

Karola Pitsch

# Mensch-Roboter-Interaktion als Forschungsinstrument der Interaktionalen Linguistik

## Situiertheit von Interaktion und das Design referenzieller Praktiken

**Abstract:** Ausgehend von einer Konvergenz zwischen den aktuellen Herausforderungen, die sich zum einen einer multimodal orientierten Interaktionalen Linguistik und zum anderen im Bereich der Social Robotics stellen, schlagen wir einen neuartigen methodologischen Ansatz vor. Dieser besteht darin, Mensch-Roboter-Interaktion als ein Forschungsinstrument der Interaktionalen Linguistik zu konzipieren. Dieses wird am Beispiel des Phänomens ‚Referenzielle Praktiken‘ und anhand von Daten aus mehreren Studien ausbuchstabiert, in denen ein humanoider Roboter (Nao) in der Rolle eines Museumsführers vorbeikommenden Besuchern Informationen zu ausgewählten Exponaten anbietet, auf die er sie durch multimodal gestaltete lokal-deiktische Verweise zu orientieren versucht.

**Keywords:** Interaktionale Linguistik, Mensch-Roboter-Interaktion, Social Robotics, semi-experimentelle Studien, Multimodalität, Situiertheit, Situated Action, referenzielle Praktiken, Deixis

## 1 Einleitung: Multimodalität und Situiertheit in der Interaktionalen Linguistik

In dem Maße, in dem in der Interaktionalen Linguistik (Couper-Kuhlen & Selting 2018) vermehrt die multimodale Eingebettetheit von Sprache als Ressource zur Organisation von Kommunikation und sozialer Interaktion wichtig wird (Mondada 2014, Keevalik 2018), stellt sich die methodologische Frage, wie das Zusam-

---

**Karola Pitsch**, Universität Duisburg-Essen, Institut für Kommunikationswissenschaft, Universitätsstraße 12, 45141 Essen, GERMANY, karola.pitsch@uni-due.de

menspiel der verschiedenen kommunikativen Ressourcen – Sprache, Blick, Kopf-orientierung, Gestik, Körperpositur etc. – systematisch beschrieben werden kann. Geht man grundlegend von der Situiertheit sozialer Interaktion (Goodwin 2000, Hausendorf 2015) aus, dann versteht man die emergierenden sprachlichen wie verkörperten Strukturen als komplexe holistisch-multimodale Gestalten, die von den Interaktionsbeteiligten dynamisch und flexibel hervorgebracht werden und als kontingente Lösungen zur Bearbeitung der anstehenden kommunikativen Aufgaben lokal mobilisiert werden. Im praktischen Handlungsvollzug werden sie im kleinschrittigen interaktiven Wechselspiel mit dem/den Ko-Partizipanten und im Rahmen der lokalen ‚ecology‘ sukzessive konfiguriert und rekonfiguriert (Goodwin 2000, Mondada 2014: 139), so dass ihre genaue Ausgestaltung situativ bedingt und damit nicht in ihren Details vorhersehbar ist (Schegloff 1996). Als methodologische Herausforderungen für eine moderne Interaktionale Linguistik stellen sich damit insbesondere die folgenden Fragen (vgl. Mondada 2014: 139; Keevallik 2018, Couper-Kuhlen 2018): Wie lässt sich das Zusammenspiel der verschiedenen Ressourcen als Teil von „multimodal packages“ unter verschiedenen Bedingungen beschreiben? Woran orientieren sich Interaktionsbeteiligte bei der Rezeption von multimodalen Gestalten und welche Funktionen übernehmen dabei die einzelnen Ressourcen? Wie kann man angesichts der Variabilität von multimodalen Gestalten systematische Kollektionen von vergleichbaren Fällen erstellen?

Diese aktuellen Fragen einer multimodal orientierten, von „situated actions“ ausgehenden Interaktionalen Linguistik stellen zugleich auch zentrale Herausforderungen im Forschungsbereich der Social Robotics bzw. Human-Robot-Interaction (Breazeal et al. 2016, Dautenhahn 2014) dar. So stellt Breazeal (2003: 167) fest: „The success of these robots hinges not only on their utility but also on their ability to be responsive to and interact with people in a natural and intuitive manner“. Verfolgt man das Ziel, Roboter, virtuelle Agenten und andere technische Systeme mit Ressourcen auszustatten, die es Menschen erlauben, diese intuitiv mit den Mitteln natürlichsprachlicher Kommunikation zu nutzen, stellt sich daher die Frage, wie das kommunikative und interaktive Verhalten des Roboters/Agenten gestaltet werden soll und wie dieses mit der Situiertheit sozialer Interaktion umgehen kann (vgl. Schegloff 1996).

An dieser interdisziplinären Schnittstelle setzt der vorliegende Beitrag an. Er verfolgt das Ziel, auszuloten inwieweit das Modellieren von Interaktionsverhalten für einen humanoiden Roboter und dessen Erproben in semi-experimentellen Studien zur Mensch-Roboter-Interaktion als Instrument für die Erforschung von multimodaler Kommunikation und sozialer Interaktion dienen kann (vgl. Pitsch

et al. 2012, 2013b, 2014a, 2014b 2016).<sup>1</sup> Dieser Ansatz greift zum einen Suchmans (2006: 186) Gedanken auf: „Just as the project of building intelligent artifacts has been enlisted in the service of a theory of mind, the attempt to build interactive artifacts, taken seriously, could contribute much to an account of situated human action and shared understanding.“ Zum anderen ist er anschlussfähig an aktuelle Vorschläge in Konversationsanalyse und Interaktionaler Linguistik, die die Fokussierung auf authentische Alltagsinteraktion um Studiendesigns mit gezielter Manipulation von Bedingungen in semi-experimentellen Labor-Experimenten ergänzen (Pitsch et al. 2014a, Kendrick 2017, de Ruiter & Albert 2017, Heath & Luff 2018). Dieses Ziel wird im Folgenden am Phänomen referenzieller Praktiken bzw. Lokal-Deixis und im Szenario eines Museums-Roboters nachgegangen: Wie lässt sich das Design von referenziellen Praktiken für einen humanoiden Museumsroboter von Erkenntnissen aus der menschlichen Interaktionsforschung inspirieren? Wie gehen Museumsbesucher in der „real world“ mit den kommunikativen Angeboten eines autonomen Roboters um?

Zur Beantwortung dieser Fragen werden wir zunächst den Ansatz der Konversationsanalyse und ihr Verhältnis zu methodischen Ansätzen der Manipulation von Interaktionsbedingungen beleuchten (Kap. 2) und auf dieser Basis unseren Ansatz präsentieren, Mensch-Roboter-Interaktion als ein neuartiges methodisches Instrument zur Erforschung von situierter multimodaler Kommunikation zu verwenden (Kap. 3). Auf einen kurzen Überblick zum Forschungsstand zu ‚Referenziellen Praktiken‘ und dessen zentralen Desideraten (Kap. 4) wird – nach einer Beschreibung der Studien (Kap. 5) – in zwei analytischen Schritten untersucht, wie Museumsbesucher die multimodal gestalteten referenziellen Verweise eines humanoiden Museumsroboters behandeln (Kap. 6 und 7). Abschließend werden kurz die Möglichkeiten und Grenzen des präsentierten Ansatzes diskutiert (Kap. 8).

---

<sup>1</sup> Die Studien wurden in den Projekten „Interaktion & Raum“ (Volkswagenstiftung, PI: K. Pitsch) und „Incremental Coordination in Human-Robot-Interaction“ (DFG-EXC-277 CITEC, PI: K. Pitsch & S. Wrede) unter Mitwirkung von R. Gehle, T. Dankert, M. Görlich und J.-C. Seele durchgeführt. Die Autorin dankt zudem Vivien Ebben für ihre Unterstützung bei der Anonymisierung der Standbilder.

## 2 Multimodale Kommunikation und soziale Interaktion: methodische Instrumente

Das Empirieverständnis der ethnomethodologischen Konversationsanalyse beinhaltet traditionell eine Fokussierung auf die Untersuchung von Daten aus authentischen Interaktionssituationen. In jüngerer Zeit zeigt sich darüber hinaus, dass ihre Forschungsergebnisse viel Anschlusspotenzial in interdisziplinären Kontexten bieten, wie z.B. Psycholinguistik, Robotik, Conversational AI etc., die sich mit interaktiven Fragestellungen in primär quantitativen oder modellierenden Forschungsparadigmen beschäftigen. Daher stellt sich die Frage, inwieweit die ethnomethodologische Konversationsanalyse gewinnbringend mit anderen Ansätzen und Paradigmen kombiniert werden kann (s. auch Pitsch et al. 2014a, Kendrick 2017).

### 2.1 Konversationsanalyse und die Manipulation von Bedingungen sozialer Interaktion

Entsprechend ihrer ethnomethodologischen Provenienz und ihres empirischen Grundverständnisses interessiert sich die Konversationsanalyse traditionell für die sozialen Praktiken und Verfahren, mit denen Interaktionsbeteiligte in authentischen Alltagssituationen ihre kommunikativen Aufgaben bearbeiten. Angesichts der Dynamik und Variabilität in der konkreten Ausgestaltung dieser Praktiken als multimodale Gestalten stellt sich – wie Mondada (2014: 139) formuliert – die grundlegende Frage, „how participants in social interaction mobilize a set of resources for the locally situated, intersubjective and methodic organization of action“. In methodologischer Hinsicht besteht also die zentrale Frage darin, herauszufinden, (i) wie die verschiedenen kommunikativen Ressourcen in je unterschiedlichen emergent-kontingenten Situationen zu multimodalen Gestalten konfiguriert und dynamisch re-konfiguriert werden, (ii) woran sich Interaktionsbeteiligte bei ihrer Rezeption orientieren, (iii) welche Funktionen die einzelnen Ressourcen in der je spezifischen Situation und Ökologie einnehmen und (iv) wie man unter den Bedingungen solcher Variabilität systematische Kollektionen von vergleichbaren Fällen erstellen kann (vgl. Mondada 2014; Stukenbrock 2015, Keevallik 2018, Couper-Kuhlen 2018).

Ein naheliegender Ansatz dafür, das Zusammenspiel der Ressourcen im Rahmen sozialer Praktiken tiefergehend zu erforschen, besteht darin, ein möglichst großes Korpus verschiedenster Alltagssettings zu untersuchen: „for the study of the respective calibration of multimodal resources, the diversification of settings



and activities to analyze is crucial, in order to go beyond face to face conversation and explore a variety of multimodal praxeological configurations“ (Mondada 2014: 140). In diesem Sinne ist z.B. Hirschauers (1999) Vorschlag, den Fahrstuhl als ein „soziologisches Forschungsinstrument“ (222) zu nutzen paradigmatisch. Durch die spezifische räumliche Konfiguration dieses „Soziotops“ (223) – d.h. insbesondere seine Enge und Umbauthet – werden spezifische Interaktionsbedingungen hergestellt und damit die Grundparameter der sozialen Ordnung in einem authentischen Alltagssetting variiert.

Ein alternativer Ansatz besteht in der gezielten Herstellung spezifischer Interaktionsbedingungen. Dieser steht allerdings in Gegensatz zum traditionellen Primat des Untersuchens authentischer Interaktionssituationen in der Konversationsanalyse. Erst in jüngster Zeit beginnt eine methodologische Diskussion darüber, inwieweit die Konversationsanalyse mit semi-experimentellen Arbeitsweisen kombinierbar ist (Kendrick 2017, de Ruiter & Albert 2017). Gleichzeitig kann dabei auf frühe und wegweisende Studien zurückgegriffen werden. Zum einen hat bereits Garfinkel (1967) in den „Breaching Experiments“ seine Interaktionspartner mit absichtlich unerwarteten Gesprächsbeiträgen konfrontiert, um so die Normalität sozialer Ordnung und Handlungen aufzudecken sowie die Verfahren sichtbar zu machen, mit denen sie die entstehenden Brüche bearbeiten. Zum anderen liegen Suchmans (1987/2006) wegweisender Studie zu „Plans and Situated Action“ Videoaufzeichnungen von Situationen zugrunde, die im Rahmen einer semi-experimentellen Studie entstanden sind: Zwei Personen werden zu Nutzern eines neuen, technischen Geräts (hier: Kopierer) und gebeten, für den Zweck der Untersuchung ein Set an definierten Aufgaben damit zu erledigen. Methodologisch zentral in unserem Kontext ist der Umgang mit diesen Daten (vgl. zur Argumentation Pitsch et al. 2014a): Die Art, wie die Teilnehmer die Aufgaben erledigen, ist nicht weiter vorgegeben, und die Fragestellung und Analyse der entstehenden Interaktionen folgen keiner quantifizierend-experimentellen Logik. Vielmehr werden anhand der entstandenen Daten die – unter den geschaffenen Bedingungen – authentischen Handlungen der Interaktionsbeteiligten ethnographisch-konversationsanalytisch untersucht und daran explorativ der grundlegende Unterschied zwischen den situierten und kontingenten menschlichen und den planbasierten Handlungslogiken des technischen Geräts herausgearbeitet. Auf diesen Gedanken aufbauend beschreiben auch Heath & Luff (2018) ihren explorativen Ansatz zu Design und Erforschung von Workplace-Technologien unter dem Label eines „quasi-naturalistic experiment“. Hierunter werden Labor- wie Feldexperimente verstanden, die – entsprechend dem Interesse der Autoren an sozialen Praktiken im Workplace-Setting – in die diesbezügliche ethnographisch-konversationsanalytische Feldforschung eingebettet sind. Sie werden als

Bestandteil eines „broader program of naturalistic research“ (467) mit genuin explorativem Charakter verstanden: „to explore and assess the impact of a particular set of practices, techniques, or technologies“ (467) und „to expose the unknown or unexpected aspects of social organization“ (469).

Vor dem Hintergrund dieser ersten Ansätze, semi-experimentelle Arbeitsweisen mit konversationsanalytischer Herangehensweise zu verbinden, argumentiert Kendrick (2017) programmatisch für eine Öffnung der Konversationsanalyse hin zu einem methodischen Pluralismus, der die naturalistische Beobachtung durch experimentelle Studiendesigns, Labor-Settings und Quantifizierung ergänzt. Als Gewinn der Überwindung der bisherigen methodischen Dichotomie führt er eine Reihe an neuen methodischen Möglichkeiten an. Diese konvergieren mit unserem Ansatz, Mensch-Roboter-Interaktion als ein Instrument der Interaktionalen Linguistik zu verstehen (vgl. Pitsch et al. 2012, 2013b, 2014a, 2014b, 2016, 2020): (a) das Überprüfen der Robustheit von Erkenntnissen aus menschlicher sozialer Interaktion unter anderen, spezifischen (hier: technologisierten) Bedingungen, (b) die interdisziplinäre Anschließbarkeit von konversationsanalytischen Erkenntnissen an Informatik und Social Robotics; (c) der Zugewinn neuer Aufzeichnungs- und Datenerhebungsmethoden; sowie (d) die Frage des Verhältnisses von Ergebnissen aus Alltagswelt und Laborkontexten. Die Möglichkeiten und Grenzen eines solchen Verfahrens sollen im vorliegenden Text eruiert werden.

## 2.2 Mediated Communication und Mensch-Maschine-Kommunikation

In der linguistischen wie gesprächsanalytischen Forschung besteht eine lange Tradition in der Berücksichtigung der medialen Bedingungen von Kommunikation, wie sie durch den Einsatz verschiedener Kommunikationstechnologien entstehen (vgl. Computer-Mediated Communication, Herring 1996). Dieses gilt sowohl für mündliche Kommunikationsformate vom Telefon über Videotelefonie bzw. Videokonferenz-Systeme und Media Spaces bis hin zu Augmented Reality-Szenarien als auch für an schriftsprachliche Bedingungen anschließende Formate wie Chat, SMS/WhatsApp oder Twitter. In diesen Arrangements werden jeweils durch die technologische Verfasstheit spezifische Bedingungen induziert, unter denen die menschlichen Interaktionsbeteiligten miteinander interagieren. Diese betreffen insbesondere verschiedene Dimensionen von Zugänglichkeit – etwa Zeitlichkeit, Sichtbarkeit und Räumlichkeit –, an deren Bearbeitung sich die Interaktionsbeteiligten in ihren Beiträgen orientieren. In einem solchen Ansatz der ‚Mediated Communication‘ ist allerdings der Umgang mit einem technischen

Agenten bislang weitgehend unberücksichtigt geblieben (s. aber Lotze 2016). Erst in jüngster Zeit plädieren Arminen et al. (2016) für eine breitere Perspektive auf Mediatisierung. Davon ausgehend, dass die Wahrnehmung der Welt immer mediatisiert ist, schlagen sie anstelle eines Dualismus zwischen ‚unmediatisierten‘ (face-to-face) und mediatisierten Situationen ein Kontinuum mit den zwei Polen „Medium Immersed“ und „Tool Using“ vor. Hierüber sollen verschiedene materielle Konstellationen abbildbar sein, die jeweils daraufhin befragt werden können, wie „the production of particular sequences may be accomplished, enabled, constrained, or inhibited“ (293). In einem solchen Kontinuum lassen sich aus Sicht der Autoren auch hybride/intermediäre Formen verorten, wie z.B. Mensch-Roboter-Konstellationen. Wenngleich also neuerdings die Konstellation Mensch-Agent/Roboter prinzipiell als ein Aspekt mediatisierter Kommunikation mit angedacht wird, ist allerdings noch völlig ungeklärt, wie der Spezifik einer solchen „sozio-technischen Konfiguration“ (Rammert 2008) Rechnung getragen werden kann. Welchen Status weisen die menschlichen Interaktionsbeteiligten einem Roboter zu? Wie (wenn überhaupt) wird er möglicherweise als Ko-Partizipant definiert?

Im Bereich der linguistischen Erforschung von Mensch-Maschine-Kommunikation stand bis dato primär die Art und Weise im Fokus, wie Menschen mit Technologie sprechen (z.B. Weingarten 1989, Fischer 2006, 2010, 2011). Die multimodale Eingebundenheit von Sprache sowie insbesondere die interaktive Dimension – d.h. die Koordinierung von Handlungen – blieben dabei weitgehend unberücksichtigt. In jüngerer Zeit beginnen Pitsch et al. (2012, 2013) zu zeigen, wie die verbalsprachlichen Erklärungen und manipulativen Handlungen eines menschlichen Tutors in ihrer Sequenzialität und konkreten Handlungsoptionen von den Aktivitäten des Roboters mit-konfiguriert werden. Ähnlich interaktiv fokussierend zeigen Porcheron et al. (2018) wie die menschlichen Nutzer im Umgang mit dem Voice Interface ‚Alexa‘ ihre Äußerungen so designen, dass sie in die jeweiligen, durch das System vorgesehenen Slots hineinpassen. Solche an die Situiertheit von Kommunikation anschließenden Herangehensweisen bieten neue Perspektiven für die linguistische Dimension von Mensch-Maschine-Kommunikation.

## 2.3 Konversationsanalyse und Entwicklung neuer Technologien

Im Rahmen der Workplace Studies hat die Konversationsanalyse ein traditionsreiches Interesse am Erforschen von neuen Technologien in interaktiven und sozialen Situationen wie auch am Mitwirken bei ihrer Gestaltung und Entwicklung

(Suchman 2006, Heath & Luff 2018, Dourish & Button 1998). So haben frühe Studien das Potential von konversationsanalytischen Ideen für die Gestaltung von text-basierten Dialogsystemen herausgestellt und vorgeschlagen, die Annahmen und Prinzipien von sequenziellen Strukturen und des lokalen Managements aus der Mensch-Mensch-Interaktion für die technische Modellierung zu übernehmen. Dieses anhand von Phänomenen wie Turn-Taking, Nebensequenzen oder Pre-Closings ausbuchstabierend, zielen Frohlich & Luff (1990:189) darauf „to see how productively the technology of conversation can be used to reproduce the details of actual, naturally occurring conversations between people, in conversations between people and computers“. Während solche Versuche einerseits als „simulacrum of conversation“ kritisiert wurden (Button 1990), bringen sie andererseits den Vorschlag ein, die Erkenntnisse über das Funktionieren von sozialer Interaktion für die Gestaltung von technischen Systemen fruchtbar zu machen.

Wenn wir uns heute mit der konzeptuellen Perspektive und dem methodischen Ansatz von Konversationsanalyse und Interaktionslinguistik im Bereich der ‚Social Robotics‘ beschäftigen und die Mensch-Roboter-Interaktion als ein Instrument für die Erforschung der Details von situierter Interaktion explorieren wollen, dann ist nicht das Ziel, die Möglichkeiten und Funktionsweise von menschlicher Kommunikation in einem technischen System zu reproduzieren. Vielmehr steht das Erforschen einzelner kleiner ‚Building Blocks‘ im Fokus, anhand derer das multimodale Zusammenspiel verschiedener kommunikativer Ressourcen im Rahmen der Dynamik von situierter Interaktion erprobt werden kann. Das Design der Strukturen ist dabei von den Erkenntnissen aus Interaktionslinguistik, Gesprächsforschung, Konversationsanalyse etc. inspiriert, wird aber in eine neue, nun sozio-technische Konstellation eingebunden, in der zunächst grundsätzlich eigenständige Bedingungen herrschen. Inwieweit die aus der menschlichen Interaktion bekannten Verfahren auch in diesem Kontext produktiv sind, ist eine offene, empirisch zu beantwortende Frage.

### **3 Mensch-Roboter-Interaktion als methodisches Instrument zur Erforschung von ‚Situating Action‘**

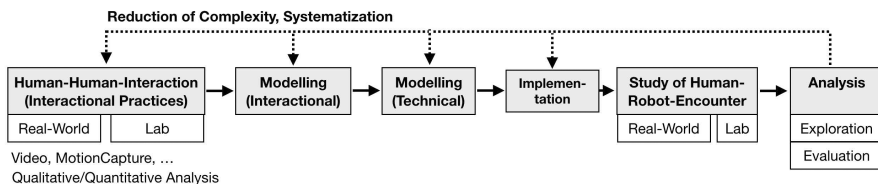
Basierend auf der Feststellung, dass einerseits eine zentrale Herausforderung in der interaktionslinguistischen Erforschung von multimodaler Kommunikation darin besteht, die Dynamik der Konfiguration von „multimodal packages“ zu verstehen, und andererseits eine methodische Öffnung der Konversationsanalyse zu

semi-experimentellen Ansätzen (vgl. Kendrick 2017) neue Möglichkeiten zur Frage der Robustheit interaktiver Phänomene bieten kann, schlage ich vor, Mensch-Roboter-Interaktion als ein methodisches Instrument zur Erforschung der Situiertheit von multimodaler Kommunikation und sozialer Interaktion einzusetzen. Ein solcher Ansatz ist an der interdisziplinären Schnittstelle von Interaktionaler Linguistik, Konversationsanalyse, Multimodaler Interaktion und Informatik, Robotik, Künstlicher Intelligenz angesiedelt und trifft auf die aktuelle Herausforderung im Bereich der Social Robotics, technische Systeme so mit kommunikativen Fähigkeiten auszustatten, dass sie vom menschlichen Nutzer mit Mitteln natürlicher Kommunikation bedienbar sind.

Verwendet man Mensch-Roboter-Interaktion als ein methodisches Instrument zur Erforschung von situierter multimodaler Kommunikation, so lassen sich Erkenntnisse auf verschiedenen Ebenen generieren. Diese umfassen mindestens die folgenden:

- (1) Grundlegende ‚Building Blocks‘ von situierter Interaktion mit ihren je verschiedenen Konstellationen des Zusammenspiels kommunikativer Ressourcen zu ‚multimodal packages‘ in Relation zu spezifischen Interaktionsaufgaben, Beteiligtenstrukturen, Teilnehmerinterpretationen und dynamisch-kontingenten Interaktionssituationen (Fokus in diesem Text),
- (2) Theorien und Konzepte von Interaktion,
- (3) sozio-technische Konstellationen,
- (4) menschliche Sozialität und moralische Ordnungen insbesondere in technikbezogenen Situationen (vgl. Pitsch 2020),
- (5) Integration neuer Technologien in die Alltagsökologie ihrer Nutzer und damit verbundene gesellschaftliche Fragen (vgl. Kramer et al. 2013).

Diese Erkenntnisse können nicht nur durch die Auswertung der Daten aus einer Studie zur Mensch-Roboter-Interaktion im Labor oder in der realen Welt gewonnen werden. Vielmehr betrifft der Ansatz, Mensch-Roboter-Interaktion als methodisches Instrument der Interaktionalen Linguistik einzusetzen, grundsätzlich alle Etappen eines iterativen Forschungsprozesses (s. Abb. 1 unten). Denn durch den interdisziplinären Zuschnitt dieses Unterfangens – bei dem Fragestellungen und Herausforderungen aus verschiedenen Disziplinen an einem gemeinsamen Gegenstand zusammentreffen – kann die konversationsanalytische Interaktionsforschung aufgrund ihrer Expertise in multimodaler Kommunikation und sozialer Interaktion einen Beitrag zur Gestaltung der kommunikativen Kompetenzen der technischen Systems leisten.



**Abbildung 1:** Mensch-Roboter-Interaktion als methodisches Instrument für die Erforschung von Kommunikation – iterativer Forschungsprozess

Dieser Zuschnitt betrifft (i) Erkenntnisse über kommunikative Verfahren und multimodale Ressourcen, die als Inspiration für insbesondere (ii) die interaktionale wie möglicherweise auch (iii) die technische Modellierung von Interaktionsfähigkeiten dienen können,<sup>2</sup> und auf (iv) der Implementierung in einem Roboter-System basierend (v) die Durchführung einer Studie zur Mensch-Roboter-Interaktion sowie (vi) die Auswertung der Daten, die entweder als Evaluation der implementierten Verfahren oder explorativ im Hinblick auf in der Situation neu zu entdeckende Phänomene erfolgen kann. Eine derartige Integration konversationsanalytischer Methoden, Konzepte und Erkenntnisse in den gesamten Forschungsprozess der Social Robotics hat gleichzeitig auch Auswirkungen auf ihre Arbeitsweise. Durch das Ziel der Modellierung entstehen neue methodische Herausforderungen für die Konversationsanalyse: Es werden neue Formen der Systematisierung und der Reduktion von interaktiver Komplexität notwendig, die für technische Disziplinen anschließbar sind. Dafür ist eine Kombination von qualitativen mit quantifizierenden Verfahren hilfreich;<sup>3</sup> und Videoaufzeichnungen von Interaktionsverhalten sind durch von technischen Systemen gut aus-

<sup>2</sup> Die grundlegende Unterscheidung zwischen interaktionaler und technischer Modellierung basiert auf der inspirierenden Diskussion der Autorin mit S. Wrede im gemeinsamen Projekt.

<sup>3</sup> Die Quantifizierung kann auf unterschiedlichen Ebenen ansetzen und ist künftig weiter in ihren verschiedenen Aspekten zu detaillieren. So sind z.B. Kenntnisse über die Häufigkeit des Auftretens eines Verfahrens im Vergleich zu anderen (vgl. Schegloff 1993) hilfreich, um sich – aus Sicht einer an einem funktionierenden technischen System interessierten Robotik – mit den dringlichsten Problemen zu beschäftigen oder – aus konversationsanalytischer Perspektive – häufig latent vorhandene Quantifizierungen auf eine nachprüfbare Basis zu stellen. Auf eine sequenzielle Beschreibung zielend, kann das Aufdecken von „Interactional Paths“ und die Quantifizierung ihrer einzelnen Schritte (ohne die Grundidee der sukzessiven lokalen Aushandlung zu negieren) Hinweise auf die Robustheit eines Phänomens liefern (vgl. Pitsch et al. 2014a). Korrelationen verschiedener interaktionsinterner und -externer Parameter können neue, über rein konversationsanalytisch Rekonstruktionen hinausgehende Perspektiven eröffnen (vgl. Stivers 2015).

wertbare Datenformate zu ergänzen, wie z.B. Motion-Capture-Verfahren oder Eye-Tracking (vgl. Pitsch et al. 2014). Damit gewinnt die Konversationsanalyse nicht nur neue methodologische Impulse und den Zugriff auf bisher kaum verfügbare und integrierbare Datenformate, sondern sie erhält auch die Möglichkeit, ihre bisherigen Erkenntnisse in einen semi-experimentellen Forschungsprozess zu integrieren. Durch die Implementierung von interaktiven Verfahren als kleine, lokale ‚Building Blocks‘ kann z.B. das genaue Zusammenspiel verschiedener kommunikativer Ressourcen unter definierten Bedingungen erforscht und/oder die Robustheit aus menschlicher Interaktion bekannter Verfahren für den Kontext der Mensch-Maschine-Konstellation überprüft werden. Die in einem solchen Prozess erarbeiteten Ergebnisse können wiederum – im Sinne eines iterativen Forschungsdesigns – als Ausgangspunkt für neue Fragestellungen, veränderte Modellierungen oder erweiterte Studiendesigns sowohl in menschlicher Kommunikation als auch in sozio-technischen Konstellationen dienen.

Diese Herangehensweise wurde in Auseinandersetzung mit drei verwandten, sich in ihren Details unterscheidenden Ansätzen entwickelt, die jeweils einzelne Aspekte eines solchen Ansatzes beleuchten.

Zum einen schlagen Yamazaki et al. (2007, 2010) und Kuno et al. (2007) – soweit ich sehe – als erste einen interdisziplinären, inkrementellen Ansatz zur Mensch-Roboter-Interaktion vor, in dem (a) menschliche Interaktion mit ethnographischen und konversationsanalytischen Methoden untersucht wird, (b) auf den gewonnenen Erkenntnissen basierend ein Roboter-System entwickelt wird und (c) der Austausch zwischen Mensch und Roboter untersucht sowie im Hinblick auf dessen Effizienz evaluiert wird, um anschließend die Entwicklung des Roboter-Systems zu verfeinern (Yamazaki et al. 2007: 72).

Zum zweiten entwickeln Pitsch et al. (2013, 2014a) einen Ansatz, in dem Videoaufzeichnungen aus einer semi-experimentellen Studie zum ‚Tutoring‘ in Eltern-Kind-Interaktion mit dem Ziel untersucht werden, daraus Erkenntnisse für die Entwicklung eines (in Zukunft: lernenden) Robotersystems zu gewinnen. Dabei wird in einem ersten Schritt in einer auf multimodaler Konversationsanalyse basierenden Analyse der interaktive Zusammenhang von Blickpraktiken des Kindes und manipulativen Tätigkeiten des Erwachsenen herausgearbeitet, wobei zur Visualisierung der Bewegungstrajektorien zusätzlich ein Motion-Capture-Verfahren in die Analyse einbezogen wird. Die so in explorativer Analyse gewonnenen Erkenntnisse werden in einem zweiten Schritt in ein hypothesengeleitetes Verfahren überführt, in dem mit mathematischen Methoden eine Korpusanalyse und Hypothesenprüfung durchgeführt wird. In einem dritten Schritt werden die so herausgearbeiteten interaktiven Zusammenhänge als Inspiration für die Modellierung von Roboterverhalten genutzt, dessen Umsetzung in einer Studie zur

Mensch-Roboter-Interaktion mit qualitativen (Pitsch et al. 2013) wie quantifizierenden Methoden (Vollmer et al. 2014) untersucht wird. Während in diesem Ansatz im Vergleich zum von Yamazaki et al. (2007, 2010) vorgeschlagenen Verfahren insbesondere die Möglichkeiten und Implikationen der Integration von konversationsanalytischen Methoden in andere methodologische Zugänge erprobt wird, liegt der Fokus hinsichtlich des Roboterhaltens auf der Frage, wie der Roboter möglicherweise das Verhalten des menschlichen Nutzers durch seine eigenen Handlungen beeinflussen kann. Die Frage nach den interaktiven Anschlussmöglichkeiten vonseiten des Roboter-Systems wird hier nicht explizit gestellt. Diese nächste Komplexitätsstufe des Erforschens von ‚Building Blocks‘ sequenzieller Handlungsstrukturen wird anschließend – im Rahmen der diesem Aufsatz zugrundeliegenden Projekte – betrachtet, und zwar dergestalt, dass der Roboter nicht nur den ‚first turn‘ übernimmt, sondern prinzipiell im ‚third turn‘ auch an die kommunikativen Angebote des Menschen (‚second turn‘) anschließen muss (Pitsch 2016, Pitsch et al. 2016).

Zum dritten fokussiert Fischer (2016) auf die Rolle des Roboters innerhalb einer Mensch-Roboter-Studie. Sie schlägt vor, Roboter als „interactional confederates“ zu verstehen, deren Vorteil darin bestehe, „reliable“ und „controllable“ (3) zu sein und „to behave exactly in the way programmed and identically for each participant“ (4). Damit lassen sich möglicherweise Aspekte der von Kuhlen & Brennan (2013) angeführten Schwierigkeiten von Confederates bearbeiten, allerdings erscheint diese Sicht – zumindest beim aktuellen Stand autonomer Robotersysteme – noch stark idealisiert und mit praktischen wie konzeptuellen Schwierigkeiten verbunden: Unsicherheiten in der Perzeption der Umwelt erschweren die Präzision in den Kontrollmöglichkeiten; fest vorprogrammierte Strukturen erschweren das interaktive ‚fitting‘; und Wizard-of-Oz-Ansätze (Riek 2012) sind unter Umständen nicht unmittelbar auf autonome Systeme anwendbar.

Insgesamt erweitert unser Ansatz<sup>4</sup> also die erste und zweite Forschungsperspektive an verschiedenen Stellen. Mit der Idee, interaktive Verfahren im Roboterverhalten zu implementieren und sie anschließend in einer Studie zur Mensch-Roboter-Interaktion zu erproben, verstehen wir ebenfalls einen Roboter als „interactional confederate“, wobei wir allerdings dessen Manipulierbarkeit in Teilen etwas kritischer betrachten.

---

<sup>4</sup> Eine frühe Version haben wir ab 2012 in den Projektskizzen „Interaktion & Raum“ und „Incremental Coordination in Human-Robot-Interaction“ beschrieben und auf diversen Konferenzen vorgestellt. In vorliegender Form wurde sie insbesondere auf dem Workshop „Interacting with Robots and Social Agents?“ auf der „Mensch & Computer“ (MuC) 2019 präsentiert.



## 4 Referenzielle Praktiken und Lokal-Deixis

Für die Umsetzung des Ansatzes, Mensch-Roboter-Interaktion als Instrument für die multimodale Interaktionsforschung zu verwenden, betrachten wir im Folgenden das Phänomen des Verweisens auf Objekte. Solche referenziellen Praktiken stellen eine grundlegende Aufgabe in sozialer Interaktion dar, deren linguistische, interaktionale und multimodale Dimension in den letzten Jahren für menschliche Face-to-face-Interaktion ausbuchstabiert worden ist. Für die intrapersonelle Koordinierung zwischen Sprache und Gestik gilt als grundlegendes Prinzip, dass die Geste dem „lexical affiliate“ zeitlich vorangeht (Schegloff 1984). Das Design der Zeigegesten orientiert sich am Interaktionspartner, sie werden „produced and timed with respect to the activities of the co-participants, such that they are in a position to be able to see the pointing gesture in the course of its production“ (Hindmarsh & Heath 2000). Für dieses Prinzip wird von Mondada (2012) anhand von französisch-sprachigen Daten eine systematische multimodale – Sprache, Gestik und Blick involvierende – Praktik beschrieben, in der das lokal-deiktische Pronomens „ici“ in Kombination mit einer Zeigegeste entweder zur Einführung eines neuen Referenten (bei einem aufmerksamen Interaktionspartner) oder zur Aufmerksamkeitsorientierung eines (aktuell anderweitig orientierten) Interaktionspartners dient. Diese Dimensionen zusammenführend, schlägt Stukenbrock (2015) ein Set von Parametern vor, anhand dessen sich solche lokal-deiktische Handlungen systematisch beschreiben lassen: Fokussierte Interaktion, Körper, deiktischer Ausdruck, Zeigegeste, Suchraum, Ziel, Referenten, Reziprozität der Wahrnehmung, Display von Verstehen. Da in deiktischen Praktiken die Aufmerksamkeitsfokussierung der Beteiligten eine zentrale Rolle spielt, beginnt Stukenbrock (2018) die Rolle des Blicks innerhalb dieser Verfahren mittels mobiler Eye-Tracking Brillen detaillierter zu untersuchen. Diese Studien zeigen zum einen die multimodale und interaktive Komplexität des Phänomens ‚referenzielle Praktiken‘ auf, bieten erste Hinweise auf das systematische Zusammenspiel der verschiedenen kommunikativen Ressourcen und verdeutlichen zum anderen auch gleichzeitig die Dynamik und Flexibilität von Interaktionsbeteiligten in ihrer Orientierung an den situativen Kontingenzen. Offen bleibt aber die Frage, wie sich das Zusammenspiel der multimodalen kommunikativen Ressourcen in der interaktiven Praxis weiter ausbuchstabieren lässt in Bezug auf die dynamische Formierung von ‚multimodal packages‘, die Funktionen der einzelnen Ressourcen und ihre Interpretation durch die Rezipienten. Welche Funktionen übernehmen z.B. jeweils Blick bzw. Geste im Rahmen des multimodalen Zusammenspiels?

Der Bedarf systematischer Beschreibungen zeigt sich insbesondere dann, wenn ein Modell referenzieller Praktiken für einen Roboter entwickelt bzw. implementiert werden soll. Studien im Bereich der Social Robotics zeigen zwar, dass kombinierte ‚Kopf-und-Armbewegungs‘-Strategien des Roboters erfolgreicher sind als die Verwendung einzelner Modalitäten (St. Clair et al. 2011) und orientieren sich für die Bestimmung der Haltedauer einer Zeigegeste an mittleren Werten von menschlichen Interaktionsbeteiligten in einem vergleichbaren Szenario (Huang & Mutlu 2012). Aber auf der Basis dieser Informationen lassen sich nur schwerlich referenzielle Praktiken für einen Roboter modellieren. Pitsch et al. (2014b) erproben eine individualistische Modellierung der referenziellen Praktiken für einen Museumsroboter in interaktivem Kontext und zeigen daran dessen Grenzen auf und formulieren erste Implikationen für nächste Schritte in Richtung eines interaktiven Zugangs. Insgesamt besteht also neben Kenntnissen zum konkreten Zusammenspiel der multimodalen Ressourcen auch ein Desiderat bezüglich der interaktiven Dimension, der Rolle des Interaktionspartners und der situativen Anpassungsleistungen für die systematische Modellierung von referenziellen Praktiken für ein Roboter-System.

## 5 Studien: Roboter, Design des Verhaltens, Daten

An die beschriebenen Kenntnisse und Desiderate zur Beschreibung referenzieller Praktiken anknüpfend, haben wir in einer Reihe aufeinander aufbauender Studien deiktische Praktiken in einem humanoiden Roboter (Nao) implementiert und erprobt. Der Roboter agiert in der Rolle eines Museumsguide und bietet vorbeikommenden Besuchern zunächst an, in ein „Focused Encounter“ einzusteigen, dann Informationen zu mehreren Exponaten zu geben und schließlich das Encounter wieder zu beenden. Im Rahmen der Exponaterläuterungen wurden jeweils verschiedene Situationen des Herstellens von Referenz auf ein Objekt integriert, die jeweils spezifische räumliche Konstellationen und Beteiligungsformen bieten.

### 5.1 Robotersystem

Für die Durchführung der Studien wurde eine humanoide Roboter-Plattform ausgewählt, die nicht nur im Forschungslabor, sondern insbesondere auch im „real world“-Kontext des Museums einsetzbar ist. D.h. sie soll sowohl freundlich auf die Besucher wirken als auch transportabel, robust, bezahlbar und in der Com-

munity gut verortet sein. Daher fiel die Wahl auf einen Roboter des Typs NAO (Aldebaran, version 3+), der konstant über die verschiedenen Studien hinweg eingesetzt wurde. Erst 2017 wurden für eine weitere Studie die vorhandenen Elemente auf einen Roboter des Typs Pepper übertragen. Um die Größe des NAO (52 cm) für die Interaktionsstudien zu kompensieren, wurde der Roboter im Ausstellungsraum auf einem Tisch positioniert. Nach einer ersten höhenverstellbaren Tischversion mit kleiner Grundfläche wurde eigens für den Roboter ein Tisch angefertigt, der sowohl Laufbewegungen des Roboters ermöglicht als auch – für den Fall von technischen Komplikationen – eine Auffangmöglichkeit für den Roboter integriert, um nicht Gefahr zu laufen, möglicherweise im Museum Exponate zu beschädigen.

Das Roboter-System ist so programmiert, dass es im Museum (s. Studien 1 und 2) autonom agiert. Es ist mit einer basalen Middleware-Architektur ausgestattet, die dezidierte modulare Systemkomponenten (z.B. Personen-Tracking, Spracherkennung, Dialog u.a.) verbindet. Die Handlungsabfolge des Roboters folgt einem prädefinierten Ablauf-Skript, bei dem die einzelnen Phasen durch Finite State Automaten (mit sich entwickelnder Fein-Granularität) modelliert wurden.

**Studie 1:** In der hier verwendeten ersten Version des Roboter-Systems ist die Perzeption ausschließlich visuell (ohne Spracherkennung) mittels eines externen infrarotbasierten Motion-Capture-Systems (Vicon) realisiert. Hierfür wurden Infrarot-Kameras unter der Decke des Museumsraums installiert, mittels derer die Position im Raum, Bewegung und Ausrichtung der Besucher detektiert, auf einem externen PC verarbeitet und als Personen-Informationen an die Software des Roboter-Systems übergeben. Damit die Besucher vom Motion-Capture-System detektiert werden können, wurden sie vor dem Betreten des Raums mit (bis zu maximal fünf) Mützen ausgestattet, auf denen jeweils mehrere Vicon-Marker in einem spezifischen Muster angeordnet waren. Dieses Verfahren wurde verwendet, um zum einen die Bewegung der Besucher im Raum als Teil des Daten-Korpus mit aufzeichnen zu können und zum anderen als initiale ‚Abkürzung‘ für künftige, mit den Bordmitteln des Roboters zu realisierende Perzeptionskomponenten. Gleichzeitig hat die Integration des Vicon-Systems zu zwei unbeabsichtigten, aber für die Interaktionsforschung fruchtbaren Konstellationen geführt: (a) Zeitverzögerung in der Verarbeitung der Daten, was zu unbeabsichtigt langen Pausen zwischen verschiedenen Roboter-Aktivitäten führt; (b) Lokalisierung der Personen mit einer leichten seitlichen Abweichung, was zu spezifischen Dynamiken in der Beteiligungsstruktur der Besucher führt.

**Studie 2:** In dieser Version des Roboter-Systems wird für die visuelle Perzeption die roboter-interne VGA-Kamera des NAO verwendet. Zusätzlich wurde ein

externes Mikrofon unterhalb des Roboter-Tischs platziert, dessen Daten für die Spracherkennung per Keyword-Spotting genutzt wird (vgl. Pitsch et al. 2016).

5.2 Design der referenziellen Praktiken

Das Design der im Roboter implementierten referenziellen Praktiken orientiert sich an interaktionsanalytischen Erkenntnissen (s. Kap. 3) und Konzepten interaktiver Koordinierung. Insbesondere sind hierbei zentral das Konzept der „multimodal packages“ (Goodwin, Mondada), die Beobachtung zur intrapersonellen Koordinierung derzufolge „the gesture precedes the lexical affiliate“ (Schegloff 1984) sowie die Orientierung am Interaktionspartner (Hindmarsh & Heath 2000). Gleichzeitig müssen die – zum Zeitpunkt der jeweiligen Studie aktuellen – Möglichkeiten der technischen Realisierung berücksichtigt und in das Design der referenziellen Praktiken einbezogen werden. Dieses führt einerseits zu Limitierungen im Design, befördert andererseits aber die Beobachtung spezifischer interaktiver Dynamiken.

**Studie 1:** Im Rahmen der vorprogrammierten Museumsführung wurden drei verschiedene räumlich-situative Konstellationen von deiktischer Referenz auf ein Exponat integriert (s. Pitsch et al. 2014). In der im Folgenden näher betrachteten Konstellation verweist der Roboter auf ein Exponat, das sich auf der gegenüberliegenden Seite des Raumes befindet (s. Abb. 2, P6). Durch die vorangehenden Erklär-Aktivitäten des Roboters werden zudem Bedingungen hergestellt, unter denen die Aufmerksamkeit der Besucher in etwa in Richtung des Roboters ausgerichtet sein dürfte.

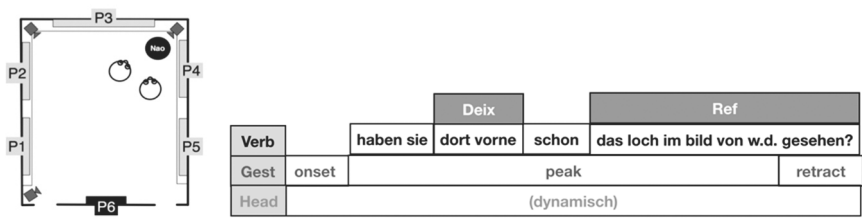


Abbildung 2: Setting (1)    Abbildung 3: Beispiel „Walther Dahn“

Die deiktische Referenz des Roboters wurde als „multimodal package“ mit den folgenden Eigenschaften definiert (s. Abb. 3):

**(a) Intrapersonelle Koordinierung von Sprache und Gestik des Roboters:**

Zu Beginn der Äußerung wird ein lokaldeiktischer Ausdruck („dort vorne“) mit einer Zeigegeste kombiniert, deren Onset leicht dem verbalen Ausdruck vorangeht (s. Schegloff 1984). Hierauf folgt die Benennung des Referenten („das loch im bild von walther dahn“) währenddessen der Peak der Zeigegeste gehalten wird. Mit Ende der verbalen Äußerung wird auch die Geste in ihre Home-Position zurückgeführt.

**(b) Kopforientierung des Roboters richtet sich dynamisch an der räumlichen Position der Besucher aus:**

Die technischen Möglichkeiten des Motion-Capture-Systems nutzend, ist der Roboter so programmiert, dass er seinen Kopf jeweils dynamisch auf die sich verändernde Position des dem Roboter am nächsten stehenden Besuchers ausrichtet. Aus technischen Gründen musste dieses als ein permanent mitlaufendes Verhalten realisiert werden, so dass sich die Steuerung der Kopforientierung nicht in die Gestaltung der ‚multimodal packages‘ in Situationen lokal-deiktischer Referenz integriert werden konnte. Diese getrennte Verarbeitung von Sprache/Gestik und Kopforientierung führt zu einer experimentellen Situation, in der Blick und Geste des Roboters – je nach der von den Besuchern eingenommenen Position im Raum – während der Deixis voneinander dissoziiert sein können.

**Studie 2:** Analog zur ersten Studie wurden auch bei Studie 2 drei verschiedene räumlich-situative Konstellationen von deiktischer Referenz auf ein Exponat in die Erläuterungen des Roboters integriert (s. Pitsch et al. 2016). In der im Folgenden näher betrachteten Konstellation verweist der Roboter auf ein Exponat zu seiner rechten Seite (s. Abb. 4, EX-1). Auch hier wurden durch die vorangehende Interaktion Bedingungen hergestellt, unter denen die Aufmerksamkeit der Besucher initial primär auf den Roboter ausgerichtet ist.

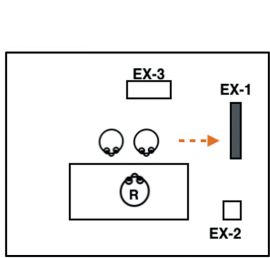


Abbildung 4: Setting (2)

Deix		Ref	
Verb	dort	drüben könnt ihr sehen	wer damals auf der sparrenburg gewohnt hat
Gest		onset	peak retract
Head	@Besucher		

Deix		Deix-Präzisierung	
Verb		dort drüben	auf dem großen bild
Gest	onset		peak retract
Head	>>>		@EX-1 <<<

Abbildung 5: Beispiel „Auf der Sparrenburg“

Im Vergleich zu Studie 1 bestehen die zentralen Unterschiede im Design der deiktischen Referenz des Roboters in den folgenden Aspekten:

**(a) Kopforientierung des Roboters als Teil des ‚multimodal package‘:**

Durch die veränderten technischen Bedingungen kann nun die Kopfausrichtung des Roboters als integrierter Bestandteil in der Gestaltung der ‚referenziellen Praktik‘ mitmodelliert werden.

**(b) Monitoring und Repair:** Entsprechend des zugrundeliegenden Ansatzes der interaktiven Koordinierung ist das Verhalten des Roboters so gestaltet, dass er das Verhalten der Besucher beobachtet, dieses im Hinblick auf Erfolg bzw. Misserfolg seiner deiktischen Referenz bewertet und ggf. einen zweiten Verweis auf das Exponat (d.h. ein Repair) anbietet.

**(c) Design der referenziellen Praktiken – Upgrade in der zweiten Version:** Beim ersten deiktischen Verweis zeigt der Roboter per Armgeste auf das Exponat, während der Kopf auf die Besucher ausgerichtet ist (analog zu Studie 1). Diese Dissoziation der Ressourcen ist erforderlich, damit der Roboter – dessen Kamera frontal im Roboterkopf eingebaut ist – die Reaktionen der Besucher beobachten kann. Der (wenn notwendig) zweite Verweis auf das Exponat, d.h. eine Reparatur, ist als Upgrade gestaltet: Der Roboter zeigt per Armgeste auf das Exponat und wendet ebenfalls parallel dazu seinen Kopf dorthin (s. Abb. 5).

### 5.3 Studien & Daten: Roboter als Museumsguide

Im vorliegenden Text werden Daten aus zwei (von fünf sukzessive aufeinander aufbauenden) Studien verwendet.

**Studie 1:** Studie 1 wurde im Jahr 2010 in einem Kunstmuseum durchgeführt (Pitsch et al. 2014b). Als Teilnehmer der Studie wurden reguläre Museumsbesucher rekrutiert, die in einem Vorraum des Museums vom Forscherteam angesprochen wurden. Sie wurden darüber informiert, dass im Nachbarraum eine wissenschaftliche Studie stattfindet, sie – beim Betreten des Raumes – mittels Videokameras und eines Motion-Capture-Systems aufgezeichnet würden und gebeten einen Hut mit Vicon-Markern für diese Zeit aufzusetzen. Weitere Informationen zu Art und Inhalt der Studie wurden ihnen nach der Teilnahme erläutert sowie ihre schriftliche Einwilligung eingeholt (andernfalls wurden die Aufnahmen gelöscht). Dementsprechend bestehen die Studienteilnehmer vornehmlich aus Kleingruppen von Erwachsenen. Die Begegnungen zwischen Roboter und Besuchern dauerten ca. 3 Minuten, wobei verschiedene Durchläufe aufgezeichnet wurden. Diese wurden mittels zwei bzw. drei Videokameras und Motion-Capture-Verfahren (Vicon) aufgezeichnet.

**Studie 2:** Studie 2 fand im Jahr 2014 in einem Historischen Museum (Pitsch 2016, Pitsch et al. 2016) im Rahmen eines Wissenschaftsfestivals statt, in dessen Programmheft der Museumsroboter im Historischen Museum angekündigt wur-

de. Die Teilnehmer der Studie sind dementsprechend vornehmlich Kleingruppen bestehend aus Kindern und Eltern, die aufgrund eines spezifischen Interesses an Technik/Robotern ins Museum gekommen sind. Auch hier wurden die Besucher im Vorraum des Museums vom Forscherteam begrüßt, sie über das Stattfinden einer wissenschaftlichen Studie informiert und ihre schriftliche Einwilligung eingeholt. Erläuterungen zu Art und Zielen der Studie wurden ihnen im Anschluss an die Studienteilnahme angeboten, wofür Mitglieder des Forschungsteams bereitstanden. Die Begegnungen zwischen Roboter und Besuchern dauerten ca. 4 Minuten, wobei 72 verschiedene Durchläufe mittels vier Videokameras und zwei externen (nicht mit dem Roboter-System verbundenen) Motion-Capture-Kameras (Kinect) aufgezeichnet wurden.

## 6 Umgang mit deiktischen Angeboten des Roboters

In einem ersten Schritt (Studie 1) wird anhand von zwei Fallanalysen der Frage nachgegangen, wie die Museumsbesucher mit den kommunikativen Angeboten des Roboters spontan in einem alltagsweltlichen Setting umgehen. Wie behandeln sie in einer solchen sozio-technischen Konstellation die referenziellen Verweise des Roboters? Welche Hinweise lassen sich auf das Zusammenspiel der kommunikativen Ressourcen – unter den Bedingungen dieser sozio-technischen Konstellation – erkennen? Welche Implikationen lassen sich aus diesen Beobachtungen für die weitere Entwicklung der Interaktionskompetenzen für das technische System ableiten?

### 6.1 Parallelität von Kopfausrichtung und Zeigegeste – Sukzessive Lokalisierung (Fallanalyse 1)

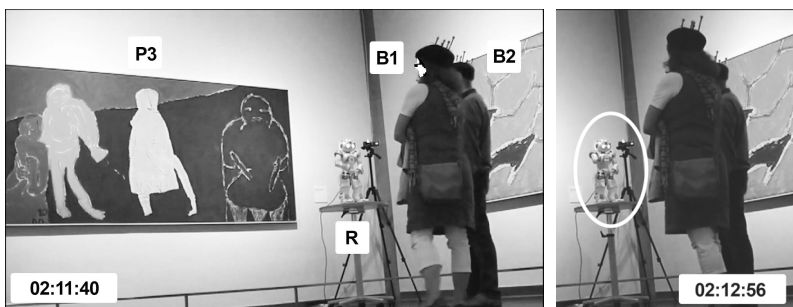
Untersuchen wir ein erstes Beispiel (Studie 1, VP 222), in dem zwei Besucher die referenziellen Orientierungsangebote des Roboters unmittelbar befolgen und sich erfolgreich auf das Exponat hin orientieren. Dabei folgen sie den Angeboten des Roboters kleinschrittig und zeigen sich wechselseitig ihre jeweils aktuellen Interpretationen der Roboter-Aktivitäten auf.

**(a) Deiktische Referenz – Parallelität von Kopfausrichtung und Zeigegeste (R):** Zu Beginn des hier interessierenden Ausschnitts sind eine Besucherin (B1) und ein Besucher (B2) körperlich in Richtung des Roboters bzw. des Exponats P3 ausgerichtet (#02:11:40). Der Kopf des Roboters ist in ihre Richtung wie

auch auf ein einige Meter hinter ihnen an der Wand hängendes Bild (P6, Walther Dahn) ausgerichtet. Der Roboter äußert die Frage „haben sie dort VORne schon das LOCH im bild von walther DAHN gesehen,“, die durch eine Zeigegeste in Richtung der Besucher und des Exponats P6 begleitet wird (Z. 01–02, #02:12:56).

01 R-ver: |haben sie dort VORne schon das LOCH  
 R-gaz: |@P6/B1/B2  
 R-ges: |pt@P6-onset |pt@P6-peak  
 |#02:11:40 |#02:12:56

02 R-ver: im bild von walther DAHN gesehen, |  
 R-ges: .. |



**(b) Behandlung als Frage-Antwort-Sequenz (B2):** B2 beantwortet die Frage des Roboters mit „NEIN;“ (03) und etabliert damit auf gesprächsstruktureller Ebene eine Frage-Antwort-Sequenz.

03 R-ver: (1.0) | (1.0)  
 R-ges: .. |  
 B2-ver: |NEIN; |

**(c) Sukzessive Lokalisierung des Referenten (B1):** B1 hingegen behandelt die Äußerung des Roboters als Aufforderung, sich auf die bezeichnete Referenz hin zu orientieren. Sie dreht sich um 180° in die durch Geste und Kopf angezeigte Richtung zu P6 (#02:17:32), wendet ihren Kopf kurz zu P1 (#02:17:68), wieder zurück zu P6 (#02:18:28) und zeigt schließlich auf P6 (#02:18:72). Damit wird ihr Versuch sichtbar, die multimodale Referenz des Roboters aufzulösen, d.h. sukzessive das Zeigeziel in ihrer ‚domain of scrutiny‘ zu verorten und inhaltlich zu interpretieren. Dieser Vorgang wird gerahmt als soziales Ereignis, indem die In-



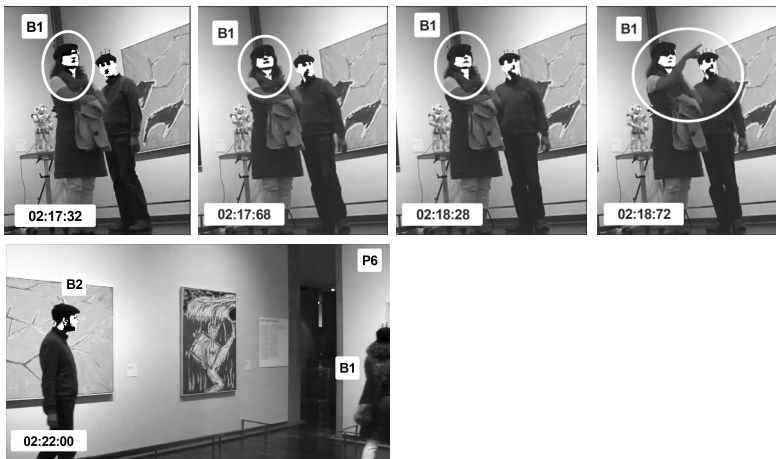
terpretationsschnitte für B2 sichtbar durchgeführt werden und schließlich B1 und B2 gemeinsam zum Exponat (P6) schreiten (#02:22:00).

```

03  R-ver:  (1.0)          | (1.0)          |
     R-ges:  ..            |                |
     B2:      | NEIN;    |                |
     B2-act:  | (turns round) |                |
     B2-gaz:  |                | @P6          |
     B1-act:  (turns round) |                |
     B1-gaz:  |                | @6P          |
           | #02:17:32 | #02:17:68 | #02:18:28

04  R-ver:  |auch das ist ein |stilmittel |der mülheimer freiheit
     R-ges:  |pt@P6-retr      |                |
     B1-act:  | (pt>P6)      |                |
           | #02:18:72          |                | #02:22:00

```



Insgesamt deckt dieses Beispiel nicht nur auf, dass die Besucher auf die referenziellen Angebote des Roboters reagieren, sondern auch wie sie schrittweise den Orientierungshinweisen folgen, aktiv und als soziale Praktik gestaltet daran arbeiten, den designierten Referenten zu etablieren.

## 6.2 Dissoziation von Kopfausrichtung und Geste – Rückorientierung zum Roboter (Fallanalyse 2)

Während im ersten Beispiel relativ einfache strukturelle Bedingungen gegeben sind – Parallelität von Kopfausrichtung und Geste des Roboters; räumliche Konstellation, in der beide Besucher und das designierte Exponat in einer Fluchtlinie stehen – ermöglicht ein zweites, komplexeres Beispiel (Studie 1, VP 043) eine weitere Ausdifferenzierung der referenziellen Praktiken, des Zusammenspiels der kommunikativen Ressourcen sowie ihre interaktiven Implikationen. Der zweite Fall bietet eine leicht veränderte Konstellation: Kopfausrichtung und Geste des Roboters weichen leicht voneinander ab; und zwei Besucherinnen (B1, B2) stehen mit ca. 1 m Abstand zueinander, d.h. in der Fluchtlinie zu unterschiedlichen Exponaten.

**(a) Aufmerksamkeitsorientierung und Relevantsetzung der Ressource ‚Blick‘ (R):** Wenn wir in die Situation einsteigen, sind die beiden Besucherinnen teils auf den Roboter, teils auf die Exponate P3 (B1) und P2 (B2) orientiert. Der Kopf des Roboters (R) ist in Richtung B2 bzw. des Exponats P1 (#02:38:00) orientiert, dann wechselt die Kopforientierung kurz in Richtung B1 (#02:39:80) und – parallel zum Beginn des verbal-gestischen Verweises auf P6 (Z. 01) – zurück zu B2 bzw. P1 (#02:41:76). Auf technischer Ebene ist diese wechselnde Kopforientierung des Roboters dadurch motiviert, dass sich das System jeweils dynamisch auf die (aus technischer Sicht) am nächsten zum Roboter stehende Person ausrichtet, wobei die Orientierung systematisch nach rechts abweicht (s.o.). Auf Interaktionsebene setzt diese wechselnde Kopforientierung – von Exponat P1 zu den Besucherinnen und zurück zu P1 – beim Beginn des referenziellen Verweises die Ressource ‚Blick‘ relevant. Mindestens für B1 ist dieses aufgrund ihrer körperlichen Ausrichtung wahrnehmbar.

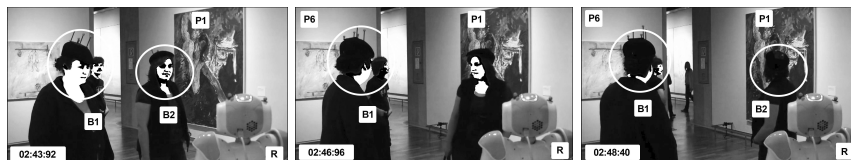
```
01 R-ver:                |haben sie dort VORne |schon |das LOCH
R-gaz:   @P1 |@V2        |@P1.....
R-ges:    |pt@P6-onset   |pt@P6-peak .....
V1-gaz: |@P3                                |@R
V2-gaz: |@P2                                |>>>>>|@R
          |02:38:00          |02:41:76          |02:43:92
          |02:39:80
```



### (b) Deiktische Referenz – Dissoziation von Kopfausrichtung und Geste (R):

Der Verweis auf das Exponat „haben sie dort VORne schon das LOCH im bild von walther DAHN gesehen,“ (Z. 01-02) wird von B1 und B2 sukzessive mitvollzogen. Im Anschluss an das verbale und per Zeigegeste begleitete Orientierungsangebot „dort VORne“ wenden B1 und B2 ihre visuelle Aufmerksamkeit auf den Roboter (#00:43:92). Nach der inhaltlichen Präzisierung („LOCH im bild von walther DAHN“) versuchen sie sukzessive den Referenten zu lokalisieren: B2 dreht sich im Uhrzeigersinn zunächst zu P6 (d.h. dem Bild von Walther Dahn an der gegenüberliegenden Raumseite) und dann weiter zu P1, auf dem ihr Blick schließlich ruhen bleibt. B1 dreht sich gegen den Uhrzeigersinn zunächst zu P1 (#02:46:96), dann weiter in Richtung P6 (02:48:40). Diese sukzessiven Re-Orientierungen zwischen P1 und P6 als mögliche Referenten spiegeln den Umstand wider, dass die Zeigegeste des Roboters auf P6 verweist (Bild von Walter Dahn), wohingegen dessen Kopfausrichtung – entsprechend der technisch bedingten leicht versetzten Orientierung auf B2 – in Richtung auf P1 geht. Hier werden also die kommunikativen Ressourcen Kopforientierung/Blick und (Zeige-)Geste in der Verkörperung des Roboters voneinander dissoziiert, wobei sie auf unterschiedliche Objekte (P6 vs. P1) im Ausstellungsraum verweisen.

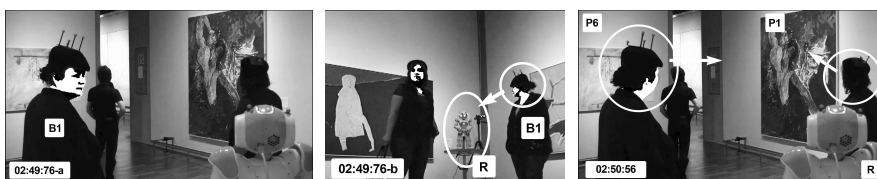
```
02 R-ver:  im bild von walther DAHN gesehen, | (1.0) | (1.4) |
R-gaz:  .....
R-ges:  ..... |pt@P6-retr |
V1-gaz:  >> |@P1 |@P6 |>> @P1 >>
V2-gaz:  >> |@P6|@P1
          |#02:46:96
          |#02:48:40
```



**(c) Umgang mit Ambivalenz – Fehlorientierung (B2) vs. Re-Orientierung zum Roboter (B1):** Diese Ambivalenz im Verweisraum des Roboters wird von den Besucherinnen unterschiedlich bearbeitet. B2s Blick bleibt (fälschlicherweise) auf P1 haften (02:49:76-a). B1 hingegen dreht sich zurück zum Roboter (02:49:76-a), wo weitere Disambiguierungshinweise prinzipiell erwartet werden könnten. Allerdings hat der Roboter zu diesem Zeitpunkt seine (korrekt auf P1 ausgerichtete) deiktische Geste bereits wieder zurück in die Home Position (Sacks) geführt (02:49:76-b), so dass sie als (korrekter) Orientierungshinweis auf P6 nicht mehr für B2 zur Verfügung steht. Zur Orientierung bleibt also nur noch die Kopfausrichtung des Roboters (02:49:76-b), der B2 konsequenterweise folgt und entsprechend – wie auch B1 – fälschlicherweise das Exponat P1 betrachtet (02:50:56).

```
03 R-ver: |auch das ist ein stil|mittel der MÜHLheimer freiheit |
R-gaz: ..... |
V1-gaz: |@R..... >>|@P1
V2-gaz: |..... |
        |#02:49:76-a/b |#02:50:56
```

```
04 R-gaz: |@P6/V2
```



Insgesamt zeigt dieses Beispiel also, dass Besucher im Fall der Dissoziation von Kopforientierung und Geste Schwierigkeiten haben, den korrekten Referenten zu identifizieren. Gleichzeitig wird bezüglich der zeitlichen Struktur erkennbar, dass eine fixe, vorprogrammierte Dauer einer Zeigegeste kaum den Anforderungen von Situiertheit genügen kann. Dieses führt gleichzeitig zur Beobachtung eines Alltagsverfahrens: Im Fall des Zweifels re-orientiert sich die Besucherin zum Roboter.

### 6.3 Erste Beobachtungen und Implikationen

Aus diesen Beobachtungen lassen die folgenden Ergebnisse festhalten und erste Implikationen für das Design referenzieller Praktiken ableiten:

**(a) Interaktionale Relevanz von Orientierungshinweise des Roboters:**

Interaktionsbeteiligte reagieren auf die Orientierungshinweise des Roboters in seiner Rolle als Museumsführer.

**(b) Dissoziation der Ressourcen Blick und Geste:** Es finden sich erste Hinweise darauf, dass sich Interaktionsbeteiligte an einzelnen Aspekten bzw. Modalitäten einer ‚kommunikativen Gestalt‘ orientieren. Dieses ist korpusbasiert weiter zu analysieren.

**(c) Praktik der Besucher:** Im Fall der Ambivalenz und Unsicherheit, auf welches Exponat vom Roboter verwiesen wird, findet sich eine Nutzer-Praktik ‚Re-Orientierung zum Roboter‘, von dem weitere Disambiguierungshinweise prinzipiell erwartet werden könnten.

**(d) Dauer einer Zeigegeste:** Da nicht in allen Details vorhersehbar ist, wann genau Nutzer auf den Roboter hin orientiert sind und folglich seine visuellen Orientierungshinweise wahrnehmen können, ist die erforderliche Haltedauer einer Zeigegeste schwer im Vorfeld definierbar. Sie allein in Relation zur verbalsprachlichen Ebene der Deixis zu designen, greift zu kurz, vielmehr ist ein interaktiver Ansatz erforderlich.

**(e) Implikationen für das Design eines Building Blocks ‚referenzielle Praktik‘ – Explizitheit vs. Interaktivität:** Aus diesen Beobachtungen ergeben sich als Implikation für das Design eines Building Blocks ‚referenziellen Praktiken‘ für einen humanoiden Roboter zunächst zwei prinzipielle Varianten. Als **Variante (1)** könnte eine Möglichkeit für das Design von referenziellen Praktiken in einer verbesserten **Explizitheit** des initialen Orientierungsangebots liegen. Dieses hätte einerseits den Vorteil, dass der Roboter anschließend weniger mit möglichen auftretenden Schwierigkeiten umgehen müsste. Andererseits würde ein solches Verfahren wenig Übertragbarkeit und Flexibilität bieten. Demgegenüber besteht **Variante (2)** im Ausbau der **interaktiven, auf Monitoring basierenden** Komponente. Das technische System müsste im Anschluss bzw. während des Ausführens der roboterseitigen deiktischen Referenz die Auswirkungen der eigenen Handlungen mit beobachten und darauf dynamisch reagieren. Konkret müsste hier ein solches Verfahren bestehen in: (i) dem Angebot einer lokal-deiktischen Referenz, (ii) der Beobachtung, ob die Nutzer seinem Angebot folgen oder nicht, (iii) bei Bedarf dem Angebot eines erneuten Orientierungshinweises bzw. einer Reparatur. Ein so konzipiertes, auf Interaktivität und Monitoring basierendes Verfahren erscheint interaktional interessant und wurde in einer Folgestudie (s. Abschnitt 7) umgesetzt und erprobt.

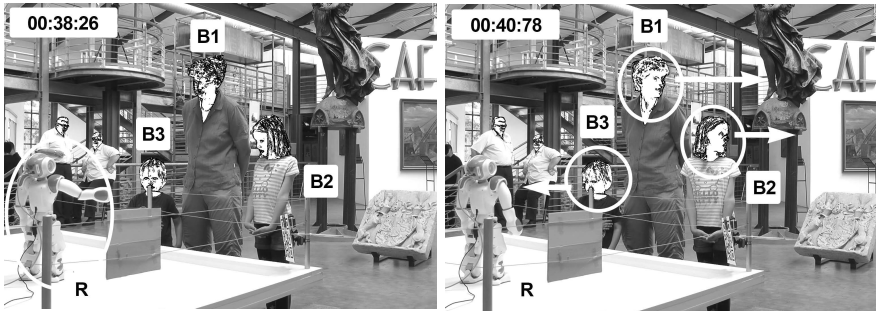
## 7 Referenzielle Praktik als sequenzstruktureller ‚Building Block‘?

Die anhand der ersten Studie entwickelten Implikationen wurden in die nächste Iteration für das Design der referenziellen Praktiken des Roboters übernommen und im Rahmen einer Folgestudie (Studie 2) erprobt. Der Fokus liegt hier insbesondere auf dem Ansatz eines lokalen ‚Building Blocks‘, der auf den Ideen von Monitoring und Interaktivität basiert: (i) Angebot einer lokal-deiktischen Referenz, (ii) Beobachtung, ob die Nutzer seinem Angebot folgen oder nicht, (iii) bei Bedarf Angebot eines erneuten Orientierungshinweises bzw. einer Reparatur. Hiermit wird ein erster Versuch konzipiert und implementiert, referenzielle Praktiken für die Mensch-Roboter-Interaktion als sequenzielle Strukturen zu verstehen. Anhand einer Fallanalyse (Studie 2, 4\_007) wird der Frage nachgegangen, inwiefern sich ein solcher Ansatz als umsetzbar erweist. Hierzu dient die folgende Analyse:

**(a) Deiktische Referenz (R):** In unserem Beispiel sind drei Besucher – eine Frau mit zwei Kindern – vis-à-vis dem Roboter positioniert und auf diesen orientiert (#00.38.26). Der Roboter lädt – per Sprache und Geste und mit Kopfausrichtung auf die Besucher – ein, das Exponat zu seiner rechten Seite (EX-1) zu betrachten: „dort drüben könnt ihr sehen wer damals auf der sparrenburg gewohnt hat“ (Z. 01-02, #00.38.26). B1 und B2 folgen diesem Angebot sofort, B3 hingegen bleibt auf den Roboter fokussiert (#00.40.78).

```
01 R-ver: |dort |drüben könnt ihr sehen |wer      |
    R-ges:          |d-onset                |d-peak |
    R-hea: |@Besucher ...

                                         |#00:38:26
02 R-ver: |damals auf der sparrenburg gewohnt |hat|
    R-ges: |d-retract                        |
    R-hea: ...                               |#00:40:78
```



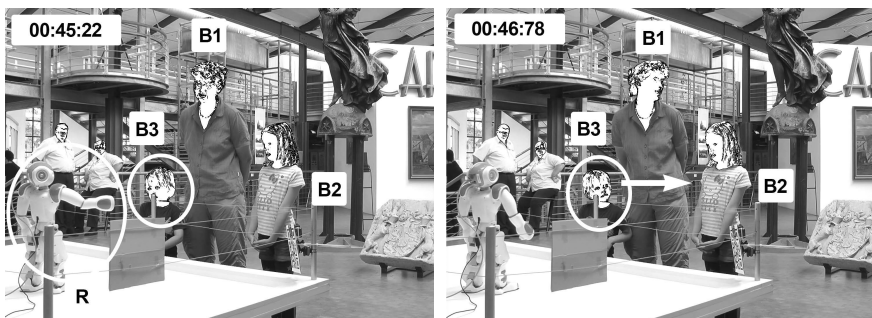
**(b) Monitoring von Erfolg/Misserfolg (R):** Entsprechend des implementierten Prinzips, dass der Roboter den Erfolg/Misserfolg seiner Handlungen beobachtet stellt sich die obige Situation aus Sicht des Roboters wie folgt dar: Das technische System kann zumindest eine Person – die hier besonders interessante B3 (die der deiktischen Referenz *nicht* folgt) – durchgängig stabil detektieren (dieses wird in der Videoaufzeichnung dadurch angezeigt, dass eine ‚bounding box‘ um den Kopf von B3 inkl. Kategorisierung „Nao“ erscheint (am unteren Bildrand unter der ‚bounding box‘, leicht abgeschnitten)) und damit beobachten, dass sie auch nach dem deiktischen Verweis auf das Exponat weiterhin auf den Roboter orientiert bleibt (#00.38.26; 00:40:78).



**(c) Zweiter deiktischer Verweis als Reparaturangebot:** Der implementierten Handlungslogik folgend entscheidet sich das technische System für die Relevanz eines Reparaturangebots und bietet einen zweiten deiktischen Verweis an: „dort drüben auf dem großen bild“ (Z. 03-04, #00:45:22), wobei hier nun auch die Kopforientierung des Roboters auf das Exponat verweist. Auf diesen zweiten Orientierungshinweis reagiert B3, indem er nun auch den Kopf in Richtung des Exponats wendet (#00:46:78).

```
03 R-ver: |(0.5) |dort drüben auf dem |großen  
R-ges: |d-on |d-peak  
R-hea:      |>>>>>>>>>>>>>>>>>>|EX-1  
                                           |#00:45:22
```

```
04 R-ver: bild | (0.7) |
    R-ges:      | d-retract
    R-hea:      | <<<<<<<< |
                                     | #00:46:78
```



Insgesamt lässt sich also festhalten, dass sich der gewählte Ansatz als grundsätzlich funktional erweist. Damit ist ein erster Schritt in Richtung des Erprobens von kleinen ‚Building Blocks‘ für sequenzielle Handlungsstrukturen erfolgt, deren verschiedene Facetten und Details künftig weiter zu untersuchen sind. Erste Hinweise finden sich auch auf Herausforderungen in Bezug auf Mehrpersonen-Konstellationen, Unsicherheit und Entscheidungen bei weniger eindeutigen Situationen (s. z.B. beginnend in Pitsch et al 2016, Pitsch 2016).

## 8 Zusammenfassung und Diskussion

Im Verlauf des Texts haben wir Mensch-Roboter-Interaktion als ein Forschungsinstrument für die Interaktionale Linguistik eingeführt und dieses anhand der Frage ausbuchstabiert, wie referenzielle Praktiken für einen Roboter designt werden können, der in engem Kontakt mit Menschen agiert und für diese intuitiv interpretierbar sein soll. Diese Frage speist sich zum einen aus einer multimodal orientierten Interaktionslinguistik, für die eine Aufgabe darin besteht, das komplexe multimodale Zusammenspiel verschiedener kommunikativer Ressourcen zu „multimodal packages“ unter den Bedingungen von Situiertheit, Interaktivität



und Kontingenz systematisch zu beschreiben. Zum anderen trifft die Forschungsfrage auf aktuelle Herausforderungen in den Social Robotics, die darauf zielen, Robotersysteme so mit Interaktionsfähigkeiten auszustatten, dass sie von menschlichen Nutzern in Alltagskontexten mit den Mitteln natürlichsprachlicher Interaktion intuitiv bedienbar sind.

Dieser Frage sind wir empirisch anhand von zwei aufeinander aufbauenden Studien nachgegangen, in denen ein humanoider Roboter in der Rolle eines Museumsführers agiert und vorbeikommenden Besuchern verschiedene Exponate erklärt. Dabei konnten die folgenden Ergebnisse herausgearbeitet werden:

**(a) Dissoziation der Ressourcen Blick und Geste:** Es finden sich erste Hinweise darauf, dass sich Interaktionsbeteiligte an einzelnen Aspekten bzw. Modalitäten einer ‚kommunikativen Gestalt‘ orientieren. Dieses ist korpusbasiert weiter zu analysieren.

**(b) Praktik der Re-Orientierung im Fall von Unsicherheiten:** Im Fall der Ambivalenz und Unsicherheit, auf welches Exponat vom Roboter verwiesen wird, findet sich eine Nutzer-Praktik ‚Re-Orientierung zum Roboter‘, von dem weitere Disambiguierungshinweise prinzipiell erwartet werden könnten.

**(c) Dauer der Zeigegeste als interaktive Herstellungsleistung:** Da nicht in allen Details vorhersehbar ist, wann genau Nutzer die visuellen Orientierungshinweise des Roboters wahrnehmen können, ist die erforderliche Haltedauer einer Zeigegeste schwer im Vorfeld definierbar. Vielmehr ist ein interaktiver Ansatz erforderlich.

**(d) Implikationen für das Design referenzieller Praktiken – Monitoring und Interaktivität:** Als Implikation für das Design eines ‚Building Blocks‘ ‚referenzielle Praktiken‘ für einen humanoiden Roboter wurde ein auf Monitoring und Interaktivität basierendes Verfahren beschrieben, das in einer Folgestudie implementiert und erprobt wurde: Das technische System soll im Anschluss bzw. während des Ausführens der roboterseitigen deiktischen Referenz die Auswirkungen der eigenen Handlungen mit beobachten und darauf dynamisch reagieren. Konkret wurde das Folgende implementiert: (i) Angebot einer lokal-deiktischen Referenz, (ii) Beobachtung, ob die Nutzer seinem Angebot folgen oder nicht, (iii) bei Bedarf Angebot eines erneuten Orientierungshinweises bzw. einer Reparatur. Eine erste Auswertung der Studiendaten zeigt, dass sich dieser Ansatz grundsätzlich als funktional erweist und so ein erstes Angebot für das Erproben sequenzieller Handlungsstrukturen vorliegt. Verschiedene Facetten und Details sind künftig weiter zu untersuchen, wie z.B. die Herausforderungen einer Mehrpersonen-Konstellation, Unsicherheit und Entscheidungen bei weniger eindeutigen Situationen (s. z.B. Pitsch 2016).

Zusammenfassend sollten diese empirischen Beobachtungen am Beispiel der Lokal-Deixis als ein Proof of Concept den Mehrwert von semi-experimentellen Studiendesigns und der Mensch-Roboter-Interaktion als einem Instrument zur Erforschung der Situiertheit von multimodaler Kommunikation erkennbar werden lassen: Die Modellierung – und damit höchst systematisch-präzise Beschreibung – von interaktionalen ‚Building Blocks‘ und die Überprüfung der „Risiken und Nebenwirkungen“ von Interaktionsverfahren in konkreten Interaktionsstudien bringen zum einen die Grenzen vorliegender Beschreibungen und Ergebnisse ans Licht. Zum anderen ermöglichen solche Studien im Sinne Garfinkelscher Krisen-Experimente durch das (un/beabsichtige) Aufbrechen gewohnter Alltagsstrukturen einen vertieften Einblick in die Funktionsweise interaktiver Verfahren und Zusammenhänge.

## Literatur

- Arminen, I./Licoppe, C./Spagnolli, A. (2016): Respecifying Mediated Interaction. In: *Research on Language and Social Interaction* 49 (4), 290–309.
- Breazeal, C. (2003). Toward sociable robots. In: *Robotics and Autonomous Systems* 42 (3-4), 167–175.
- Breazeal, C./Dautenhahn, K./Kanda, T. (2016): Social Robotics. In: Siciliano, Bruno/Khatib, Oussama (Hg.): Springer Handbook of Robotics. 2. Auflage. Berlin, Heidelberg: Springer, 1935–1971.
- Couper-Kuhlen, E. (2018): Finding a Place for Body Movement in Grammar. In: *Research on Language and Social Interaction* 51 (1), 22–25.
- Couper-Kuhlen, E./Selting, M. (2018): Interactional Linguistics. Studying Language in Social Interaction. Cambridge: Cambridge University Press.
- Dautenhahn, K. (2014): Human-Robot-Interaction. In: Dam, R. F./Soegard, M. (Hg.): The Encyclopedia of Human-Computer-Interaction. 2. Auflage. Aarhus, Denmark: The Interaction Design Foundation. URL: <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-robot-interaction>
- Dourish, P./Button, G. (1998): On „Technomethodology“. Foundational Relationships between Ethnomethodology and System Design. In: *Human-Computer Interaction* 13 (4), 395–432.
- Fischer, K. (2006): What computer talk is and isn't: Human-Computer Conversation as Intercultural Communication. Weinberg: AQ-Verlag.
- Fischer, K. (2010): Why it is interesting to investigate how people talk to computers and robots: Introduction to the special issue. In: *Journal of Pragmatics* 42 (9), 2349–2354.
- Fischer, K. (2011): How People Talk with Robots: Designing Dialogue to Reduce User Uncertainty. In: *AI Magazine* 32 (4), 31–38.
- Fischer, K. (2016): Robots as Confederates. How robots can and should support research in the humanities. *Paper presented at the Robophilosophy 2016. Frontiers in Artificial Intelligence and Applications*.

- Frohlich, D./Luff, P. (1990): Applying the technology of conversation to the technology for conversation. In: Luff, P./Nigel, D./Frohlich, D.M. (Hg.): *Computers and Conversation*. London etc.: Academic Press, 187–220.
- Garfinkel, H. (1967): *Studies in Ethnomethodology*. Englewood Cliffs: Prentice-Hall.
- Goodwin, C. (2000): Action and embodiment within situated human interaction. In: *Journal of Pragmatics* 32 (10), 1489–1522.
- Hausendorf, H. (2015): Interaktionslinguistik. In: Eichinger, L. (Hg.): *Sprachwissenschaft im Fokus. Positionsbestimmungen und Perspektiven*. Berlin, München, Boston: De Gruyter, 43–69.
- Heath, C./Luff, P. (2018): The Naturalistic Experiment. Video and Organizational Interaction. In: *Organizational Research Methods* 21 (2), 466–488.
- Herring, S. (Hg.) (1996): *Computer-Mediated Communication. Linguistic, social and cross-cultural perspectives*. Amsterdam, Philadelphia: Benjamins.
- Hindmarsh, J./Heath, C. (2000): Embodied Reference: A Study on Deixis in the Workplace. In: *Journal of Pragmatics* 32, 1855–1878.
- Hirschauer, S. (1999): Die Praxis der Fremdheit und die Minimierung von Anwesenheit. Eine Fahrstuhlfahrt. In: *Soziale Welt* 50 (3), 221–245.
- Huang, C. M./Mutlu, B. (2012): Modeling and evaluating narrative gestures for humanlike robots. In: *Robotics: Science and Systems* 2013. URL: <http://roboticsproceedings.org/rss09/p26.pdf>
- Keevallik, L. (2018): What Does Embodied Interaction Tell Us About Grammar? In: *Research on Language and Social Interaction* 51 (1), 1–21.
- Kendrick, K. H. (2017): Using Conversation Analysis in the Lab. In: *Research on Language and Social Interaction* 50, 1–11.
- Kuno, Y./Sadazuka, K./Kawashima, M./Yamazaki, K./Yamazaki, A./Kuzuoka, H. (2007): Museum Guide Robot Based on Sociological Interaction Analysis. In: *CHI 2007*, San Jose, 1191–1194.
- Kramer, M./Yaghoubzadeh, R./Kopp, S./Pitsch, K. (2013): A conversational virtual human as autonomous assistant for elderly and cognitively impaired users? Social acceptability and design considerations. In: *INFORMATIK 2013, Workshop „Wer hat Angst vor autonomen Maschinen?“*, 1105–1119.
- Kuhlen, A. K./Brennan, S. E. (2013): Language in Dialogue. When Confederates Might be Hazardous to Your Data. In: *Psychonomic Bulletin and Review* 20, 54–72.
- Lotze, N. (2016). *Chatbots. Eine linguistische Analyse*. Frankfurt/Main: Peter Lang.
- Mondada, L. (2012): Deixis. An integrated interactional multimodal analysis. In: Bergmann, P./Brenning, J./Pfeiffer, M./Reber, E. (Hg.): *Prosody and Embodiment in Interactional Grammar*. Berlin: De Gruyter, 173–206.
- Mondada, L. (2014): The local constitution of multimodal resources for social interaction. In: *Journal of Pragmatics* 65, 137–156.
- Pitsch, K. (2016): Limits and Opportunities for Mathematizing Communicational Conduct for Social Robotics in the Real-World? – Towards enabling a Robot to make use of the Human’s Competences. In: *AI & Society* 31 (4), 587–593.
- Pitsch, K. (2020): Answering a robot’s questions. Participation dynamics of adult-child-groups in encounters with a museum guide robot. In: *Réseaux* (Themenheft: „Ethnographie des agents conversationnels“) 2-3 (220-221), 113–150. URL: <https://doi.org/10.3917/res.220.0113>

- Pitsch, K./Dankert, T./Gehle, R./Wrede, S. (2016): Referential practices. Effects of a museum guide robot suggesting a deictic 'repair' action to visitors attempting to orient to an exhibit. In: *Ro-Man 2016*, New York, 225–231.
- Pitsch, K./Lohan, K. S./Rohlfing, K./Saunders, J./Nehaniv, C.L./Wrede, B. (2012): Better be reactive at the beginning. Implications of the first seconds of an encounter for the tutoring style in human-robot-interaction. In: *Ro-Man 2012*, 974–981.
- Pitsch, K./Vollmer, A.-L./Mühlig, M. (2013): Robot feedback shapes the tutor's presentation. How a robot's online gaze strategies lead to micro-adaptation of the human's conduct. In: *Interaction Studies* 14 (2), 268–296.
- Pitsch, K./Vollmer, A.-L./Rohlfing, K./Fritsch, J./Wrede, B. (2014a): Tutoring in adult-child-interaction: On the loop of the tutor's action modification and the recipient's gaze. In: *Interaction Studies* 15 (1), 55–98.
- Pitsch, K./Wrede, S. (2014b): When a robot orients visitors to an exhibit. Referential practices and interactional dynamics in real world HRI. In: *Ro-Man 2014*, Edinburgh, 36–42.
- Porcheron, M./Fischer, J. E./Reeves, S./Sharples, S. (2018): Voice Interfaces in Everyday Life. In: *CHI 2018*, Montréal, QC, Canada, paper 640.
- Rammert, W. (2008): Where the action is: Distributed agency between humans, machines, and programs. In: Seifert, U./Kim, J. H./Moore, A. (Hg.): *Paradoxes of Interactivity. Perspectives of Media Theory, Human Computer Interaction, and Artistic Investigations*. Bielefeld: Transcript, 62–92.
- Riek, L. (2012): Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines. In: *Journal of Human-Robot Interaction* 1 (1), 119–136.
- Ruiter, J. P. de/Albert, S. (2017): An Appeal for a Methodological Fusion of Conversation Analysis and Experimental Psychology. In: *Research on Language and Social Interaction* 50, 90–107.
- Schegloff, E. A. (1984): On some gestures' relation to talk. In: Heritage, J./Atkinson, J.M. (Hg.): *Structures of Social Action*. Cambridge: Cambridge University Press, 266–296.
- Schegloff, E. A. (1993): Reflections on Quantification in the Study of Conversation. In: *Research on Language and Social Interaction* 26 (1), 88–128.
- Schegloff, E. A. (1996): Issues of Relevance for Discourse Analysis: Contingency in Action, Interaction, and Co-Participant Context. In: Hovy, E.H./Scott, D.R. (Hg.): *Computational and Conversational Discourse: Burning Issues – An Interdisciplinary Account*. Berlin, Heidelberg: Springer, 3–38.
- Stivers, T. (2015): Coding Social Interaction: A Heretical Approach in Conversation Analysis? In: *Research on Language and Social Interaction* 48 (1), 1–19.
- St Clair, A./Mead, R./Mataric, M.J. (2011): Investigating the effects of visual saliency on deictic gesture production by a humanoid robot. In: *RO-MAN 2011*, 210–216.
- Stukenbrock, A. (2015): *Deixis in der face-to-face-Interaktion*. Berlin: De Gruyter.
- Stukenbrock, A. (2018): Mobile dual eye-tracking in face-to-face interaction. The case of deixis and joint attention. In: Bröne, G./Oben, B. (Hg.): *Eye-tracking in interaction. Studies on the role of eye gaze in dialogue*. Amsterdam: Benjamins, 265–302.
- Suchman, L. (2006): *Human and Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press.
- Vollmer, A.-L./Mühlig, M./Steil, J.J./Pitsch, K./Fritsch, J./Rohlfing, K.J./Wrede, B. (2014): Robots Show Us How to Teach Them: Feedback from Robots Shapes Tutoring Behavior during Action Learning. In: *Plos One* 9 (3), e91349.

- Weingarten, R. (1989): Die Verkabelung der Sprache. Grenzen der Technisierung von Kommunikation. Frankfurt a.M.: Fischer.
- Yamazaki, K./Kawashima, M./Kuno, Y./Akiya, N./Burdelski, M./Yamazaki, A./Kuzuoka, H. (2007): Prior-to-request and request behaviors within elderly day care. Implications for developing service robots for use in multiparty settings. In: *ECSCW 2007*, 61–78.
- Yamazaki, A./Yamazaki, K./Burdelski, M./Kuno, Y./Fukushima, M. (2010): Coordination of verbal and non-verbal actions in human-robot interaction at museums and exhibitions. In: *Journal of Pragmatics* 42 (9), 2398–2414.

