellen Entsprechung zeitlich aligniert noch intrapersonell gelöst werden. Vielmehr liegt in der Verzögerung ko-verbaler Bedeutungsgenerierung ein ko-konstruktives Potenzial, das von den Teilnehmenden genutzt wird, um potenzielle Vervollständigungen anzubieten und Verstehen anzuzeigen. Es entsteht der Eindruck eines kollaborativ entwickelten mentalen Raums, der auf den gleichen *backstage* ablaufenden Raumaktualisierungen beruht.

Im folgenden Kapitel soll anhand einer einzelnen längeren Sequenz der Aspekt interpersoneller gestischer Raumevozierung mit unterschiedlichen gestisch-deiktischen Zentren, aber gleichem Bezugssystem genauer untersucht werden. Mit Bezugssystem meine ich im Folgenden nicht nur den situativen Kontext der Interaktion (Bühler [1934]1999: 113),

sondern auch die lokale Bedeutungsstruktur als deiktisches Ziel. Die Bedeutungsstruktur wird in der Interaktion teilweise auf verbaler, vor allem aber auf gestischer Ebene ko-konstruiert. Die Teilnehmenden befinden sich währenddessen in der körperlichen Spiegelung des Gegenübers und damit in der perspektivischen Spiegelung des entstehenden *surrogate spaces*. Durch das *grounding* erschaffen die Teilnehmenden einen physisch wahrnehmbaren Ersatzraum im *reality space*, dessen Entitäten in der körperlich-gestischen Spiegelung des Gegenübers repräsentiert sind.

4.2 Perspektivübernahme als deiktische Spiegelung

Der folgenden Interaktion geht eine Einigungssequenz voraus, in der festgelegt wird, dass das Bühnenbild eine Art Drehbühne umfassen soll. Durch einfaches Verschieben und Drehen der einzelnen Wände sollen andere Elemente der Oper zum Vorschein kommen. So sollen dieselben Bühnenelemente, die im ersten Akt das Elternhaus von Hänsel und Gretel zeigen, auf ihrer Rückseite das Hexenhaus verbergen, das im dritten Akt benötigt und durch das Wenden sichtbar gemacht wird.

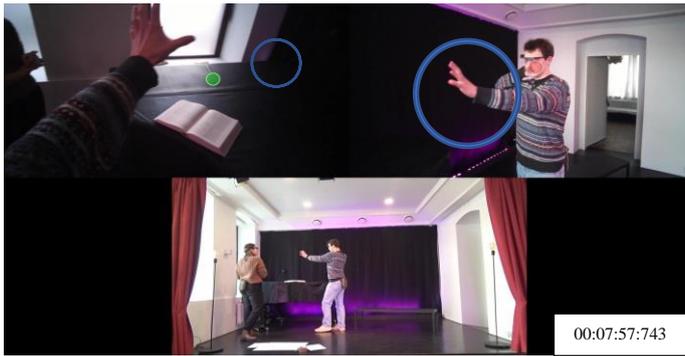
(3) Wände ((00:07:56 – 00:08:28 min))

→01 Tim: oder du hast HIER (-) auf
 02 <<pp>hm_hm->
 →03 du hast hier ZWEI wände quasi,
 04 Ana: ja,
 05 Tim: VORne is (-) is Inneneinrich
 06 also is an_ner wand geMALT mehr oder we[niger-]
 07 Ana: [<<p>o>]
 08 Tim: [INnen]einrichtung [(-) des hAUses,]
 09 Ana: [ja;]
 →10 [da könnte man auch eine (.) KÜCHen]insel
 11 äh_die küche könnte man auch DRAUFtu[n dann;]
 12 Tim: [zum BEIsp]iel,
 13 (0.7)
 14 und dann: kannst du_s KLAPpen?
 15 Ana: ja;
 →16 Tim: (0.8) dann is hier IMmer noch ne wAnd,=
 →17 Ana: =ja und [auf der ANdern seite,]
 18 Tim: [die das HEXenhau]s is
 oder=oder=[oder=oder]WALD,
 19 Ana: [oder vielleicht WALD?]
 20 weil wenn man [das so diagoNAL] [macht,]
 21 Tim: [und geNAU;]
 22 [und dann] is das: [(-) nämlich]
 äh auf der ANdern seite-
 23 Ana: [<<p>ha_JA;>]
 24 Tim: wä[re DAS:]
 25 Ana: [ja der WAL]D;
 26 das_is ja EIgentlich=
 27 =das is_ne GUTE idee;

Zu Beginn der Sequenz stehen sich Ana und Tim direkt gegenüber, ihre Körper sind einander zugewandt. Tim eröffnet einen zur vorangegangenen Sequenz alternativen (oder (Z. 01)) mentalen Raum und richtet die interpersonelle Koordinierung durch den lokaldeiktischen Ausdruck *HIER* (Z. 01) perzeptorisch auf einen Bereich des Wahrnehmungsraumes aus (vgl. Stukenbrock 2015: 65). Gleichzeitig zeigt sich spätestens in Zeile 03, dass schon die deiktische Gestik in Zeile 01 als *space builder* fungiert (Standbild 03) und die mentale Versetzung in den Vorstellungsraum markiert. Mit *HIER* (Z. 01) und analog zum Gestenhöhepunkt seiner verweisenden Hand *grounded* Tim den temporären mentalen Raum in einem physisch zugänglichen Ort des *reality space* und besetzt diesen damit lokal (Standbild 03). Nach seinem Abbruch der sequenzeinleitenden Phrase (Z. 01) und der verzögernd wirkenden Äußerung in Zeile 02 macht Tim in Zeile 03 verbal deutlich, welches imaginierte Objekt er im Sinne einer gemeinsamen Imagination zur Disposition stellt: die *wände* (Z. 03) des Bühnenbilds. Analog dazu verkörpern seine vor dem Körper ausgestreckten, parallel zueinander stehenden Arme als ikonische Geste die (Ausrichtung der) *Wände* (Standbild 04).¹⁵ Durch die körperliche Repräsentation und die explizite deiktische Verortung im Wahrnehmungsraum projiziert Tim einen *surrogate space*. Dieser ‚Stellvertreterraum‘, in den Ana ab Zeile 10 durch ihre körperlich-gestische Spiegelung von Tims Gestik miteinsteigt, bleibt bis Zeile 18 visuell bestehen. Der *surrogate space* wird für den folgenden Interaktionsausschnitt zum temporären Basisraum, auf den bedeutungskonstitutiv hierarchisch aufgebaut und zu dem immer wieder sprunghaft zurückgekehrt wird (dazu auch Zima 2021: 105). In der Aufnahme lässt sich dieser hierarchische Aufbau und die wiederkehrende Bezugnahme auf den aktuellen Basisraum nicht auf der verbalen, sondern vor allem auf der gestisch-körperlichen Ebene nachvollziehen. Die Teilnehmenden betonen Aspekte des *surrogate space*, indem sie den hinteren (von der Kamera abgewandten) Arm in exponierter – und hier raumkonstitutiver – Position belassen und vornehmlich mit dem vorderen Arm raummodifizierende Aspekte gestisch darstellen.

¹⁵ Wenngleich die vorbereitende Bewegungsphase dieser Gestenphrase schon im von Zeile 01 auslautenden *auf* und schon in Zeile 02 seinen quasistatischen Zustand erreicht, lässt sich ihr Gestenhöhepunkt in der taktstockartigen Betonung ausmachen, die parallel zum Nucleusakzent (*ZWEI*) von Zeile 03 stattfindet.

Standbild 03



01 Tim: oder du hast HI#ER (-) auf
 Tim-G: -----▲----->>

Standbild 04



03 Tim: du hast hier ZWEI# wände quasi,
 Tim-G: >>-----▲-----

Standbild 05



10 Ana: [da könnte man auch eine
 Ana-G: -----
 (.) KÜCHEn]insel#
 -----▲

Im Folgenden gehe ich genauer auf die Spiegelungshandlungen ein und diskutiere, inwiefern sich darin Perspektivübernahmen als Verstehensdokumentationen erkennen lassen oder ob es sich nicht vielmehr um turnkompetitive Aktionen¹⁶ handelt. Körperlich befindet Ana sich in der Spiegelung von Tims *grounding*-Geste und rezipiert spiegelnd seine Perspektive. Eine raummodifizierende Elaboration erhält der Vorstellungraum erst dadurch, dass Ana die körperlich-gestische Spiegelung von Tims physischer *mental space*-Repräsentation nutzt, um vom selben Fokusraum aus, eigene Ideen zur Disposition zu stellen.

¹⁶ Für einen Überblick über allgemeine Mechanismen des kompetitiven Turn-Taking auf syntaktischer Ebene, verweise ich auf Auer 2013: 238f. oder Brenning 2015: 229f.

Mit Zeile 10 befinden sich die Teilnehmenden also erstmals in der direkten körperlich-gestischen Spiegelung des Gegenübers und im Ausgangspunkt des *surrogate space* (Standbild 05). Ana nutzt die Pause in Tims Turn (nach der bereits geäußerten Nukleusakzentsilbe (Z. 08)), einen weiteren hypothetischen mentalen Raum zu eröffnen (Z. 10), der auf Tims Vorstellung der Inneneinrichtung aufbaut. An dieser Stelle wird die „Kohärenzrelation“ (Zima 2021: 105), in der hierarchisch aufgebaute mentale Räume in der Regel zueinander stehen, plastisch: Anas Ansatz der gestischen Spiegelung von Tims *grounded mental space* baut auf seiner ikonischen Geste auf, indem sie ebenfalls mit dem von der Kamera aus hinteren Arm die Wandkone nachahmt und mit dem anderen Arm zu einer gestischen Darstellung eines tischartigen Objekts ansetzt, das sie als KÜCHENinsel (Z. 10) bezeichnet (Standbild 05). Das anaphorisch gebrauchte Deiktikum *da* (Z. 10) sowie *DRAUFtun* (Z. 11) verweisen zusätzlich am Phantasma in den von Tim eröffneten Fokusraum der Inneneinrichtung. Tim verharrt währenddessen im *grounding* des temporären Basisraums. Seine ikonische Geste hält er in exponierter Position bei und somit die Relevanz ‚seines‘ Fokusraums aufrecht (Streeck 2007: 187). Nach Äußerung der Nukleusakzentsilbe in Zeile 11, kann Tim das Ende der Intonationsphrase antizipieren (Brenning 2015: 30) und sichert sich in der Ratifizierungsposition, noch in Überlappung, das Rederecht wieder. Nach einer Pause von 0,7 Sekunden – in der er weiterhin im „Hold“ (Streeck 2007: 168) seiner Geste verharrt – erläutert Tim dann, was es mit der zweiten, hinteren Wand auf sich hat: *und dann: kannst du_s KLAPpen,* (Z. 14). Mit dem Aufklappen der vorderen Wand, soll die hintere erscheinen, auf der das Bild des Hexenhauses als zentrales Bühnenelement für einen späteren Akt sichtbar wird (vgl. Z. 14, 16, 18). Zeitgleich mit *KLAPpen,* (Z. 14) demonstriert er mit dem Arm, der zuvor die vordere Wand im *reality space* verankert hat, das Auseinanderziehen der Wände. Während Tims Geste den *surrogate space* elaboriert, steigt Ana in die Geste des Aufziehens mit ein, deutet sie aber eher im symbolisierenden Modus eines *token space* an und bricht außerdem die gestische Spiegelung auf, indem sie mit der hinteren, statt wie Tim mit der vorderen Hand agiert. Die verzögernde Pause, phraseneinleitend zu Zeile 16, nutzt Ana dann, um sich körperlich wieder in die perspektivische Spiegelung des *surrogate spaces* der Wände zu begeben. Einleitend in Zeile 16 (*dann is hier IMmer noch ne wAnd, =*) setzt Tim an, seine Idee zur Rolle der hinteren Wand zu erläutern und betont mit kleinen, taktstockartigen Auf-und-Ab-Bewegungen, des weiterhin in der Repräsentation der hinteren Wand befindlichen Arms, von welcher Wand er gerade spricht. Die steigende Intonation am Ende der Intonationsphrase nutzt Ana als Angebot (Brenning 2015: 213), um im schnellen Anschluss auch auf verbaler Ebene anzuzeigen, dass sie verstanden hat, wo im mentalen Raum Tim sich

aktuell befindet. Dass [auf der Andern seite-] (Z. 17) als semantische Paraphrase zur Äußerung von Tim davor zu interpretieren ist (und nicht als Vervollständigung auf semantischer und syntaktischer Ebene), wird durch ihre äußerungsaligmierte Geste deutlich, die Tims taktstockartige Auf-und-Ab-Bewegung aus Zeile 16 exakt wiederholt.

Standbild 06¹⁷

		<p>16 Tim: (0.8) dann is hier I#Mmer noch ne wAnd,= Tim-G: >>-----▲---▲---▲---▲---▲---▲---▲---▲--- Ana-G: ---▲-----</p>

Standbild 07

		<p>17 Ana: =ja und [auf der AN#dern seite,] Ana-G: -----▲---▲---▲---▲---▲---▲----- -->> Tim-G: ----- -->></p>

In der Folge wird ihre Verstehensbekundung auch nicht von Tim behandelt, stattdessen nimmt er seine eigene syntaktische Projektion in Zeile 18 und noch in Überlappung mit Anas Einwurf mit [die das HEXenhau]s is oder=oder=[oder=oder]WALD, (Z. 18) wieder auf. Die in schnellem Anschluss wiederholte Konjunktion und die dadurch entstehende Verzögerung des Redeflusses, nutzt Ana, um erneut das Rederecht zu ergreifen und mit Wiederholung der Konjunktion und hoch steigendem Intonationsverlauf [oder vielleicht WALD?] (Z. 19) ein eigenes Angebot zur Vervollständigung der Phrase

¹⁷ Der leicht verschwommene Eindruck der Hände in den Standbildern (siehe Hervorhebung durch die blauen Kreise) ist auf die schnellen taktstockartigen Bewegungen zurückzuführen. Der unmittelbare, wiederholte Wechsel von Bewegungsphase und Gestenhöhepunkt in der Gestenzeile (Z. 16 Tim & Z. 17 Ana) symbolisiert die taktstockartige Gestik. Die Anzahl der einzelnen *beats* entsprechen der Anzahl der Gipfelpunkt-Symbole. Die Gipfelpunkte sind hier nicht durch #-Symbole 1:1 mit dem Verbaltranskript abgeglichen, weil das durch die gegebene Videoauflösung kaum angemessen einsehbar ist.

anzubieten (Brenning 2015: 213). Während Ana die mögliche gesuchte Konstituente anbietet, unterbricht Tim aber seinen Redefluss nicht, sondern ‚verharrt‘ in der Wiederholung der Konjunktion, bis Ana ihre Äußerung beendet hat. Nach ihrem Phrasenabschluss wiederholt er die hauptakzentuierte Silbe *WALD*, (Z. 18) und kontextualisiert – weiterhin im passenden Projekt seiner eigenen begonnen syntaktischen Gestalt – damit sein Einverständnis. (In Zeile 21 bestätigt er diesen Eindruck nochmals durch [und *geNAU*;].) Ab Zeile 20 lösen sich die beiden aus der direkten gestischen Spiegelung, bewegen sich im Wahrnehmungsraum und beginnen mit der Ausarbeitung der auf die andere Seite des Raums verschobenen Wand.

Zusammenfassung

Der mentale Raum, der hier ko-expressiv evoziert wird, ist sehr komplex, sodass eine visuelle Orientierung der Beteiligten aufeinander zur Bedingung des Verstehens wird (Stukenbrock 2015: 232). Gleichzeitig gilt dieser körperlich-visuelle Bezug vor allem dem gemeinsamen *grounding* des Fokusraums, der im Sinne eines „lokale[n] Aufbau[s] einer gemeinsamen semantischen Struktur“ (Ehmer 2011: 48) zum aktuellen Interaktionsraum wird. Die körperliche Spiegelung fungiert dabei nicht nur als eine Verstehen anzeigende Paraphrase der Gestik des Gegenübers. Vielmehr bietet die interpersonale Alignment der gestischen Ebene die Möglichkeit, fortwährend zu kontextualisieren, *wo* im mentalen Raum sich die Teilnehmenden gerade befinden bzw. *was* aktuell im Fokus der Aufmerksamkeit steht, ohne dies verbal explizit machen zu müssen. Das wiederum ermöglicht die Ausarbeitung einer gemeinsamen Vorstellung, die interaktiv und hierarchisch auf dem lokalen Basisraum aufbauend gestaltet wird, während dieser Basisraum auch auf visueller Ebene präsent bleibt. (Die Beteiligten belassen während der verbalen wie gestischen Ausarbeitung des mentalen Raums einen Arm in der Geste, die Ort und Ausrichtung der hinteren Wand markiert.) Die Übernahmen des Rederechts in Überlappung und mit expandierten Vervollständigungen (Z. 17, 19) wirken dabei zunächst kompetitiv, durch die zustimmende Behandlung des ersten Sprechers (Tim) (Z. 12, 18, 21), erscheinen sie aber nachträglich als kollaborativ ko-konstruierte Strukturen. Unterstrichen wird der Eindruck einer gemeinsam aufgebauten semantischen Struktur (*shared cognition*) durch die ‚dichte‘ (schnelle Anschlüsse, Überlappungen, wenig Rückversicherungspartikeln) beinahe reparaturfreie Verbalebene.

Wie schon im ersten Kapitel bleibt die Überschreibung des *reality space* mit mentalen Entitäten, um Vorstellungsräume mental zu verankern und interaktiv zu etablieren, d.h. visuell wahrnehmbar und teilbar zu machen, ein zentraler Aspekt des Gestikeinsatzes. Bereiche des

Wahrnehmungsraumes werden so zunehmend durch imaginierte Objekte *besetzt* und das so dass die gemeinsam lokal aufgebauten Räume im Wissensbestand der Interaktionsteilnehmenden aktiv gehalten werden müssen, um eine Doppelbesetzung des *reality space* zu vermeiden. Im folgenden Kapitel wird sich zeigen, wie die Teilnehmenden *grounding* einsetzen, um eine *Origoversetzung* zu kontextualisieren und imaginierte Figurenhandlungen auf körperlicher Ebene zu *animieren*.

4.3 Origoversetzung und Animation

Bisher lag der Fokus der Untersuchung auf deiktischen Gesten, deren Funktion darin bestand *ad oculus* und gleichzeitig – in Kombination mit verbaler Kontextualisierung – auf imaginierte Objekte am Phantasma zu verweisen und interaktiv teilbar zu machen. Dadurch, dass die Vorstellungsräume an die physischen Ausprägungen des *reality space* angepasst sein müssen und dieser ständig mindestens latent kognitiv aktiviert bleibt und darüber hinaus auch die Deixis am Phantasma auf ein Bezugssystem im Wahrnehmungsraum referiert, konnte bisher kaum von einer Origoversetzung in den Vorstellungsräume gesprochen werden. Die abschließenden Beispielsequenzen widmen sich daher dem Aspekt, wie sich eine schrittweise und kollaborative Origoverschiebung in einen mentalen Raum entwickeln kann. Wie in den vorangegangenen Kapiteln werden Sequenzen untersucht, in denen sich das gemeinsame Imaginieren auf syntaktischer, vor allem aber auf gestischer Ebene zeigt und in ko-konstruktiven Handlungen widerspiegelt. Die lokalen mentalen Räume, die sich im Folgenden ad-hoc im Gespräch entwickeln, zeichnen sich durch einen höheren Grad an szenischer *Animiertheit* aus, als sie in den bisherigen Beispielen beobachtet werden konnten. Die körperliche Expressivität der Sprecher*innen kann dabei nach Ehmer (2011: 318) als Indiz für die Origoversetzung in den mentalen Raum bzw. in die Perspektive einer Rolle dienen. Mit einer Verschiebung der Origo in die Vorstellung – bei Bühler ([1934] 1999: 134) dem *Hingehen ins Phantasma* – übernimmt der*die Sprecher*in als perspektivisches Zentrum die vorgestellte Szene und agiert aus ihr heraus. Wichtig ist, dass für die erfolgreiche Vermittlung einer Imagination, auch die Rezipierenden die Versetzung in die Vorstellung vornehmen müssen (Ehmer 2011: 17). Dies wird in der folgenden Sequenz deutlich.

Ana ist damit beschäftigt, die Rückseite des Hexenhauses auf einem Papier zu skizzieren, die, wie in der Sequenz *Wände* besprochen, Teil des aufklappbaren Bühnenbilds ist. Die Skizze zeigt drei Bäume, perspektivisch vor der Rückseite des Hexenhauses. Tim schlägt vor, das Wurzelwerk der Bäume so anzulegen, dass aus ihnen der Käfig entsteht, in dem Hänsel im Verlauf der Oper gefangen sein wird. Er versucht, seine Idee durch Aufzeichnen auf einem

weiteren Papier zu verdeutlichen, bricht dies aber ab und gibt zu bedenken, dass er nicht glaube, dass seine Idee verständlich transportierbar sei. Anas Reaktion darauf bildet den Anfang des Transkriptausschnitts.

(4) Astgefängnis ((00:31:30 - 00:31:55 min))

01 Ana: naja ich WEISS nich mal-=
 02 =es REICHT ja auch wenn man-=
 03 =n weißt du am ANfang,
 →04 is das der WALD und dann schlafen sie so hIer,
 05 Tim: ja,
 06 Ana: weil es sollte auch schon n bIsschen realistisch sein das sie
 da_ähm UNten schlafen müssen,
 07 weil das is so UNheimlich glaub ich.
 08 Tim: ja_ja;
 09 Ana: aber was is wenn man dann HIER zwischen dIesen bäumen das macht,
 →10 und dann kommt da einfach so ÄSte kommen da so;
 →11 Tim: (0.8) die kommen da so RAU[S;]
 12 Ana: [ja.]
 →13 die kommen einfach so RAUS und_und_[und machen so-]=
 →14 Tim: [und HALten ;]
 →15 Ana: =und HALten ihn oder so;
 16 Tim: ja:-
 17 nich schl[Echt,]
 18 Ana: [weil die] [kann man auch sElber sich NEHmen,]
 19 Tim: [die muss aber auch wieder ne gewisse]
 DREIdimensionalität haben [ne,]
 20 Ana: [ja.]

In den Zeilen 01 bis 07 elaboriert Ana einen in der Vergangenheit liegenden mentalen Raum, der den Aspekt der beginnenden Oper umfasst, wo Hänsel und Gretel sich im Wald verlaufen haben und daher im unheimlichen Wald schlafen müssen. Als *space builder* dient hier vor allem die Konstruktion am Anfang (Z. 03). In Zeile 04 deutet sie parallel zu *is das der WALD ad oculus* auf die Skizze, die für die folgende Interaktion den *quasi-gegroundeten* (siehe Kap. 4.1), physischen Abgleich dieses mentalen Vergangenheitsraums bildet. Mithilfe der deiktischen Mittel fokussiert sie die Interaktion auf den imaginierten Waldraum im Kontext des Bühnenbilds und im Wahrnehmungsraum. Noch in derselben Intonationsphrase verweist sie nochmals *ad oculus* mit *und dann schlafen sie so hIer*, (Z. 04) lokal auf die Zeichnung, anaphorisch aber in den Vergangenheitsraum. Wichtig für die folgende Analyse ist daher das Verständnis der Skizze als deiktisches Medium, auf das *ad oculus* gezeigt werden kann, wodurch bestimmte Stellen auf dem Papier in den Fokus der Aufmerksamkeit gerückt werden. Gleichzeitig sind diese Verweise in die Skizze als Deixis am Phantasma zu verstehen, weil sie auf Aspekte referieren, die im Bühnenbild-Basisraum stattfinden.

In Zeile 09 aktualisiert Ana mit der Fokussierungsaufforderung durch das akzentuierte Lokaldeiktikon *HIER* (dazu Stukenbrock 2015: 52) und mit ihrer nach unten, in Richtung

Zeichnung geöffneten Hand die deiktische Verankerung des neuen Fokusraums im gezeichneten, physischen Abgleich des mentalen Waldraums. Die visuelle Aufmerksamkeit ist verbal und gestisch auf den Wahrnehmungsraum der Zeichnung gerichtet. Gleichzeitig evoziert der als Frage und durch das konditionale *wenn* gefärbte Vorschlag einen hypothetischen Vorstellungsraum und *HIER* erhält auch anaphorisch zum Vergangenheitswaldraum in Zeile 04 sowie als lokaler Bezugspunkt an den Bäumen im Zeichenraum eine multireferenzielle Bedeutung. Mit *diesen* (Bäumen) (Z. 09) wird noch in derselben Intonationsphrase ein weiteres, deiktisch verwendetes Demonstrativpronomen akzentuiert, das *ad oculus* und zeitgleich taktstockartig betonend mit beiden nach unten geöffneten Händen auf die Baumskizzen im Zeichenraum verweist.

Weitere Komplexität gewinnt die deiktische Handlung in Zeile 10, in der Ana mit ihrem Stift redegleitend zu und dann kommt da einfach so Äste die Bewegung gestisch am Phantasma der auf der Zeichnung (noch) nicht vorhandenen Äste und gleichzeitig *ad oculus* auf dem Papier symbolisierend nachfährt. Noch in perspektivischer Distanz zum Vorstellungsraum, beschreibt Ana gestisch und verbal die Handlung der Äste und evoziert damit eine weitere Dimension visuell-gestischer Zugänglichkeit mentaler Strukturen: Die Demonstration nicht-akteurgebundener Vorgänge im Vorstellungsraum (Ehmer 2011: 315). Dass die neuen Aspekte des Vorstellungsraums neben der deiktischen Multireferenzialität nicht nur aus Analysierendenperspektive komplexe Verstehensprozesse verlangen, wird an der folgenden empirisch nachweisbaren und multimodalen Bearbeitung des evozierten mentalen Raums sichtbar.

Zunächst versichert Ana sich, mit Abschluss ihrer Intonationsphrase in Zeile 10, durch einen „Kontrollblick zur Wahrnehmungswahrnehmung“ (Stukenbrock 2015: 87) Tims Aufmerksamkeit auf den aktuellen Fokusraum. Da der Kontrollblick von Ana im vorliegenden Fall vor allem im Sinne einer Verständnissicherung gedeutet werden kann, endet die Wahrnehmungshandlung hier nicht, sondern sie wird als „Wahrnehmung der Wahrnehmungswahrnehmung“ (ebd.: 89) durch Tim bestätigt, der durch seinen Blick auf den Fokusraum diesen explizit evaluiert. Die Verstehensdokumentation als sozio-kognitive Entsprechung zu seiner Wahrnehmung (ebd.: 92) macht Tim im Folgenden noch expliziter, wenn er gesamtkörperlich in die Demonstration des zuvor von Ana symbolisierend evozierten mentalen Raums einsteigt. Durch die ikonische Modifizierung verschiebt sich die Origo der Handlung in den mentalen Raum. Anaphorisch kongruent übernimmt Tim mit der zur Demonstration alignierten verbalen Beschreibung *die kommen da so RAUS:;* (Z. 11) den symbolisierenden Sprechmodus von Ana.

Standbild 08



09 Ana: aber was is wenn man dann HIE#R
zwischen dIesen bäumen das macht,
(Zeigegeste in den Skizzenraum)

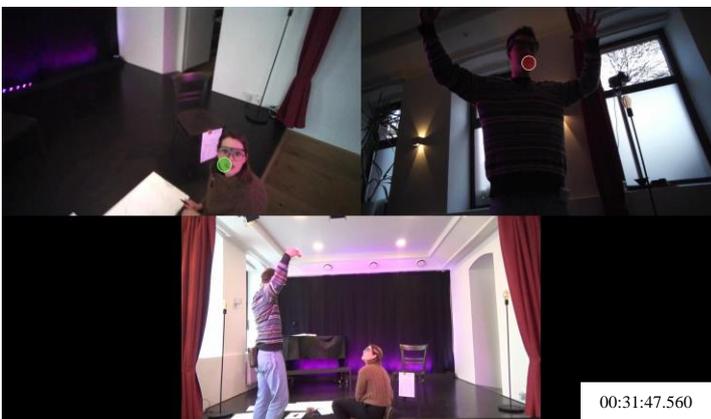
Standbild 09



10 Ana: und dann kommt da einfach
so ÄSte kommen da so;#

((Blick & Ausrichtung: Ana(rot): Wahrnehmungswahrnehmung
Tim(grün): Wahrnehmung der Wahrnehmungswahrnehmung))

Standbild 10



11 Tim: (0.8) die kommen da so RAU#[S:;]

Mit Äußerung der nukleusakzenttragenden Silbe RAUS: führt Tim am Gipfelpunkt der Demonstration eine beidarmigen, wellenartigen Geste über seinen Kopf aus, die die Bewegung der Äste nachahmt. Die körperlich-räumliche Repräsentation dieses mentalen Raums wird durch Ana – hier in der Ratifizierungsposition – noch überlappend mit einer Zustimmungspartikel als zutreffende Interpretation ihrer (also Anas) Vorstellung bestätigt (Z. 12).

Die folgenden Äußerungen (Z. 13–15) sind gleich durch mehrere Aspekte und auf multimodaler Ebene als kollaborative Handlungen zu identifizieren, die nicht nur Zustimmung und Verstehen anzeigen, sondern auch der schrittweisen, gestisch zunehmend animierten Ausgestaltung des mentalen Raums dienen. Leicht paraphrasierend wiederholt Ana auf verbaler Ebene die Konstruktion und in ihrer Gestik die Demonstration von Tim in Zeile 11. Mit Abschluss der gestischen Demonstration von Tim setzt Ana ebenfalls beidarmig, aber von ihrer Körpermitte ausgehend an, die ‚herauskommenden Äste‘ ikonisch abzubilden. Dabei expandiert Ana ihre Paraphrase und das wird besonders auf der gestischen Ebene deutlich: Auf Verbalebene möglicherweise nach Worten suchend (dreifache Wiederholung von und als Verzögerungssignal, gefolgt von einer in der Projektion des Satzes als Planungsproblem zu identifizierenden Konstruktion (Brenning 2015: 82)), zeigt sie durch eine Art umarmende, weiterhin beidarmige Geste Richtung Körpermitte, was auf verbaler Ebene kaum zum Verständnis gereicht hätte (Stukenbrock 2010: 15). Unterstrichen wird dieser Eindruck durch den gleichbleibenden Intonationsverlauf, der in Fällen wie dem vorliegenden, Unabgeschlossenheit vermuten lässt (Brenning 2015: 80).

Standbild 11



13 Ana: die kommen einfach so RAUS
und_und_[#und machen so-]=

Standbild 12



14 Tim: [und HALten;#]

Tim retrahiert zur Konjunktion und bietet die gesuchte Konstituente noch in Überlappung und in fallendem Tonhöhenverlauf mit [und HALten;] (Z. 15) als Verstehensdokumentation (Deppermann 2008: 235) an. Noch expliziter wird seine Verstehensbekundung, als er auf der letzten Silbe seiner Vervollständigungskonstituente wiederum die Geste von Ana in seine eigene größere, gesamtkörperlichere Inszenierung des mentalen Astgefängnisses aus Zeile 11 (siehe Standbild 12) überführt und das ‚(Fest)Halten‘ mit seitlich ausgebreiteten Armen abbildet. Wie der Vergleich der Standbilder 11 und 12 zeigt, ist die ikonische Demonstration der ‚Halten‘-Szene bei Tim aber durchaus anders ausgestaltet als bei Ana. Dass sich jedoch beide mit ihrer gestischen Interpretation auf die gleiche Imagination beziehen und die abweichende Demonstration für die kollaborative Repräsentation problemlos ist, zeigt sich in Zeile 15, in der Ana im schnellen Anschluss und mit der Wiederholung von =und HALten Tims verbale und gestische Vervollständigung ratifiziert und darüber hinaus die eigene syntaktische Gestalt zum Abschluss bringt. Durch ihren Objektnachtrag (und HALten ihn oder so; (Z. 15)), löst sie die syntaktische Projektion der Äußerung ein¹⁸ und modifiziert gleichzeitig den mentalen Astgefängnisraum verbal um einen weiteren, bisher nur latent im gestischen Entwurf des Raums vorhandenen Aspekt: Die Evozierung dieses mentalen Raums als *Demonstration* einer spezifischen Handlung der Hänsel-Rolle. Dass ihn (Z. 15) Hänsel bezeichnen muss – ohne dass dieser zuvor explizit als Teil des mentalen Raums eingeführt worden wäre – ist eindeutig, weil der *Frame* des Märchens bzw. der Oper nur Hänsel als männliche Rolle in der aktuellen Einbettung der Szene zulässt. Der Objektnachtrag kontextualisiert auch die Handlungsdemonstration im mentalen Raum neu. So wird deutlich, dass im Übergang der gestischen Demonstration von ‚herauskommenden Ästen‘ zu ‚(fest)halten‘, auch ein Übergang der Inszenierung nicht-akteurgebundener Handlungen zur körperlichen Animation einer im mentalen Raum verankerten Figur vollzogen werden kann. Die körperliche Ebene kann dabei als Indiz für die Origoverschiebung der Sprecher*innen gelten, die sich in eine imaginierte Szene bzw. Figur versetzen und diese körperlich animieren, ohne den symbolisierenden Modus der Verbalebene zu verlassen (dazu auch Ehmer 2011: 318).

Anders als in der Sequenz *Wände*, in der sich *Deixis am Phantasma* und *Demonstratio ad oculus* in selber Referenz von mentalem Objekt, Suchraum und Zeigeziel synchronisieren, ist die gestisch etablierte Vorstellung in der Sequenz *Astgefängnis* szenischer und deiktisch

¹⁸ oder so; (Z. 15) könnte als Vagheitsmarkierer interpretiert werden, der eine latente Dispräferenz von ‚halten‘ als semantische Einlösung der syntaktischen Gestalt äußern würde. Ich tendiere allerdings dazu, die vagheitsmarkierende Wirkung dieses Markers hier zugunsten einer auch strukturell beobachtbaren, allgemeinen Tendenz finaler Vagheitsmarkierer als redebeitragsabschließend (dazu Auer (im Druck) Kap. 6) zu vernachlässigen.

innerhalb der Interaktionspartner*innen versetzt. Den verbal höchst kollaborativen Entwurf des mentalen Raums animieren die Sprecher*innen gleichzeitig körperlich unterschiedlich. Indexikalität erhält der Vorstellungsraum hier durch das Personalpronomen *ihn*, das die Origo der Handlung von der der Sprecher*innen different konstituiert und die Gestik als Versetzung in die gleiche Rolle kontextualisiert.

Das folgende Beispiel zeigt, funktional ähnlich zur vorangegangenen Sequenz, wie durch den stets aktivierten Frame *Hänsel-und-Gretel-Oper* imaginierte Figuren latent etabliert für bestimmte Szenen sind, sodass die Teilnehmenden akteurgebundene Aktionen imaginierter Figuren uneingeleitet demonstrieren können. Ungleich komplexer als die obige Sequenz erscheint die folgende Interaktion dadurch, dass die Teilnehmenden zusätzlich zwischen einzelnen Rollenhandlungen der imaginierten Szene wechseln und verschiedene Figurenorigos etablieren. Die Animation der Figuren ist zunächst wieder vornehmlich auf der körperlichen Ebene zu beobachten, während der Aufbau der Szene auf verbaler Ebene beschrieben wird. Das interaktive Potenzial der folgenden Sequenz entfaltet sich vor allem in der wechselnden körperlichen Animation derselben imaginierten Figuren.

Der Ausschnitt stammt aus dem letzten Drittel der Aufnahme und ist eingebettet in die Diskussion darüber, wie und wo der Backofen Teil des Hexenhauses werden soll und welche Aspekte bezüglich des Opernverlaufs dafür beachtet werden müssen. Weil Gretel die Hexe am Ende des letzten Akts in den Ofen stößt, ist es wichtig, dass der Ofen mindestens so groß ist, dass eine Person hinein- bzw. hindurchpasst. Mit Tims Einwurf und der Frage, wie die Hexe publikumswirksam (was sieht denn auch BESSer aus-| wenn man die hEXe dann da nachher REINschmeißt; (Z. 01–02)) am besten in den Ofen befördert wird (vgl. Z. 01–04), führt er explizit die Hexe als raumkonstitutive Größe ein. Der gesamte folgende Ausschnitt bietet einen in sich geschlossenen Einschub zur aktuellen Aushandlungssituation (den Ort des Ofens im Bühnenbild). Das wird auch durch den *future event space builder* nachher (Z. 02) deutlich, mit dem Tim die folgende Interaktion und szenische Ausarbeitung als zukünftiges Event im Verlauf der Oper einleitet.

(5) Hexe ((00:47:28 – 00:47:56 min))

01 Tim: wAs sieht denn auch BESSer aus-
 02 wenn man die hEXe dann da nachher REINschmeißt;
 03 das so_n bisschen SEITlich zu machen,=
 04 =oder auch von HINTen;
 →05 das man wirklich [SO-] (0.6)
 06 Ana: [DAS]
 →07 [ja das]man von HINTen sie so-

08 Tim: [(ploppt mit den Lippen)]
 09 Ana: dAs die WEISST du das die
 →10 die hExe steht SO,
 11 und schaut REIN,
 12 (0.3)
 →13 sagen wir SO==
 14 =und [dann](.) hÄnsel und GREtel steh
 15 Tim: [<<p>JA;>]
 16 Ana: oder GREtel is_es ja nur,
 17 steht entweder HIER,=
 18 =oder auf DER seite,
 19 (0.4)
 →20 und macht so (--) <<creaky,f>!BA:M!,>=
 21 =und dann fliegt sie nach hinten REIN?
 22 [und dann is]_es LEbensgroß sozusagen,=
 23 Tim: [<<p>JA;>]
 24 Ana: =LEUCHtet vielleicht von innen ein bisschen-
 25 Tim: [geNAU;]
 →26 Ana: [und dann so] [<<creaky,h>AH_ah_ah,>]
 27 Tim: [und QUA]L:M-
 28 [nE]bel [und so-]
 29 Ana: [EAsy;]
 30 [JA;]
 31 (0.9)
 32 Tim: ja is SUpEr;

Tims verbale Einleitung in den Vorstellungsraum findet funktional als Frage und in der direkten Face-to-Face Zuwendung der Interaktionspartner*innen statt. Ohne ein Rezeptionssignal von Ana abzuwarten, paraphrasiert er im Verlauf der Zeilen 05 bis 08 mit seiner gesamtkörperlichen Inszenierung von ‚reinschmeißen‘ – und pragmatisch projiziert durch die Modaldeiktikonkonstruktion (Z. 05) – seine einleitende Frage durch eine körperliche Demonstration. Tims Geste respektive sein Körper bildet hier das Medium der Rollenanimation, die bisher höchstens durch das man (Z. 02 & 05) als weitere Figur dieser Szene, neben der Hexe, implizit thematisiert wurde. Vor allem aber zeigt das man eine Dissoziierung von Sprecher und Rolle bzw. Szenenhandlung an, sodass die unterschiedlichen Modi der körperlich demonstrativen und verbal symbolisierenden Animation der Vorstellung eindeutig nachvollziehbar sind. Durch die Demonstration seiner Vorstellung, wird diese leichter teilbar (Gerrig & Clark 1990), was sich auch in der Verstehensdokumentation widerspiegelt, die Ana durch ihre verbale Interpretation in Zeile 07 anbietet [ja das]man von HINten sie so-.

Der Konstruktionstyp des Modaldeiktikons *so*, mit gestisch-ikonischer Inszenierung einer Rollenhandlung (wie von Tim in Zeile 05 verwendet), scheint für die folgende Interaktion eine gewisse aktivitätsstrukturierende Projektionskraft zu besitzen. Auf diese Art der deiktisch-gestischen Verwendungsweise von *so* (dazu auch Stukenbrock 2010) wird noch vier weitere Male zurückgegriffen. Ich gehe deshalb hier genauer auf den Ablauf und mögliche Funktionen dieses Konstruktionstyps ein, der mit Stukenbrock „verbal isoliert“ (2010: 20) auf

„interpersonelle Koordination“ (ebd.: 15) sowie auf die multimodale Inszenierung von Figuren (Stukenbrock 2012) ausgelegt ist.

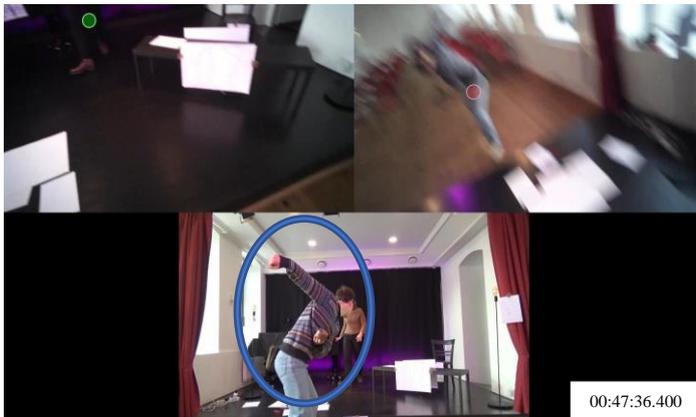
Parallel zum hauptakzentuierten SO (Z. 05) holt Tim mit beiden Armen seitlich rechts von seinem von der externen Kamera abgewandten Körper nach hinten aus, beugt dabei den Oberkörper und löst synchron mit seinem Lippenploppen (Z. 08) die Ausholbewegung in einer Schubbewegung nach vorne ein (siehe Standbild 13 im Höhepunkt der Ausholbewegung). Noch in Überlappung mit Tims SO (Z. 05), bemüht sich Ana in Zeile 06 um das Rederecht, bricht ihre Äußerung aber ab, bzw. setzt synchron zum Lippenploppen erneut an und paraphrasiert Tims Äußerung aus Zeile 04–05 eingeleitet mit einer Zustimmungspartikel. Phrasenbegleitend übernimmt sie räumlich gespiegelt auch die Körperausrichtung von Tim und setzt zur Wiederholung seiner Schubbewegung mit beiden Armen an. Der Raum, der sich dadurch zwischen beiden ergibt, erscheint als die physisch-räumliche Leerstelle der imaginierten Figur der Hexe. Diese wird zeitlich versetzt zwei Mal durch beide Sprecher*innen in einer körperlich demonstrativen Figurenhandlung einer noch nicht explizit verbal eingeführten Figur, in den imaginären Ofen geworfen. Nacheinander handeln hier beide aus der gleichen Figurenorigo, im gleichen Vorstellungsraum, wengleich aus unterschiedlichen Perspektiven im Wahrnehmungsraum.

Auf dem Höhepunkt ihrer Ausholbewegung des ‚in den Ofen Reinschiebens‘ setzt Ana aber auch zu einer neuen Intonationsphrase an, die eine Elaboration oder Spezifizierung der imaginierten Handlung einleitet (das die WEISST du das die (Z. 09)). Im Folgenden (Z. 10–24) entwickelt Ana auf verbaler Ebene im symbolisierenden Sprechmodus sequenziell einen Szenenablauf, den sie auf körperlich-gestischer Ebene gleichzeitig animiert. Dabei versetzt sie sich körperlich sowohl in die Figur der Hexe als auch in die Figur von Gretel, um visuell zu realisieren, was sie auf verbaler Ebene projiziert. Eingeleitet werden diese Animationen jeweils durch *so*-Konstruktionen (Z. 10, 13, 20, 26), die wie *space builder* fungieren und die Aufmerksamkeit auf die körperliche Performanz im *surrogate space* lenken. So demonstriert Ana in Zeile 10 und 11 (die hExe steht SO, | und schaut REIN,) Ausrichtung und Körperhaltung der Hexe im hypothetisch-zukünftigen mentalen Raum. Anschließend führt sie nun auch auf verbaler Ebene – nach einer selbstinitiierten Selbstreparatur, die Hänsel explizit aus dem lokalen mentalen Raum ‚löscht‘ (Z. 14–16) – Gretel als zweite zentrale Figur des mentalen Raums ein.¹⁹ Mit steht entweder HIER, =

¹⁹ Schon in der raumbildenden Turnkonstruktionseinheit von Tim am Anfang ist Gretel durch das geteilte Hintergrundwissen beider implizit Teil des mentalen Raums und wurde bereits von beiden Teilnehmenden durch die Schubbewegung animiert.

| =oder auf DER seite, (Z. 17–18) positioniert sich Ana in der körperlichen Animation der Gretelfigur explizit neu im *reality space* und verortet die Figurenorigo deiktisch im räumlichen Blend von Wahrnehmungsraum und Bühnenbildimagination sowie gleichzeitig abweichend von ihrer Sprecherinorigo. Daran anschließend animiert sie ‚als Gretel‘ erneut, was das Handlungszentrum des aktuellen mentalen Raums ausmacht: Das Hineinbefördern der Hexe in den Ofen. Der Perspektivenkontrast von Sprecher*in und Figur war auf verbaler Ebene bisher sehr deutlich kontextualisiert, und das vor allem durch die Behandlung der Figuren in der dritten Person. Im Folgenden verschwimmen diese Grenzen der Beschreibung auf verbaler Ebene und der Animation auf körperlicher Ebene jedoch durch zunehmende Expressivität auf sprachlicher Äußerungsebene (vgl. Ehmer 2011: 99). In Zeile 20, wieder eingeleitet durch die Demonstration projizierende Konstruktion und macht so unterstreicht Ana ihre körperliche Performance mit der onomatopoetischen Interjektion !BA:M!, (Z. 20), die sich auch auf der phonetischen Ebene nun deutlich von dem bisher verbal symbolisierend ausgestalteten mentalen Raum abhebt.

Standbild 13



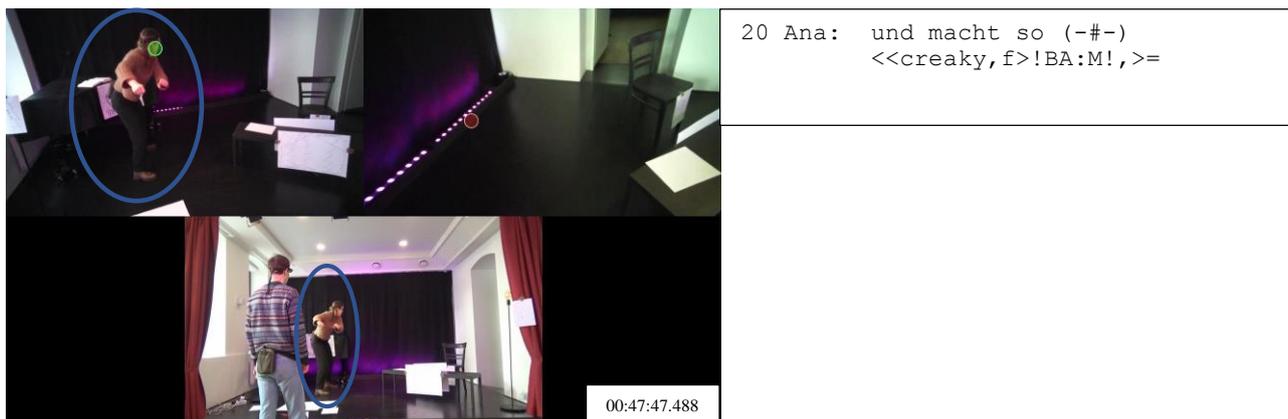
05 Tim: das man wirklich [SO-] (#0.6)

Standbild 14



07 Ana: [ja das]man von HINTen sie so-#

Standbild 15



In schnellem Anschluss wechselt sie aber in Zeile 21 auch wieder in den symbolisierenden Sprechmodus, zählt auf, was mit der Hexe dadurch im Folgenden passiere (Z. 21) und – in Überlappung mit Zustimmungspartikeln von Tim – welche Qualitäten der Ofen in dieser Szene benötige (Z. 22–25). Die Konsequenzen, die daraus entstehen, demonstriert sie daraufhin wiederum verbal und animiert in der Origo der Figur das Krächzen der verbrennenden Hexe auf verbaler und phonetischer Ebene [<<creaky, h>AH_ah_ah, >] (Z. 26). Diesen Wechsel in den animierenden Modus leitet die Sprecherin durch eine *quotativähnliche* Konstruktion²⁰ ohne Personalpronomen ein ([und dann so] (Z. 26)) und kontextualisiert die Figurenanimation durch den Wechsel in ein höchst expressives prosodisches Format (erhöhte Lautstärke, hohe Sprechspannung, kehliges, krächzendes Stimmregister) auf lexikalisch rein interjektiver Ebene. Interessant ist, dass die Animation auf verbaler Ebene mit dem funktional gleichen Konstruktionstyp wie die körperlichen Performances aus Zeile 10, 13 und 20 eingeleitet wird. Nur wird in Zeile 26 nicht der Körper der Sprecherin zum projizierten Verweisraum des Modaldeiktikons (vgl. Stukenbrock 2010: 15), sondern die performative verbale Äußerung. Zum Zeitpunkt der verbalen Animation befindet sich die Sprecherin nicht mehr in der körperlichen Animation. Hier unterstreicht sie lediglich in einem *token space* redegleitend das Krächzen der Hexe mit vor ihrem Körper wackelnden Händen.

Zusammenfassung

In diesem Kapitel wurde die körperliche Demonstration als Animation eines *surrogate space* dargestellt, der imaginierte nicht-akteur- und akteurgebundene Handlungen lebensgroß im Wahrnehmungsraum verankert. Im Unterschied zu den bisherigen Beispielen, konnte anhand der Figurenanimation die Verschiebung der Sprecher*in-Origo in den mentalen Raum der

²⁰ Zu Form und Funktion von Quotativkonstruktionen im Deutschen verweise ich auf Auer (2006).

imaginierten Szene bzw. Figur beobachtet werden. Eine einmal etablierte Figurenorigo scheint dabei kognitiv präsent zu bleiben, sodass die Sprecher*innen im Interaktionsverlauf mehrfach und ohne explizite Einleitung aus der Origo derselben Figur agieren. Besonders anschaulich ließen sich die Origoverschiebungen in der letzten Sequenz anhand der einleitenden Modaldeiktikonkonstruktionen nachvollziehen. Die Animation einer Figur kann in verschiedenen Modi unterschiedliche Grade der Expressivität aufweisen. Im vorliegenden Kapitel betrifft dies vornehmlich die Dissoziation der körperlich-pantomimischen Inszenierung einer Figurenhandlung vom symbolisierenden Modus der verbalen Äußerungsebene. Die Wahrung der symbolisierend-beschreibenden Ebene bewirkt die kognitiv-semantische Abgrenzung von Sprecher*in und Figur bei gleichzeitiger gestischer Verkörperung.

Unterschiedliche Faktoren des gemeinsamen Handelns können so ko-expressiv stattfinden. Neben syntaktischen Aspekten der Phrasenvervollständigungen und paraphrasierten Verstehensdokumentationen konnte vor allem anhand der körperlichen Demonstrationen gezeigt werden, wie Wahrnehmung, Verstehen und Wiederholung körperlich konstituiert werden können, sodass der Eindruck einer *shared cognition* entsteht. Die körperliche Übernahme der gleichen Figurenorigo in der gestischen Wiederholung der Figurenhandlung erscheint in beiden Sequenzen als Höhepunkt geteilter Imagination.

5. Fazit und Ausblick

Die vorliegende Arbeit hat die interaktiven Mittel zur Herstellung geteilter Vorstellung und gemeinsamer Imagination untersucht. Es konnte gezeigt werden, dass die mentale Repräsentation eines Raums bzw. einer Szene an die Stelle der ‚realen‘ Wahrnehmungssituation treten und so als aktuelle Realität erlebt und behandelt werden kann. Dabei spielt es keine Rolle, ob die mentalen Räume einen hypothetischen, vergangenen oder zukünftigen Status im Bezug zum Hier-und-Jetzt der Wahrnehmungssituation einnehmen.

Unter Einbezug der Erkenntnisse und Konzepte aus dem Theorieteil der Arbeit zeigen die Analysekapitel drei übergeordnete Dimensionen interaktiven gestisch-körperlichen Handelns beim gemeinsamen Imaginieren auf: Interpersonelle Einlösung gestisch-deiktischer Projektionen (Kap. 4.1), die Möglichkeit gleichzeitiger Bespielung eines gemeinsamen *surrogate space* durch körperliche Spiegelung (4.2) sowie die Versetzung in die gleiche Figurenorigo durch die körperliche Animation einer bestimmten Figurenhandlung (4.3).

Die körperliche Ebene der multimodalen Kommunikation bietet in der untersuchten Aufnahmesituation nicht nur den Vorteil, Dinge kommunizieren zu können, die schwer verbal

auszudrücken sind; ihre Leistung besteht auch darin, physische Repräsentationen kognitiver Strukturen in einer Interaktion aufzubauen, die durch das ihr zugrundeliegende spezielle Aufnahmesetting auf gemeinsame Imagination ausgelegt ist. Einsatz und Aufbau der körperlich-gestische Ebene dient hier also vor allem der physischen Repräsentation lokaler mentaler Räume, die perzeptuell ko-konstruiert werden können. Durch den Aspekt der ko-konstruierten Gestik können mentale Räume interaktiv entwickelt und origorelativ im Wahrnehmungsraum verankert werden. Deiktische Bezugspunkte können aber auch durch unterschiedliche Ebenen egozentrischer Involviertheit eines mentalen Raums kontextualisiert sein, bspw. dann, wenn eine imaginierte Handlung oder Figur *demonstriert* wird (Clark & Gerrig 1990). Der Grad der Animation einer imaginierten Figur kann dabei multimodal unterschiedlich expressiv gestaltet sein. Reflektieren verbale und körperliche Ebene dabei verschiedene Modi der Expressivität, betont das eine Dissoziation von Sprecher*in und imaginiertes Figur.

Die deiktische Gestik, die in den untersuchten Interaktionen sequenziell etabliert wird, wird häufig multireferenziell eingesetzt und verweist gleichzeitig *ad oculus* im *reality space* und in den Vorstellungsraum. Ergebnis dieser Multireferenzialität sind *blending*ähnliche Prozesse, bei denen der *reality space* mit mentalen Entitäten überschrieben wird und ein neuer Basisraum entsteht, der temporär durch gestisches *grounding* identifizierbar ist und funktional in dem durch Skizzen abgesteckten Bühnenbildraum kulminiert.

Die besondere Leistung dieser Arbeit besteht darin, statische kognitive Modelle auf multimodale Aspekte gemeinsamen sprachlichen Handelns anzuwenden. Ein Ausblick auf potenziell weiterführende Forschungsparadigmen ist deshalb schwierig, weil anzunehmen ist, dass die hochfrequente, interaktional abgestimmte gestische Ausarbeitung der mentalen Räume maßgeblich auf die spezielle Aufnahmesituation zurückzuführen ist. Allgemein vielversprechend erscheint aber die interdisziplinäre Anwendung statischer, kognitiv-linguistischer Konzepte in empirischen Analysen, um, wie bspw. im vorliegenden Fall, Ansätze gemeinsamer Kognition als interaktiv hergestellte semantische Strukturen zu erfassen.

6. Literaturverzeichnis

- Auer, Peter (2000): On-line Syntax. Oder: was es bedeuten könnte, die Zeitlichkeit der mündlichen Sprache ernst zu nehmen. In: *Sprache und Literatur* 31(1), 43–56.
- Auer, Peter (2006): Construction Grammar meets Conversation: Einige Überlegungen am Beispiel von “so“-Konstruktionen in der Interaktion. In: Günthner, Susanne/ Imo, Wolfgang (Hrsg.). *Konstruktionen in der Interaktion*. Berlin: De Gruyter.
- Auer, Peter (2007): Syntax als Prozess. In: Hausendorf, Heiko (Hrsg.). *Gespräch als Prozess. Linguistische Aspekte der Zeitlichkeit verbaler Interaktion*. Tübingen: Narr.
- Auer, Peter (Hrsg.) (2013): *Sprachwissenschaft. Grammatik – Interaktion – Kognition*. Stuttgart/ Weimar: Metzler.
- Auer, Peter (im Erscheinen): *Online-Syntax. Eine Anleitung zur Analyse der gesprochenen Sprache*. Stuttgart: Metzler.
- Barcelona, Antonio/ Valenzuela, Javier (2011): An overview of cognitive linguistics. In: Barcelona, Antonio/ Javier Valenzuela/ Mario Brdar/ Stephan Th. Gries/ Milena Zic Fuchs (Hrsg.). *Cognitive Linguistics: Convergence and Expansion*. Amsterdam: John Benjamins.
- Brenning, Jana (2015): *Syntaktische Ko-Konstruktionen im gesprochenen Deutsch*. Heidelberg: Universitätsverlag Winter.
- Bühler, Karl ([1934] 1999): *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Stuttgart: Lucius & Lucius. (3. Aufl., ungekürzter Neudruck der Ausgabe Jena, Fischer 1934).
- Clark, Herbert/ Gerrig, Richard (1990): Quotations as Demonstrations. In: *Language* 66(4), 764–805.
- Deppermann, Arnulf. (2008): Verstehen im Gespräch. In: Kämper, Hedrun/ Eichinger, Ludwig (Hrsg.). *Sprache - Kognition - Kultur: Sprache zwischen mentaler Struktur und kultureller Prägung*. Berlin/ Boston: De Gruyter.
- Ehlich, Konrad (1979): Verwendung der Deixis beim sprachlichen Handeln. Linguistisch-philologische Untersuchungen zum hebräischen deiktischen System. In: *Nordic Journal of Linguistics* 6(2), 157–162.
- Ehmer, Oliver (2011): *Imagination und Animation: Die Herstellung mentaler Räume durch animierte Rede*. Berlin/ New York: De Gruyter.
- ELAN (Version 6.3) [Computer software]. (2022): Nijmegen: Max Planck Institute for Psycholinguistics. The Language Archive. <https://archive.mpi.nl/tla/elan/download> .
- Fauconnier, Gilles ([1985] 1994): *Mental spaces: Aspects of meaning construction in natural language*. Cambridge: Cambridge University Press. (2nd Edition).
- Fauconnier, Gilles ([1997] 1999): *Mappings in Thought and Language*. Cambridge: Cambridge University Press. (Reprint).

- Fauconnier, Gilles/ Turner, Mark (2002): *The way we think: Conceptual blending and the mind's hidden complexities*. New York: Basic Books.
- Fillmore, Charles J. (1985): Frames and the semantics of understanding. In: *Quaderni di Semantica* 6, 222–254.
- Fricke, Ellen (2007): *Origo, Geste und Raum. Lokaldeixis im Deutschen*. Berlin: De Gruyter.
- Goffman, Erving (1963): *Behavior in Public Places. Notes on the Social Organization of Gatherings*. New York: Free Press of Glencoe.
- Graumann, Andrea (2013): Blending Theorie. In: Schierholz, Stephan/ Giacomini, Laura (Hrsg.). *Wörterbücher zur Sprach- und Kommunikationswissenschaft (WSK) Online*. Berlin/ Boston: De Gruyter. https://www.degruyter.com/database/WSK/entry/wsk_id80ef7232-ced1-49ad-8c1d-9cce2f0e227b/html [abgerufen am 04.02.2024].
- Hausendorf, Heiko (2003): Deixis and speechsituation revisited. The mechanism of perceived perception. In: Lenz, Friedrich (Hrsg.). *Deictic conceptualization of Space, Time and Person*. Amsterdam: Benjamins.
- Hoffmann, Thomas (2017): Multimodal constructs – multimodal constructions? The role of constructions in the working memory. In: *Linguistics Vanguard* 3(s1), <http://dx.doi.org/10.1515/lingvan-2016-0042>.
- Kendon, Adam (1980): Gesticulation and speech: Two aspects of the process of utterance. In: *The relationship of verbal and nonverbal communication* 25(1980), 207–227.
- Kendon, Adam (2004): *Gesture: Visible Action as Utterances*. Cambridge: Cambridge University Press.
- Kloiber, Rudolf/ Wulf Konold/ Robert Maschka (2016): *Das Handbuch der Oper*. 14. grundlegend überarb. Aufl. Kassel: Bärenreiter.
- Lakoff, George/ Johnson, Mark (1980): *Metaphors we live by*. Chicago: University of Chicago Press.
- Liddell, Scott K. (1995): Real, Surrogate, and Token Space. Grammatical Consequences in ASL. In: Emmory, Karen/ Reilly, Judy S. (Hrsg.). *Language, Gesture and Space*. Hillsdale: Erlbaum.
- Liddell, Scott K. (1998): Grounded blends, gestures, and conceptual shifts. In: *Cognitive Linguistics* 9(3), 283–314.
- Liddell, Scott K./ Metzger, Melanie (1998): Gesture in sign language discourse. In: *Journal of Pragmatics* 30(6), 657–697.
- McNeill, David (1992): *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
- McNeill, David (Hrsg.) (2007): *Gesture and Thought*. Chicago: University of Chicago Press.

- Mondada, Lorenza (2014): Pointing, talk, and the bodies. Reference and joint attention as embodied interactional achievements. In: Seyfeddinipur, Mandana/ Gullberg, Marianne (Hrsg.). *From gesture in conversation to visible action as utterance. Essays in honor of Adam Kendon*. Amsterdam/ Philadelphia: John Benjamins.
- Selting, Margret/ Peter Auer, Dagmar Barth-Weingarten/ Jörg Bergmann/ Pia Bergmann/ Karin Birkner/ Elizabeth Couper-Kuhlen/ Arnulf Deppermann/ Peter Gilles/ Susanne Günthner/ Martin Hartung/ Friederike Kern/ Christine Mertzlufft/ Christian Meyer/ Miriam Morek/ Frank Oberzaucher/ Jörg Peters/ Uta Quasthoff/ Wilfried Schütte/ Anja Stukenbrock/ Susanne Uhmman (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10, 353–402.
- Streeck, Jürgen (2007): Geste und verstreichende Zeit: Innehalten und Bedeutungswandel der ‚bietenden Hand‘. In: Hausendorf, Heiko (Hrsg.). *Gespräch als Prozess. Linguistische Aspekte der Zeitlichkeit verbaler Interaktion*. Tübingen: Narr.
- Stukenbrock, Anja (2009): Herausforderungen der multimodalen Transkription: Methodische und theoretische Überlegungen aus der wissenschaftlichen Praxis. In: Birkner, Karin/ Stukenbrock, Anja (Hrsg.). *Die Arbeit mit Transkripten in Fortbildung, Lehre und Forschung*. Mannheim: Verlag für Gesprächsforschung.
- Stukenbrock, Anja (2010): Überlegungen zu einem multimodalen Verständnis der gesprochenen Sprache am Beispiel deiktischer Verwendungsweisen des Ausdrucks „so“. In: *InList* 47, <https://www.researchgate.net/publication/266997134>.
- Stukenbrock, Anja (2012): Zur Beredsamkeit des Körpers. Figurendarstellung und Figurenwissen als multimodale Alltagsinszenierung. In: Lampert, Fabian (Hrsg.). *Figurenwissen. Funktionalisierung und Repräsentation von Wissen bei der narrativen Figurendarstellung*. Berlin/ New York: De Gruyter.
- Stukenbrock, Anja (2014): Pointing to an ‘empty’ space: Deixis am Phantasma in face-to-face interaction. In: *Journal of Pragmatics* 74, 70–93.
- Stukenbrock, Anja (2015): *Deixis in der face-to-face-Interaktion*. Berlin/ München/ Boston: De Gruyter.
- Turner, Mark (1991): *Reading Minds: The Study of English in the Age of Cognitive Science*. Princeton: Princeton University Press.
- Wittenburg, Peter/ Hennie Brugman/ Albert Russel/ Alex Klassmann/ Han Sloetjes (2006): ELAN: a professional framework for multimodality research. In: *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC’06)*. Genoa, Italy: European Language Resources Association (ELRA).
- Zima, Elisabeth/ Bergs, Alexander (2017): Multimodality and construction grammar. In: *Linguistics Vanguard*, 3(s1), <https://doi.org/10.1515/lingvan-2016-1006>.

Zima, Elisabeth (2021): *Einführung in die gebrauchsbasierte Linguistik*. Berlin: De Gruyter.

7. Anhang

Abbildungsverzeichnis

Abbildung 01: Modelldarstellung <i>blending</i>	7
Abbildung 02: Darstellung der Splitscreens mit Gazepunkten.....	14
Abbildung 03: Zusammenstellung Gestentranskriptionssystem.....	15
Abbildung 04: Abbildung des angefertigten Bühnenbilds.....	15
Abbildung 05: Skizze des Ofens im Bühnenbild.....	20
Abbildung 06: Modelldarstellung <i>grounded blends</i>	22
Standbild 01.....	21
Standbild 02.....	24
Standbild 03.....	28
Standbild 04.....	28
Standbild 05.....	28
Standbild 06.....	30
Standbild 07.....	30
Standbild 08.....	35
Standbild 09.....	35
Standbild 10.....	35
Standbild 11.....	36
Standbild 12.....	36
Standbild 13.....	41
Standbild 14.....	41
Standbild 15.....	41