

Rita Tamara Vallentin, Dorothea Horst (Hg.)

SPRACHE ENTGRENZEN

Beiträge zu einer kulturwissenschaftlichen Linguistik



[transcript]

STUDIEN DER KULTURWISSENSCHAFTLICHEN
GESELLSCHAFT

Rita Tamara Vallentin, Dorothea Horst (Hg.)
Sprache entgrenzen

Editorial

Die Studien der Kulturwissenschaftlichen Gesellschaft versammeln innovative Beiträge, die aus theoretisch-programmatischer oder empirischer Perspektive kulturwissenschaftlichen Fragestellungen nachgehen. Die Buchreihe nimmt Sammelbände und Monographien auf, die vornehmlich den einzelnen Sektionen der Kulturwissenschaftlichen Gesellschaft (KWG) entstammen und folgende thematische Bereiche umfassen: Materielle Kulturen, Kulturphilosophie und Kulturtheorie, Kulturwissenschaftliche Ästhetik, Kulturwissenschaftliche Border Studies, Medienkulturen, Naturen/Kulturen, Sprache und kommunikative Praktiken, Transkulturelle Lebenswelten, Raum und Kultur, Wissenskulturen und Kritische Methodologie. Über die Arbeiten der KWG hinaus begrüßt die Reihe Forschungsarbeiten, die zur Stärkung methodischer, theoretischer und anwendungsorientierter Ansätze im Umgang mit kulturellen Praktiken beitragen.

Rita Tamara Vallentin (Dr.), geb. 1983, lehrt und forscht an der kulturwissenschaftlichen Fakultät der Europa-Universität Viadrina Frankfurt (Oder) an der Schnittstelle zwischen Linguistik und Soziologie zu sprachlichen Praktiken, Zugehörigkeit und Mensch-Maschine Interaktion.

Dorothea Horst (Dr.), geb. 1982, arbeitet am Wissenschaftszentrum Berlin für Sozialforschung (WZB). Sie lehrt und forscht an der Schnittstelle zwischen angewandter Sprach- und Medienwissenschaft zu Fragen medienpezifischen Wahrnehmens und Verstehens sowie zu epistemologischen Implikationen medialer Forschungspraktiken.

Rita Tamara Vallentin, Dorothea Horst (Hg.)

Sprache entgrenzen

Beiträge zu einer kulturwissenschaftlichen Linguistik

[transcript]

Dieser Sammelband wurde finanziell von der Kulturwissenschaftlichen Gesellschaft e.V. (KWG) gefördert.

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <https://dnb.dnb.de/> abrufbar.



Dieses Werk ist lizenziert unter der Creative Commons Attribution-ShareAlike 4.0 Lizenz (BY-SA). Diese Lizenz erlaubt unter Voraussetzung der Namensnennung des Urhebers die Bearbeitung, Vervielfältigung und Verbreitung des Materials in jedem Format oder Medium für beliebige Zwecke, auch kommerziell, sofern der neu entstandene Text unter derselben Lizenz wie das Original verbreitet wird.

<https://creativecommons.org/licenses/by-sa/4.0/>

Die Bedingungen der Creative-Commons-Lizenz gelten nur für Originalmaterial. Die Wiederverwendung von Material aus anderen Quellen (gekennzeichnet mit Quellenangabe) wie z.B. Schaubilder, Abbildungen, Fotos und Textauszüge erfordert ggf. weitere Nutzungsgenehmigungen durch den jeweiligen Rechteinhaber.

2025 © Rita Tamara Vallentin, Dorothea Horst (Hg.)

transcript Verlag | Hermannstraße 26 | D-33602 Bielefeld | live@transcript-verlag.de

Umschlaggestaltung: Maria Arndt, Bielefeld

Umschlagabbildung: Hotpot API (mit KI generiert)

Druck: Majuskel Medienproduktion GmbH

Print-ISBN: 978-3-8376-7061-5

PDF-ISBN: 978-3-8394-7061-9

Buchreihen-ISSN: 2749-8050

Buchreihen-eISSN: 2749-8069

Gedruckt auf alterungsbeständigem Papier mit chlorfrei gebleichtem Zellstoff.

Inhalt

Sprache entgrenzen

Dorothea Horst, Rita Tamara Vallentin 7

Sprache, Sprechen und Interagieren zwischen Mensch und Maschine

»Es ist vor allen Dingen dieser Ton«

Sprechen mit Alexa, die vergessene Macht der Laute und warum Sprache kein
abstraktes System ist

Britta Schneider 19

Einstimmen mit Sprachassistenten

Einblicke in monolinguale Bias in mehrsprachigen Settings

Didem Leblebici 47

Die Heterogenität der Sprechenden

Die Entgrenzung von Zeit und Raum in Mensch-Maschine-Kommunikation

Marie-Theres Fester-Seeger 75

»Man hat schon so einen Austausch irgendwie«

Konzepte von Kommunikation und Interaktion zwischen Mensch und Maschine im
Diabetesmanagement

Rita Tamara Vallentin 105

Sprache, Sprechen und Interagieren in mediatisierten Kontexten

Kommunikationspraktiken mit Vertreter*innen der Spezies Oktopus

Filmische, literarische und naturwissenschaftliche Evidenz

Konstanze Jungbluth 141

Die Stimme als Interface

Entgrenzungen von Mensch und Maschine anhand von Spike Jonzes Film HER

Dorothea Horst 179

Sprache-Sprechen, Hören, Sehen, Fühlen

Eine medienästhetische Perspektive auf die Multimodalität des Sprechens in
audiovisuellen Medien

Cornelia Müller 213

Sprechen in audiovisuellen Medien

Systematik für eine medienästhetische Analyse multimodaler Kommunikation

Clara Kindler-Mathôt, Cornelia Müller, Jana Katharina Junge, Katerina

Papadopoulou 253

Sprache entgrenzen

Dorothea Horst, Rita Tamara Vallentin

»Diese KI quatscht und quatscht und quatscht« (Kremp 2024). Als ein Reporter des Spiegel die menschlich sprechende Stimme von Google, Gemini Live, testet, fragt er sie nach der Spezifik eines E-Bass-Modells aus den 1960er Jahren. Was dann passiert, ist ein kaum zu stoppender Redefluss »über den wunderbaren Klang dieser Instrumente und über Details der verwendeten Materialien, sie kam vom Hundertsten ins Tausendste« (ebd.). Der Reporter vermag die Informationsflut nur mit einer Nachfrage über die damalige Farbwahl zu bremsen und nutzt hier bezeichnenderweise aktive Verben der Kommunikation für die Maschine, mit der er zu kommunizieren meint: das System »berichtet«, »versteht« oder »fragt nicht nach«. Trotz dieser evidenten Anthropomorphisierung von Gemini Live als lebendigem kommunikativem Gegenüber scheint sich der Autor ob dieser konzeptuellen Entgrenzung mahndend zur Raison zu rufen: »Umso wichtiger ist es, Anstand zu wahren. So menschlich sie auch wirkt: Gemini Live ist mehr Maschine als Mensch.« Der Rückzug in kategoriale Grenzen von Menschlichem und Nicht-Menschlichem bringt (vorläufig) Beruhigung.

Die Auseinandersetzung mit solchen Grenzziehungen rückt zunehmend in den Fokus sprach-, sozial- und kulturwissenschaftlicher Forschung. Dabei stehen häufig Nicht-Menschen im Sinne von Tieren (Lind 2022; Bolton/Loovers 2023), Maschinen (Coeckelberg 2011; Lind 2022), Medien und Technologien (Müller/Kappelhoff 2018; Horst 2023) oder Pflanzen (Stobbe et al. 2022) im Mittelpunkt. Der Mensch wird in diesen Gefügen nicht als abgeschlossenes Zentrum von an ihn angrenzenden Um-Welten, sondern in seinen Beziehungen und Verflechtungen zu nicht-menschlichen Entitäten als *Mit-Welten* gedacht. Damit einher gehen Fragen nach Erscheinungsformen dieser Mit-Welten sowie nach dem Wesen, der Materialität und auch der Medialität von Sprache, Sprechen und Interaktion in diesen Verflechtungsprozessen. Nicht zuletzt verbinden sich damit konzeptuelle Themen: Wie

bedingen, beeinflussen und verändern Interaktionen mit diesen Mit-Welten ›den Menschen‹ sowie Vorstellungen von ihm? Die Philosophin Eva Weber-Guskar beschäftigt sich zum Beispiel in ihrem 2024 erschienenen Buch mit der dazu naheliegenden Frage im Untertitel, »[w]ie wir mit emotionaler KI unser Leben verändern«. Ausgehend von technischen Entwicklungen im Bereich der Nachbildung menschlicher Emotionalität und der Diskussion um die Möglichkeit technologischer Formen von Bewusstsein, sieht sie tiefgreifende Veränderungen im menschlichen Gefühlsleben voraus und plädiert für eine Auseinandersetzung mit »Gefühle[n] der Zukunft«. Dass diese als grundlegende Transformationen menschlicher Gefühlswelten längst Alltag sind, ist für Weber-Guskar unstrittig. Was fehle, sei eine kritische Reflexion der Wünschbarkeit und des Grades der Veränderung im menschlichen Dasein. Im Kern kreist ihr Buch also um eine normative Frage: Wie weit sollen wir uns emotional mit Maschinen oder Medien verflechten?

Für die Philosophin Weber-Guskar zeichnen sich die Beziehungen von Menschen zu Technologien wie künstlicher Intelligenz dadurch aus, dass sie einseitig sind im Hinblick auf die Fähigkeit, emotional zu empfinden. Obwohl Sprachassistenzsysteme wie Alexa im stimmlichen Ausdruck mittlerweile zwischen verschiedenen Stimmungsmodi wechseln können (Peters 2019) oder Emotionserfassungssoftware menschlichen Gesichtsausdrücken auf Bildern oder im Verhalten Befindlichkeiten und Gefühlszustände zuordnen kann, basieren diese Fortschritte nicht auf affektivem Erleben der Maschinen, sondern auf Algorithmen. Durch die Formalisierung von menschlicher Sprache und durch maschinelles Lernen wird das Wahrscheinlichste ausgerechnet bzw. das Passendste symbolisch zugeordnet. »Wahre Emotionen sind das nicht, es fehlt die phänomenologische Komponente: wie es sich anfühlt« (Keller 2024: 45). Weber-Guskars »Gefühle der Zukunft« mögen vor diesem Hintergrund zwar auf den ersten Blick ähnlich scheinen wie das, was beispielsweise unter den Begriffen »Beziehung« oder »Freundschaft« bekannt ist, aber der entscheidende Unterschied besteht darin, dass es keine Wechselseitigkeit im (emotionalen) Empfinden gibt. Wir richten unsere Gefühle auf Dinge, die diese – qua Unvermögen – nicht erwidern können.

Man könnte der Philosophin vorwerfen, sie reproduziere damit im Kern die Vorstellung des Menschen als universellem Zentrum aller Existenz. Er allein kann umfassend fühlen, er allein kann (und soll) entscheiden, wie weit er sich den neuen Technologien aussetzen und ihnen Raum geben will. Ein gewisses Unbehagen ist Weber-Guskars Ausführungen zwischen den Zeilen allerdings anzumerken. Wo soll es hinführen, wenn wir den Dingen einfach ih-

ren Lauf lassen?! Ein apokalyptisches Szenario von Autonomieverlust, totaler Kontrollierbarkeit und in letzter Instanz Existenzbedrohung scheint auf. Ob wir mit Weber-Guskars Annahmen letztlich mitgehen oder nicht – Tatsache ist: Sie unterstreicht die grundsätzliche *Verflechtung* von Mensch und Technologien, mit denen alltäglich umgegangen wird, und aus der körperliche, emotionale und gesellschaftliche Transformationen resultieren. Das Bild von Technologien oder Medien im Sinne bloßer Handhabungsdinge, die uns – und denen wir – neutral gegenüberstehen, ist Vergangenheit. Zu tiefgreifend und *spürbar* ist der Wandel, den wir mit ihnen und durch sie durchlaufen: Gesundheitstracker, die unseren Schlaf vermessen und medizinische Werte überprüfen, audiovisuelle Bilder, die uns affizieren, Sprachassistenten, mit denen wir sprechen – in unserem Alltag ist diese Verflechtung über die Mediatisierung und Materialisierung von Sprache und Zeichenhaftigkeit omnipräsent.

Die Anerkennung dieser grundsätzlichen Verflechtung von Mensch und Umwelt bildet einen wichtigen Anknüpfungspunkt zu posthumanistischen Perspektiven auf Sprache und Kommunikation, die sowohl kommunikative anthropomorphe oder nicht-anthropomorphe Gegenüber wie auch die situativen Rahmenbedingungen eines mediatisierten oder Face-to-Face-, eines privaten oder öffentlichen Kontexts berücksichtigen. Im Gegensatz zu dem bei Weber-Guskar durchscheinenden Unbehagen angesichts einer drohenden Dezentrierung des Menschen und der damit verbundenen Notwendigkeit, die neuen Technologien zu regulieren, macht der Posthumanismus den Anti-Anthropozentrismus zum Kern seiner Überlegungen. Er dezentriert und zerlegt den Menschen als sich selbst genügende, wichtigste und höchste (Bewusstseins-)Instanz: »The figure of Vitruvian Man [...] has broken through its enclosing circle and square, and spread across the cosmos« (Hassan 1977: 843). Diese Dekonstruktion impliziert einen empirischen und einen konzeptuell wissenschaftlichen Aspekt: Mit der Kritik am humanistischen Konzept des Subjekts, das universell, autonom und mit sich selbst identisch ist, eröffnen posthumanistische Perspektiven eine »Vielfalt von neuen Untersuchungsgegenständen, wie Tieren, Pflanzen oder technologischer Materie« (Scheerer 2022: 229) und ihrer Verflechtung mit »dem«, und Konstituierung »des«, Menschen. Gleichzeitig bedeutet dies nicht notwendigerweise das Ende der Menschheit.

It signals instead the end of a certain conception of the human, a conception that may have applied, at best, to that fraction of humanity who had the wealth, power, and leisure to conceptualize themselves as autonomous

beings exercising their will through individual agency and choice. (Hayles 1999: 286)

Der Posthumanismus eröffnet einen Reflexionsraum für Praktiken der Wissensproduktion, alltagsweltliche wie wissenschaftliche, rekonstruierend wie prospektiv. Indem er von etablierten anthropozentrischen Überzeugungen und Gewissheiten abrückt, tritt Uneindeutiges, Dynamisches und Vorläufiges ins Zentrum. Anstelle fester, abgeschlossener Strukturen, unveränderbarer Naturalisierungen und Dualismen (vgl. Braidotti 2013: 104) weisen Netzwerke und lebendige Systeme auf Verflechtungen, Wechselbeziehungen und *Inter-Aktionen* hin. Ein dergestalt holistischer Blick auf reale Lebensbedingungen trägt der Komplexität technologisch vermittelter Körper und sozialen Formen der menschlichen Leiblichkeit Rechnung (vgl. ebd.).

Damit knüpft die posthumanistische Perspektive an kulturwissenschaftlichen Perspektiven auf Sprache an, die diese genuin als kulturelle Praxis begreifen, die in körperliche, materielle, mediale und technologische Umgebungen eingebettet ist. Dabei stehen Fragen zur Möglichkeit und Form der Verständigung im Vordergrund (Metten 2014). Solche kommunikativen Praktiken, die nicht an Artengrenzen von ›Menschlich‹ vs. ›Nicht-Menschlich‹ aufhören, eröffnen einen Blick auf bisherige Grenzziehungen und ihre kritische Reflexion: ›Der Mensch‹ als exklusiv sprachliches Wesen, das Sprechen von menschlichem Angesicht zu menschlichem Angesicht als Standardfall, standardisierte Schriftsprache als internalisiertes universelles Sprachwissen, das situativ zur Anwendung kommt – diese Konzepte werden durch hybride Kommunikationskontexte zwischen Mensch und Maschine, Mensch und Tier oder mediatisierte Interaktionen befragt und zur Diskussion gestellt. Hier ergeben sich auch Anschlusspunkte an kulturwissenschaftliche Grenzforschung. Anstatt Grenzen als Trennung zwischen einem klar umrissenen Innen und Außen aufzufassen, fragt diese explizit nach dem »Liminale[n], Periphere[n], Marginale[n] und Transgressive[n]« (Gerst/Klessmann/Krämer 2021: 15). Indem so die Grenze selbst sowie ihre Überschreitung in den Blick rücken, wird ein Möglichkeitsraum geschaffen, in dem Neues und Vorläufiges entstehen kann.

Als ein solcher Möglichkeitsraum versteht sich dieser Sammelband. In allen darin versammelten Beiträgen von Sprachwissenschaftlerinnen der Europa-Universität Viadrina in Frankfurt (Oder) wird die Vorstellung von Sprache als kognitivem, abstraktem System empirisch wie theoretisch übersritten. Anhand von konkreten Interview- oder Interaktionsdaten in der

Mensch-Maschine-Interaktion eröffnen zum einen soziolinguistische Konzepte von Sprachideologien, (räumlichen) Repertoires, *Languaging*, *Distributed Cognition* und ein Blick auf konkrete Materialisierungen von Zeichen Sprache als in räumlichen, sozialen und zeitlichen Verortungen und Verflechtungen gefasst. Im Rahmen mediatisierter Kontexte wiederum bilden verkörperte und multimodale Verflechtungen die Grundlage für *Sinmgebung*: Ganzheitliche Erfahrung – Fühlen, Hören, Sprechen, Sehen, Riechen, Tasten –, Affekt und Affizierung werden hier als zentrale Grundlagen von Interaktion verstanden. Dass diese beiden inhaltlichen Schwerpunkte, die sich im Band als »Sprache, Sprechen und Interagieren zwischen Mensch und Maschine« und »Sprache, Sprechen und Interagieren in mediatisierten Kontexten« wiederfinden, einander nicht ausschließen, sondern wechselseitig ergänzen, ist Ausdruck einer genuin kulturwissenschaftlichen Linguistik, wie sie an der Europa-Universität Viadrina seit Jahren betrieben wird.

Der Teil »Sprache, Sprechen und Interagieren zwischen Mensch und Maschine« wird durch Britta Schneider eröffnet. Sie widmet sich in ihrem Beitrag dem Phänomen der Lautlichkeit in der Bedienung von Computern und den damit stattfindenden Interaktionen. Dabei stellt sie die Bedeutsamkeit zeichenhafter Materialität heraus, da diese maßgeblich zu (traditionellen) Definitionen, als auch zu normativen und sozialen Vorstellungen von »Sprache« beiträgt. Mit einem kritischen Blick auf Sprachkonzepte der Moderne verdeutlicht sie die »Wiederkehr« der (maschinellen) Stimme und welchen Einfluss sie auf die Sprachwahrnehmung und Normenvorstellung von menschlichen Nutzenden hat.

Didem Leblebici befasst sich mit dem Begriff des räumlichen Repertoires, welches Sprache als verkörpert, kontextuell eingebettet und sozial, temporal und räumlich distribuiert fasst. Sie verknüpft dabei semiotische und pragmatische »Einstimmungen« (*attunements*) in mehrsprachigen Interaktionen zwischen Menschen und Sprachassistenten mit sprachideologischen Vorstellungen, die von den Nutzenden in die Geräte eingeschrieben werden. Dabei zeigt sie, wie die Geräte nicht nur materielle, sondern auch kulturelle und soziale Ressourcen für räumliche Repertoires in der Mensch-Maschine-Interaktion darstellen und wie sich Aspekte populärer Diskurse über KI in den Einstimmungen der Nutzenden wiederfinden.

Aus der Perspektive der *Distributed Cognition* und eines Verständnisses von Sprache als *Languaging* entwirft Marie-Theres Fester-Seeger einen analytischen Blick auf die Herstellung von Bedeutung zwischen Mensch und Maschine. Bedeutung wird durch räumliche, zeitliche und soziale Prozesse

hervorgebracht, die sich sowohl in der Maschine als solcher manifestieren (zum Beispiel über Programmierung, Materialität, Affordanzen und Marketing für Sprachassistenten) als auch über die körperlichen und kognitiven Prozesse des ›Menschen‹ in der Interaktionssituation. Daraus ergibt sich eine komplexe Dynamik, die traditionelle Modelle von Kommunikation im Sinne zweier Informationen austauschender Entitäten in Frage stellt.

Rita Tamara Vallentin betrachtet Konzepte von Kommunikation und Interaktion unter den Vorzeichen sich verändernder Partner*innendynamiken in Austauschbeziehungen mit Maschinen. Für den spezifischen Fall des Diabetesmanagements befindet sie anhand von Interviews mit Nutzer*innen dieser Maschinen, dass einschlägige Begriffe von Interaktion im Sinne von wechselseitigen und bewussten Handlungen zwischen mindestens zwei Akteur*innen neu gedacht werden müssen. Sie zeichnet Suchbewegungen der Nutzenden nach, die versuchen, diese neuen Beziehungsformen zu versprachlichen.

Im zweiten Teil »Sprache, Sprechen und Interagieren in mediatisierten Kontexten« kommen vor allem medienwissenschaftlich phänomenologische Perspektiven zum Tragen. Konstanze Jungbluth befasst sich in ihrem Beitrag mit einem liquiden Konzept von Sprache in Mensch-Tier- und Mensch-Mensch-Interaktionen, welches zeitliche, räumliche, haptische, olfaktorische und visuelle Phänomene einschließt. Am Beispiel der Mensch-Tier-Interaktion im Film »Mein Lehrer, der Oktopus«, aber auch von Berichten zu Mensch-Oktopus-Beziehungen, erkundet sie entgrenzte Vorstellungen von Vergemeinschaftung durch Sprache, die durch »liquide« kommunikative Praktiken eine heran»astende« Verständigung zwischen Mensch und Oktopoden möglich machen.

Die Stimme ist das zentrale Konzept von Dorothea Horsts Beitrag. Aus filmwissenschaftlicher Perspektive diskutiert sie anhand einer technischen, körperlosen, aber doch menschenähnlichen Stimme im Film »Her« etablierte Vorstellungen von stimmlicher Menschlichkeit. Sie zeigt, wie mit einem post-humanistischen Blick das Infragestellen dieser grundlegenden Annahmen zur Stimme auch an etablierten Konzepten von »Sprache« rüttelt.

Cornelia Müller schlägt in ihrem Beitrag ein genuin multimodales Verständnis von Sprache in audiovisuellen Medien (aber auch darüber hinaus) als »Sprache-Sprechen« vor. Hierfür entwickelt sie sowohl konzeptuell als auch methodisch eine interdisziplinäre linguistisch-filmwissenschaftliche Perspektive. Sie zeichnet Parallelen zwischen gesprochener Sprache und audiovisuellen Bildern nach und formuliert Grundsätze einer Medienästhetik,

die Bedeutungsherstellung in ihrer sequentiellen und bewegten Entfaltung sowie in sich affizierenden fühlenden Körpern verortet.

Der letzte Beitrag von Clara Kindler-Mathôt, Cornelia Müller, Jana Katharina Junge und Katerina Papadopoulou zeigt das analytische Potenzial dieser konzeptuellen und methodischen Ideen. Anhand von politischen Reden im Deutschen Bundestag demonstrieren die Autorinnen eine systematische Analyse von multimodalem Sprechen in audiovisuellen Medien. Indem sie dieses mediatisierte Sprechen als multimodale Orchestrierung von Ausdrucksbewegungen fassen, zeichnen sie nach, wie diese Gestalten Affektdynamiken hervorbringen und immer im Kontext ihrer medialen Ökologien zu verstehen sind.

Allen Beiträgen liegt ein Verständnis zugrunde, das Sprache als sich fortwährend verändernde kulturelle Praxis begreift: in ihrem kreativen *doing* durch Nutzer*innen und neue Formen der Mediatisierung ebenso wie durch die Rasanz technologischer Innovationen. Als unendlich vielfältiges und lebendiges Tun ist sie nie vollständig fassbar. Der Band versteht sich folglich als Kaleidoskop einer sich wandelnden Gestalt und als im besten Sinne vorläufig. Indem er so etablierte Vorstellungen von Sprache hinterfragt, die den öffentlichen und akademischen Diskurs in der westlichen Kultur seit der frühen Moderne dominiert haben, »eröffnet [er] Räume für ein Andersdenken und weist so neue und andere Handlungsoptionen auf« (Metten 2016: 13), im Alltag ebenso wie in der Wissenschaft. Die Herausgeberinnen danken den Autorinnen für ihre vielfältigen Perspektiven als Beitrag zur Eröffnung und Gestaltung dieser Denkräume, ebenso wie Carlynn Löwisch für ihre hilfreiche Unterstützung bei der formalen Vorbereitung des Bandes.

Literaturangaben

- Bolton, Maggie/Jan P. L. Looovers (2023): *Sentient Entanglements and Ruptures in the Americas: Human-Animal Relations in the Amazon, Andes, and Arctic*, Leiden: Brill.
- Braidotti, Rosi (2013): *The Posthuman*, Cambridge: Polity Press.
- Coeckelbergh, Mark (2011): *Humans, Animals, and Robots: A Phenomenological Approach to Human-Robot Relations*. In: *International Journal of Social Robotics* 3:2, 197–204. doi.org/10.1007/s12369-010-0075-6.
- Gerst, Dominik/Maria Klessmann/Hannes Krämer (2021): *Einleitung*. In: Gerst, Dominik/Maria Klessmann/Hannes Krämer (Hg.), *Grenz-*

- forschung. Handbuch für Wissenschaft und Studium, Baden-Baden: Nomos, 9–26.
- Hassan, Ihab (1977): Prometheus as Performer: Toward a Posthumanist Culture? In: *The Georgia Review* 31:4, 830–850.
- Hayles, N. Katherine (1999): *How we became Posthuman. Virtual Bodies in Cybernetics, Literature and Informatics*, Chicago: Chicago Press.
- Horst, Dorothea (2023): Digitale Werkzeugkritik als essentielle Bedingung einer innovativen digitalen Lehr-Lernkultur. In: Mrohs, Lorenz/Miriam Hess/Konstantin Lindner/Julia Schlüter/Sven Overhage (Hg.): *Digitalisierung in der Hochschullehre – Perspektiven und Gestaltungsoptionen*. Bamberg: University of Bamberg Press, 47–64.
- Keller, Paula (2024): Erfindet endlich neue Gefühle! *DIE ZEIT* 34/2024: 45. Abrufbar unter: <https://www.zeit.de/2024/34/gefuehle-der-zukunft-eva-wober-guskar-ki-philosophie>
- Kremp, Matthias (2024): Diese KI quatscht und quatscht und quatscht. *Spiegel Online* 15.08.2024. Abrufbar unter: https://www.spiegel.de/netzwelt/gadgets/gemini-live-von-google-diese-ki-ist-kaum-zu-stoppen-a-do91f6cb-e55b-493e-9ffe-e37e6204effb?sara_ref=re-xx-cp-sh
- Lind, Miriam (2022): *Mensch – Tier – Maschine. Sprachliche Praktiken an und jenseits der Außengrenze des Humanen*, Bielefeld: transcript.
- Metten, Thomas (2014): *Kulturwissenschaftliche Linguistik. Entwurf einer Medientheorie der Verständigung*. Berlin/Boston: De Gruyter.
- Metten, Thomas (2016): Konturen der Kulturwissenschaft/en – einleitende Überlegungen. In: *Kulturwissenschaftliche Zeitschrift* 1, 5–16.
- Müller, Cornelia/Hermann Kappelhoff (2018): *Cinematic Metaphor. Experience – Affectivity – Temporality*. Berlin/Boston: De Gruyter Mouton.
- Peters, Jay (2019): Alexa’s voice can now express disappointment and excitement. *The Verge* 27.11.2019. Abrufbar unter: <https://www.theverge.com/2019/11/26/20984629/amazon-alexa-voice-disappointment-empathetic-happy-excited-newscaster-music-us-australia>
- Scheerer, Katharina (2022): Posthumanismus: Eine Einführung. In: Berlich, Sebastian/Holger Grevenbrock/Katharina Scheerer (Hg.), *Where Are We Now? – Orientierungen nach der Postmoderne*, Bielefeld: transcript, 225–242.
- Stobbe, Urte/Anke Kramer/Berbeli Wanning (2022): *Literaturen und Kulturen des Vegetabilen. Plant Studies – Kulturwissenschaftliche Pflanzenforschung*, Berlin: Peter Lang.

Weber-Guskar, Eva (2024): Gefühle der Zukunft. Wie wir mit emotionaler KI unser Leben verändern, Berlin: Ullstein.

Dr. Dorothea Horst
Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
dorothea.boehme@wzb.eu
<https://www.wzb.eu/de/personen/dorothea-boehme>

Dr. Rita Tamara Vallentin
Europa-Universität Viadrina
Große Scharrnstrasse 59
15230 Frankfurt (Oder)
vallentin@europa-uni.de
<https://www.kuwi.europa-uni.de/de/professuren-mitarbeitende/wimi-vallentin/index.html>

Sprache, Sprechen und Interagieren zwischen Mensch und Maschine

»Es ist vor allen Dingen dieser Ton«

Sprechen mit Alexa, die vergessene Macht der Laute und warum Sprache kein abstraktes System ist

Britta Schneider

Abstract *In diesem Beitrag wird die Rolle von lautlich-tonalen Praktiken im Sprechen mit stimmungsgesteuerten Computern auf der Grundlage von Interviews mit Nutzenden dieser Geräte untersucht. Die Bedienung von Geräten durch Laute erweckt in vielen der Befragten den Wunsch, bestimmte tonale Qualitäten – beschrieben als ›höflich‹ und ›freundlich‹ – in der Mensch-Maschine-Interaktion zu produzieren. Lautlichkeit wird eine wichtige Rolle zugeschrieben, um sozial angemessene Räume im privaten Umfeld zu schaffen. Außerdem scheint die lautliche Interaktion die Herstellung von Sozialbeziehungen, auch zu nicht-menschlichen Objekten, zu fördern. Diese Beobachtungen werden dazu genutzt, kritische Überlegungen zu Sprachideologien anzustellen, in denen Sprache als abstraktes, denotationales System konstruiert wird. Diese haben ihren Ursprung in Schriftkulturen der Moderne, die die Materialität von Zeichen und deren Konsequenz für linguistische Epistemologie, Sprachnormen und Vergemeinschaftung weitgehend unbeachtet gelassen haben.*

Keywords *Mensch-Maschine-Interaktion; Lautlichkeit; Posthumanismus; Sprachideologie*

1 Sprechen mit Computern und die Affordanzen von Lautlichkeit

Sprachassistenten sind in vielen Haushalten Alltag geworden. Sie helfen uns, das Licht an- und auszuschalten, erzählen uns, wie das Wetter wird, und spielen Lieder oder Hörspiele vor. Sie tun also Dinge, die wir vorher nicht durch lautliches Sprechen, sondern händisch in Gang gebracht haben, zum Beispiel durch die Benutzung von Lichtschaltern, Anschaltknöpfen oder Computer-

tastaturen. Befragt man Nutzende von lautgesteuerten Computern, fällt auf, dass lautliche Kommunikation mit bestimmten sozialen Anforderungen einhergeht – so berichten diese, dass sie höfliches und freundliches Verhalten im Umgang mit den Geräten für wichtig halten. Lautliche Kommunikation mit Computern hat also andere Affordanzen als händische Kommunikation mit elektrischen oder elektronischen Geräten. Auf der Grundlage von Interviewdaten diskutiere ich in diesem Beitrag, welche Vorstellungen von pragmatischer Angemessenheit Menschen in der Kommunikation mit Sprachassistenten haben. Diese Beobachtungen nutze ich, um Überlegungen zur medial-technologischen Abhängigkeit von sprachlicher Epistemologie anzustellen. Mit anderen Worten, die Beobachtung, dass Menschen Höflichkeit und Freundlichkeit auch von Computern einfordern, wenn sie mit ihnen durch Laute kommunizieren, zeigt, dass die Materialität sprachlicher Zeichen einen Einfluss auf die wahrgenommene Funktion und so auch auf die Ontologie von Sprache hat. Die affektiven, sozialen und auch poetischen Charakteristika von Sprache rücken in lautlicher Kommunikation in den Vordergrund.

Zugleich zeigen diese Effekte von Lautlichkeit die Limitationen von Vorstellungen von Sprache als abstraktem, grammatischem System von referentiellen Zeichen auf. Im Kontext des Projektes der ›Entgrenzung‹ von Sprache sehen wir, dass die Auseinandersetzung mit der Materialität von Zeichen traditionelle linguistische Theorien infrage stellt, insbesondere Konzepte von menschlicher Sprache als verbaler Denotation, verstanden als Umsetzung eines immateriellen, rationalen, grammatischen Systems. Damit werden etablierte sprachwissenschaftliche Grenzziehungen – von Sprachsystem und Sprachgebrauch, von *langue* und *parole*, und von menschlicher und nicht-menschlicher Sprache – zunehmend fragwürdig. Verändern sich unsere Vorstellungen davon, was Sprache ist, wer oder was eine Interaktionspartnerin und wer Teil unserer sprachlichen Gemeinschaften, wenn Lautsprache von Maschinen produziert wird? Welche Effekte haben Veränderungen von Sprachkonzepten auf unsere Ideen von sprachlicher Autorität, das Verhältnis von Laut, Körper und Geist, oder auf das Verständnis der Rolle nicht-menschlicher Umwelt in menschlicher Kultur?

Um diesen Fragen näher auf den Grund zu gehen, führe ich im Folgenden theoretische Konzepte aus dem Bereich der Sprachideologieforschung und der posthumanistischen Linguistik ein und wende mich bestehender Forschung zur Rolle von Lautlichkeit in Sprache zu. Im dritten Abschnitt des Kapitels führe ich in meinen methodischen Zugang ein und analysiere dann Beispiele von Interviewdaten, in denen Nutzende von sprechenden Compu-

tern über ihre Vorstellungen von angemessenem sprachlichem Verhalten in der Interaktion zwischen Menschen und Geräten berichten. Diese stelle ich dem Sprachkonzept einer Sprecherin gegenüber, die in einer oral-dominierten Sprachkultur lebt. Das Beispiel zeigt, dass das Medium der Lautlichkeit nicht nur im Falle von Sprachassistenten, sondern auch in weniger stark von schriftlichen Normierungen geprägten Kulturen einen entscheidenden Effekt auf Vorstellungen von Sprache hat – Sprachkonzepte sind also kulturell und medial geprägt. Schließlich diskutiere ich, welche Konsequenzen diese Beobachtungen für sprachwissenschaftliche Theorie und den Gegenstandsbereich der Sprachwissenschaft haben.

2 Theoretische Hintergründe und Forschungsstand – Sprachideologie, Posthumanismus, Lautlichkeit

Sprachideologieforschung (Gal/Irvine 2019; Silverstein 2014) und posthumanistische Theorien (Braidotti 2013; Ferrando 2013; Pennycook 2018) regen dazu an, Annahmen zu Sprachen als abstrakten Systemen zu hinterfragen und die materielle Wirklichkeit des Sprechens zu beleuchten. Die Konzepte der Sprachideologieforschung sind dabei älter als die Diskurse um Posthumanismus und wurden seit den 1970er Jahren insbesondere in der US-amerikanischen linguistischen Anthropologie entwickelt (z.B. Gal 1979; Silverstein 1979). Sie beschäftigen sich mit kulturell bedingten metapragmatischen Konzepten von Sprache. Dabei werden auch sprachwissenschaftliche Theorien unter sprachideologischen Gesichtspunkten betrachtet. Eine zentrale Annahme ist, dass Aussagen über und Interpretationen und Vorstellungen von Sprache immer eingebettet sind in soziale Verhältnisse und selbst soziales Handeln sind. Daher beinhalten sie unweigerlich soziale Positionierungen und werden aus diesem Grund als ›Ideologien‹ bezeichnet:

Statements about language are never only about language – and they are never only statements. [...] Statements about language always reach beyond the immediate linguistic forms; they intersect with other communicative means; they give signal about their speakers; and, inevitably, they are social actions embedded in history. [...] Statements about language are never merely statements. They entail ideological positions that are made evident

in multiple sites of life, often in contradictory and contested ways, and they have wide-ranging consequences in the material world. (Gal/Irvine 2019: 5)¹

Im Zusammenhang mit der oben eingeführten Diskussion zur Rolle von lautlicher Materialität und den daraus resultierenden Effekten auf sprachwissenschaftliche Konzepte, lässt sich zunächst feststellen, dass die Sprachideologien der europäisch basierten Sprachwissenschaft in weiten Teilen von einer Vorstellung von Sprache als abstraktem System bedeutungstragender Zeichen geprägt sind. Silverstein (2014), eine der zentralen Figuren der Sprachideologieforschung, setzt sich kritisch mit solchen denotationsfokussierten Vorstellungen auseinander, deren Wurzeln er in der europäischen Aufklärung sieht. Die Funktion von Sprache wird hier als in erster Linie referentiell und prädikativ betrachtet (ebd.: 132); sie wurde auch als »intellektualistisch« bezeichnet (Krämer 2003). In diesem bis heute in allgemeinen und in vielen linguistischen Diskursen dominanten Zugang basiert die Bedeutung von grammatisch-lexikalischen Formen auf einem »autonomous denotational mechanism«, welcher auf der Existenz von abstrakten a priori Zeichen beruht. Diese angenommenen Zeichen, die sogenannten LegiSigns (Peirce 1931:58, zitiert in Silverstein 2014: 134), werden verstanden als abstrakte, logische Klassifikationen, deren Bedeutung »native speakers« einer Sprache kennen (ebd.: 136). Zugleich geht die strukturalistische Linguistik von einer internen systematischen und autonomen Organisation dieser dem Sprechen vorgängigen Zeichen aus, die die Bedeutung von Kombinationen der Zeichen hervorbringt. Die Frage woher die Abstraktionen und Generalisierungen kommen, die ein Verständnis von Einzelvorkommen von Zeichen (auch *tokens* genannt) als zugehörig zu abstrakten Klassen (*types*) und grammatischen Kategorien erlaubt, wird nicht gestellt.

Silverstein stellt dieser, wie er sie nennt, »Enlightenment folk theory« (2014: 140) ein komplexes Theoriegerüst gegenüber, das nicht von vorgängigen Zeichen ausgeht, sondern sprachliche Zeichen als dialektale sozio-semiotische Phänomene beschreibt, die durch verschiedene, simultan existierende Formen von Indexikalität immer zugleich Text sind und Kontext herstellen (ebd.: 139). Gelungene Denotation basiert dabei nicht auf feststehenden, abstrakten

1 Die Assoziation des Begriffs *Ideologie* im alltäglichen Sprachgebrauch mit »falschen« oder radikalen Vorstellungen ist dabei nicht gemeint; vielmehr sind dies Annahmen, die einer bestimmten Gruppe zu eigen sind. Im Englischen wird das Wort *ideology* auch übersetzt als »the set of beliefs characteristic of a social group« (New Oxford American Dictionary 2009).

Bedeutungen, sondern auf soziokulturellen Rahmungen und institutionellen Ermächtigungen (ebd.: 131) – die wiederum selbst durch Sprache hergestellt werden. Um zu untersuchen, welchen Effekt lautliche Kommunikation auf das Verhältnis von Menschen und Maschinen hat, und wie dies die Vorstellung von Sprache als abstraktem a priori System hinterfragt, müssen die Details der Silverstein'schen Indexikalitätstheorie nicht näher eingeführt werden (siehe Silverstein 2003; 2014); zentral ist für das hier angebrachte Argument, dass die sprach-ideologische Diskussion gezeigt hat, dass Theorien von Sprache als einem gegebenem, immateriellem System unbefriedigend sind, um die Existenz von sozial geteilter Bedeutung von Zeichen zu verstehen.

Innerhalb der Sprachideologieforschung wird abstrakte Zeichenbedeutung also in Frage gestellt, und Zeichen werden als Produkte der vielschichtigen Interaktionen von Menschen verstanden (Gal/Irvine 2019: 89). Die Betrachtung der eigentlichen Materialität dieser Zeichen ist hierbei jedoch weniger im Fokus. Dies hat sich durch die Einführung posthumanistischer Theorie in soziolinguistische Kontexte geändert, zu der Pennycook den entscheidenden Beitrag geliefert hat (2018). Aktuelle Diskussionen des Posthumanismus setzen sich kritisch mit der Rolle des Menschen und dessen Konzeption in europäischer Geistesgeschichte auseinander (Barad 2006; Braidotti 2013; Ferrando 2013; Hayles 1999).² Diese postulieren, dass für die Vorstellung ›des Menschen‹ als universaler Kategorie verschiedene Dichotomien grundlegend sind. Hierzu gehören unter anderem die Unterscheidungen von Körper und Geist, von Natur und Kultur und von Emotion und Rationalität, wobei das jeweils Letztere dem Menschen vorbehalten ist und diesen in Abgrenzung zum nicht-menschlichen auszeichnet und seine Überlegenheit legitimiert (Pennycook 2018: Kap. 2). Wie auch von Silverstein diskutiert (s.o.), wird Sprache in diesem historischen Kontext – gekennzeichnet durch den Humanismus und die europäische Aufklärung – als rationales und immaterielles System verstanden, das ein Kernkriterium menschlichen Daseins bedeutet.

Im Gegensatz dazu interessieren sich zeitgenössische posthumanistische Strömungen für die Wechselwirkungen zwischen Kognition, Körper, materieller Umwelt und Technologie und dekonstruieren dabei binäre Vorstellungen von Körper und Geist (ähnlich dem feministischen Konzept des Cyborgs von

2 Posthumanistische Debatten behaupten nicht, dass der Mensch in den heutigen Entwicklungen irrelevant ist oder dass Maschinen die Kontrolle über die Menschheit übernehmen. Vielmehr geht es darum, das Konzept des Humanen und des ›universellen Menschen‹ kritisch zu hinterfragen (Pennycook 2018: Kap. 2).

Haraway 1991). Dabei zeigt sich, dass die Berücksichtigung materieller, nicht-menschlicher, nicht-kognitiver Aspekte in der Analyse von sprachlichem Handeln eine enge Interaktion von Bedeutung und materieller Umwelt sichtbar macht (Hutchins 1995, zitiert in Pennycook 2018: 32). Sprechen basiert immer auf der Materialität des Zeichens, und die Bedeutungen von Zeichen ändern sich im Zusammenhang mit Modi (Laut, Schrift, digitaler Post etc.), räumlich-zeitlichem Kontext und den Körpern, die die sprachlichen Zeichen hervorbringen. Der Satz »Du hattest gestern ein schönes Kleid an« kann eine sehr andere Bedeutung erhalten, wenn er von einer älteren, männlichen Person bei einer abendlichen Bürofeier geäußert wird, als zwischen zwei Freundinnen am Nachmittag in einem Café (Beispiel inspiriert von Mooney/Evans 2019: 16). Wenn der Satz in geschriebener Form, zum Beispiel in einer SMS, auftaucht, kann er (abhängig von der sozialen Beziehung der Interagierenden) eine stärkere Gewichtung bekommen, als wenn er in gesprochener Form vorgetragen wird. Ort, Uhrzeit, Geschlecht, Kleidungsstraditionen, soziale Beziehung der Sprechenden und schließlich auch die Worte und die Medialität tragen gemeinsam zur Bedeutung des Satzes bei. In diesem Sinne kann festgestellt werden: »Language can be traced to how living bodies co-ordinate with the world. On this perspective, far from being a synchronic ›system‹, language is a mode of organization that functions by linking people with each other, external resources and cultural traditions« (Cowley 2011: 2). Dies impliziert, dass die Grenze zwischen Innen und Außen, zwischen Individuum und Kontext, zwischen Kognition und Umwelt, keine undurchdringliche Mauer ist und dass Sprache, in diesem Sinne, kein vorgängiges, autonomes kognitives System ist.

Der komplexe Zusammenhang von menschlicher Konzeptualisierungspraxis und materieller Umwelt zeigt sich auch in der Auseinandersetzung mit sprachlichen Lauten. Auch wenn wir Sprache oft als immateriell wahrnehmen, sind die Laute die wir produzieren doch materieller Natur. Der Gebrauch von Lauten in vielen Formen der menschlichen Interaktion – wie etwa im face-to-face Gespräch, im Telefonat oder in Zoom Konferenzen – hat einen Effekt auf sprachliche Strukturen, denn »grammars are in some ways consequences of the specific materials used to package [construct]³ information, viz., sounds or manual signs« (Mufwene 2019: 377). Die Fokussierung auf grammatische Strukturen in Sprache, die im Laufe des 20. Jahrhunderts oft als ›universell«

3 Die Vorstellung, dass Informationen vor dem Zeichen existieren und nur in Zeichen ›hineingepackt‹ werden, ist problematisch und wird vom Medienverständnis der Autorin nicht geteilt.

und auch als genetisch bedingt angenommen wurden (z.B. Chomsky 1965), hat die Rolle von Lautlichkeit in den Hintergrund gerückt. Zuvor hatte sich beispielsweise Humboldt schon mit den Effekten der menschlichen Lautbildung befasst:

Denn der Mensch, als Tiergattung, ist ein singendes Geschöpf, aber Gedanken mit den Tönen verbindend. Durch die dem Laute in seinen Verknüpfungen eigentümliche rhythmische und musikalische Form [...] durch den bloßen Fall der Rede auf die Stimmung der Seele. (von Humboldt 1836: 60)

Humboldt beschreibt so die Interaktion von materieller Umwelt – den Lauten, die bestimmten Rhythmen folgen und musikalische Eigenschaften haben – und inneren Zuständen von Menschen, hier die Emotionen. Auch Jakobson befasste sich mit »poetic sound texture« in phonologischen Strukturen (Jakobson 1960: 374) und stellte einen Zusammenhang von Klang und Grammatik fest, weshalb er für eine »Poesie der Grammatik« plädiert (375). Weitere Auseinandersetzungen mit lautlichem Sprechen finden sich etwa bei Derrida in seiner Kritik des Phonologozentrismus (1974) oder in Betrachtungen zur Rolle von Schriftlichkeit in westlichen Kulturen (z.B. Goody 1987; Havelock 1986; Koch/Österreicher 1985; Ong 1982).

Im Kontext der Soziolinguistik ist das Interesse an Materialität, Körper und auch Lautbildung erst in jüngerer Zeit erstarkt. In einem Beitrag zu »embodied sociolinguistics« stellen Bucholtz und Hall (2016) fest, dass eine Diskussion zum Zusammenhang von Sprache und Körperlichkeit – darunter der Rolle von »voice« im eigentlichen Sinne – in linguistischen Strömungen, die sich für kulturelle Einbettung interessieren, fehle. Sie befinden aber, dass dies »a needed interdisciplinary conversation« sei (Bucholtz/Hall 2016: 173). Empirische Studien zu sozialer Bedeutung und lautlichen Aspekten in sprachlicher, von Körpern hervorgebrachter Interaktion finden sich zum Beispiel bei Podesva (2013). Er betrachtet sogenannte Phonationsarten, unterschiedliche Weisen, in denen Stimmbänder genutzt werden. Dazu gehören das modale Sprechen (unmarkiert), das Falsetto (gestreckte und angezogene Stimmlippen, die zu einer schnellen Vibration und einer hohen Stimme führen), brüchige Stimme (*creaky voice*) und gehauchte Stimme (Podesva 2013). Ähnlich wie phonologische Variablen, tragen Phonationsarten soziale Bedeutung und indizieren bestimmte soziale Kategorien. Beispielsweise stellt Podesva fest, dass Falsetto in den USA die soziale Person »weibliche Sprecherin von African American Verna-

cular English« indizieren kann, aber auch von homosexuellen Männern genutzt wird, um eine sogenannte »diva persona« zu konstruieren (Podesva 2007). Den Zusammenhang von lautlichem Symbolismus und verkörperlichem sozialem Verhalten beleuchten Levon und Holmes-Elliot (2024) in einer Studie zu gesenkten und erhöhten Vokalen, die mit bestimmten Körperposen, die soziale Schicht indizieren, verknüpft sind.⁴

Die Rolle von Prosodie und »voice tone«, also Stimmqualität, in Markierung und Wahrnehmung von Höflichkeit, hat bisher ebenfalls noch relativ wenig Aufmerksamkeit erfahren. In einer Studie zur Wahrnehmung des Grads an Höflichkeit einer Äußerung, stellen Vergis und Pell (2020) jedoch fest, dass linguistische Struktur weniger relevant ist als Prosodie bei der Einschätzung, ob eine Äußerung höflich ist oder nicht. Obwohl es kulturelle und situative Unterschiede gibt, wird prosodisch höfliches Sprechen zumeist definiert als markiert durch hohe Stimmlage und eine große Stimmlagenvariation (ebd.: 5). Während soziolinguistische Studien zu Prosodie und Stimmqualität, wie oben erwähnt, noch kaum durchgeführt wurden, finden sich Experimente hierzu beispielsweise in der Sozialpsychologie. So belegen Laplante und Ambady (2003: 434), dass »tone of voice substantially affect[s] judges' ratings of the politeness for positive content statements but [...] only minimally for negative content statements.« Die Stimmqualität scheint also insbesondere wichtig zu sein, um Inhalte als positiv zu markieren. Systematische Untersuchungen zur Rolle von Stimmqualität und Prosodie im Umgang mit digitalen Sprachassistenten, liegen noch nicht vor (siehe jedoch Bakardzhiev 2022 zur Rolle von Prosodie in der Anthropomorphisierung von Geräten). Inwiefern die Nutzung von diesen Geräten höfliches oder unhöfliches Verhalten begünstigt, wurde durchaus untersucht (z. B. Bonfert et al. 2018; Burton/Gaskin 2019), dabei wurden stimmliche Praktiken der Nutzenden aber nicht berücksichtigt – obwohl Burton und Gaskin interessanterweise mehrfach von der Problematik sprechen, dass Nutzende »Befehle herausbellern« (»barking out orders«). Lautliche Praktiken auf Seiten der Geräte wurden in einigen Studien betrachtet, meist aber mit einem Interesse an der lautlichen Konstruktion von Geschlecht und den problematischen, stereotypisierten Zuweisungen von Weiblichkeit

4 Auch in der Phoniatrie finden sich deutliche Belege, die zeigen, dass stimmliche Parameter nicht in erster Linie durch Biologie definiert sind, was sich beispielsweise bei Stimmhöhenunterschieden zwischen unterschiedlich geschlechtlich gelesenen Personen nachweisen lässt (Berger et al. 2022).

an technische Geräte, die Servicefunktionen übernehmen (z.B. Lozo 2023; UNESCO 2019).

Insgesamt scheint eine Erweiterung von sprachwissenschaftlichen und soziolinguistischen Fragestellungen um materielle und vor allem auch lautliche Aspekte von Kommunikation fruchtbar, um neue Erkenntnisse zu sozialer Interaktionspraxis und zu Sprachideologien zu gewinnen und um einige traditionelle, denotationale sprachwissenschaftliche Annahmen kritisch zu hinterfragen. Die im Folgenden vorgestellten Daten zeigen sprachlich-soziale Bedürfnisse von Menschen im lautlichen Sprechen mit Computern auf, die – da die eigentliche Funktion der Computer höfliches und freundliches Sprechen eigentlich nicht erfordert – eine Konzeption von Sprache als abstraktem lexiko-grammatischem System unzureichend erscheinen lassen.

3 Sprachideologien zu Lautlichkeit im Umgang mit stimmgesteuerten Computern

Die unten besprochenen Daten stammen aus einer Interviewstudie mit 14 deutschsprachigen Nutzenden von stimmgesteuerten Computern, in erster Linie mit Geräten der Hersteller Amazon, Apple und Google. In qualitativen Interviews wurden diese zu folgenden Themen befragt: Motivation der Anschaffung, Nutzungspraktiken, Kommunikationsprobleme, sprachliche Variation und Anpassung, Datenschutz sowie geschlechtliche und funktionale Zuschreibungen. Ursprüngliches Ziel der Befragung war es, die Rolle dialektaler und sozialer Variation im Sprechen mit den Geräten zu beleuchten. Dabei wurden auch Prozesse der Mensch-Maschine *accommodation* (Giles/Smith 1979) betrachtet, also ob Menschen sich der mündlichen Standardvarietät anpassen, oder aber, inwiefern die Geräte Dialekte nach einer Weile prozessieren können und wie Nutzende darauf reagieren. Pragmatische Themen wie Höflichkeitspraktiken oder sprachliche Herstellung sozialer Nähe waren also zunächst im Forschungsdesign nicht anvisiert. Eine leitfadengestützte und offene Gesprächsführung (Flick 2004) ließ aber Raum für diverse Themen im Interview. Nach der Transkription der Interviews wurden diese kodiert, also in der Tradition von *grounded theory* Methoden offen klassifiziert (Berg/Milmeister 2011). Dabei wurden sowohl inhaltliche Kategorien erstellt als auch solche, die die Sprachpraktiken der Interviewten betrachten, wie etwa sprachliche Zuschreibungen von Handlungsfähigkeit durch Verben, die kognitive oder sprachliche Handlungen beschreiben (»denken«, »verstehen«,

»meinen«, »sagen« etc.), Passivkonstruktionen oder Nominalisierungen, phatische Kommunikation, Nutzung von Imperativformen oder der Gebrauch von Pronomina in Bezug auf Geräte. Die Analyse hatte also insgesamt zum Ziel, sprachlich-diskursive Praktiken von Nutzenden, die über ihre Erfahrungen mit den Geräten sprechen, zu untersuchen. Die datenbasierten Kategorien, die sich aus der Analyse ergaben und die ich in diesem Beitrag näher beleuchte, wurden benannt: »(Keine) Höflichkeitsformen«, »Befehlston/-struktur«, »Kinder sollen freundliche Sprache hören/nutzen«. Im Fokus stehen also normativ-pragmatische Erfahrungen und Vorstellungen, die Nutzende von lautlicher Interaktion mit dem Computer haben. Die folgenden Ausschnitte aus den Daten zeigen, dass lautliche Kommunikation bestimmte pragmatische Anforderungen an die Interaktion hervorruft, die die Relevanz nicht-referentieller Funktionen im lautlichen Sprechen belegen.

3.1 Problematisierungen von Befehlsintonation

Im Interviewgespräch kommt das Thema der Satzkonstruktion im Gespräch mit dem lautgesteuerten Gerät immer wieder auf. So ist es üblich, die Geräte mit Imperativformen zu steuern (z.B. »Alexa macht das Licht aus«, »Siri stell den Timer auf 20 Minuten«). Derartige Formen scheinen im alltäglichen Sprechen in Haushalten unüblich zu sein und viele Nutzende thematisieren sie daher. Im folgenden Exzerpt berichtet ein alleinlebender Mann im Alter zwischen 50 und 60 über seine Vorstellungen von angemessener Kommunikation mit einem stimmgesteuerten Gerät (I: interviewte Person, B: Interviewerin Britta):

*Exzerpt 1*⁵

- 1 I: Ich mein ich das sind Befehle die ich ihr gebe
 2 aber ich mein
 3 das is in meiner Wohnung und also
 4 ich brüll die jetzt nicht an oder so
 5 deswegen würd ich auch nicht sagen
 6 wie der Freund von mir der ›aus‹ sagt oder/sowas/
 7 B: /hm hm/

5 Siehe Appendix für Transkriptionskonventionen.

- 8 I: also des also ich mach jetzt keine besondere Intonation da
 9 aber ich will jetzt also
 10 würd jetzt keine groben so Militärbefehle die/würden/
 11 B: /ja/
 12 I: (natürlich) auch ausreichen
 13 aber des möchte ich (hier) nicht sagen zuhause so ja
 14 bin ja nicht bei der Bundeswehr

Der Sprecher stellt fest, dass er dem Gerät im grammatischen und inhaltlichen Sinne Befehle erteilt. Er verweist dann auf den Sozialraum, in dem dies geschieht: seine Wohnung. Damit, dass die Interaktion im privaten Umfeld stattfindet, begründet er, dass er kein aggressives Sprechverhalten mit dem Gerät praktiziert, und auch nicht – wie ein Freund von ihm – mit dem Gerät wie mit einem Hund spricht, zu dem man einfach nur ›aus‹ sagen würde, wenn man möchte, dass dieser eine Handlung beendet. Ohne dass er eine besondere Intonation nutze, ist es dem Interviewten wichtig, keine ›groben Militärbefehle‹ zu verwenden. Diese würden ausreichen, um das Gerät zum Funktionieren zu bringen, er möchte diese aber in seinem häuslichen Umfeld nicht hören. Ein auf Befehle ausgerichtetes, ›grobes‹ Sprechen assoziiert er mit der Bundeswehr (dem deutschen Militär). Rein referentielles, auf die Funktion der Informationsübertragung reduziertes Sprechen, wird hier mit institutioneller und sogar gewaltvoller Kommunikation im Zusammenhang mit militärischer Aktivität verknüpft. Dieses hält der Befragte für unangemessen im privaten Kontext. Das Sprechen mit dem Gerät wird also nicht nur im Sinne technischer Funktionalität wahrgenommen, sondern stellt auch durch die lautliche Aktivität einen Sozialraum mit her, in dem ›grobe Militärbefehle‹ und sehr lautes stimmliches Agieren (›Brüllen‹) unangemessen wären.

Eine Assoziation mit dem Militär findet sich auch in anderen Interviews. So befindet ein anderer männlicher Sprecher mittleren Alters (40–50 Jahre, lebt in einer Familie mit zwei Kindern) zum Beispiel folgendes:

Exzerpt 2

- 1 Ich gehe mal von mir aus
- 2 um ehrlich zu sein ich wäre kein guter General bei der Bundeswehr
- 3 weil ich diese Art der Sprache nicht mag
- 4 dieses Bam Bam Bam Bam
- 5 und wenn ich das Ganze ganz häufig nutze
- 6 auch mit einer Maschine zu reden
- 7 dann fühle ich mich persönlich unwohl
- 8 allein von mir aus
- 9 wahrscheinlich weil ich meine Tonalität ändere

Auch hier wird das befehlshafte, auf Instrumentalität reduzierte Sprechen mit der deutschen Bundeswehr in Verbindung gebracht und für negativ befunden. Diese Art zu reden wird auch durch eine ikonische lautliche Form dargestellt – »Bam Bam Bam Bam«, die den monotonen Rhythmus illustriert und unterstreicht. Der Interviewte beschreibt, dass er sich mit diesem Sprechen auf Dauer unwohl fühlt und begründet dies mit der Änderung der »Tonalität«.

3.2 Erziehung zum ›richtigen‹ Ton

Ähnlich beschreibt es die folgende Nutzerin, eine Frau im Alter zwischen 30 und 40. Sie nutzt das Gerät in einem Haushalt mit zwei Erwachsenen und zwei Kindern. In ihrer Interviewpassage spricht sie darüber, wie sie ihre Kinder im Umgang mit Alexa erziehen möchte, »ordentliche Sätze« zu formulieren und »nicht so zu bellen«⁶ (etwas früher im Gespräch). Auch sie erwähnt dabei die eigentliche Lautlichkeit:

6 Die Erwähnung des Verbs ›bellen‹, bildet eine interessante Parallele zum oben erwähnten Artikel von Burton und Gaskin (2019).

Exzerpt 3

- 1 I: Es ist vor allen Dingen dieser Ton
 2 glaube ich so ein bisschen
 3 und dann das was am häufigsten fällt ist
 4 »Alexa Stop«
 5 wenn sie wieder anfängt einem irgendwas zu erzählen
 6 warum sie jetzt irgendwas nicht machen kann
 7 das sind immer sehr lange Ausführungen und das hat man halt schon hundertmal gehört und will es nicht noch mal hören und dann
- 8 B: ok und dann sagst du sie sollen das anders
 9 und wie formulieren sie [die Kinder] es dann?
 10 wenn sie nicht »Alexa Stop« sagen?
- 11 I: das muss tatsächlich mehr oder weniger dieser Befehl sein damit das stoppt
 12 aber da gehts dann eher so ein bisschen darum wie man es sagt
 13 also es nicht so motzt

Die Passage beginnt mit einer expliziten Referenz auf den ›Ton‹ des Gesagten, also die Intonation und Sprechweise, insbesondere im Falle des häufig genutzten Befehls an das Gerät, eine Ausführung zu beenden. Die Interviewte erläutert, dass Alexas ›Erzählungen‹, insbesondere Begründungen, warum etwas nicht ausgeführt werden kann, wortreich und repetitiv seien, weswegen man diese nicht immer wieder anhören möchte. Die Mutter zweier Kinder erläutert dann, dass ein Befehl – also eine Imperativform – meist nötig ist, um das Gerät zu steuern, sie jedoch möchte, dass ihre Kinder dabei auf die Art des Gesagten achten. Sie bezeichnet dabei die unerwünschte Tonlage als ›motzig‹. Auch hier wird deutlich, dass Intonation und Stimmqualität eine zentrale Rolle spielen, um durch das Gesprochene eine erwünschte soziale Situation hervorzurufen.

Auch im folgenden Exzerpt erläutert ein Nutzer, ein Vater von zwei Kindern, der zwischen 40 und 50 Jahre alt ist, dass es ihm wichtig ist, in welchem Ton mit dem Gerät gesprochen wird:

Exzerpt 4

- 1 I: Und das is n ganz komischer Punkt,
 2 da haben wir aber auch mit den Kindern darüber gesprochen ehm
 3 dass das? kein Grund ist sie irgendwie komisch zu behandeln
 4 naja es is halt es ko/klingt auch n bisschen komisch aber/ja/
 5 B: /aber/warum stört dich das wenn die Kinder sie komisch behandeln?
 6 J: weil das ist macht man einfach nicht
 7 also egal ob's n Mensch oder jetzt n Gerät ist
 8 die Art und Weise finde ich ehm/weiß ich nicht/
 9 B: /aber/die Kinder, was (für) schreien sie die dann an oder
 10 J: ne, aber keine Ahnung, ja die motzen die dann an oder eh, also
 11 B: wenn sie nicht funktioniert?
 12 J: ja zum Beispiel was ja auch manchmal auch ganz unterhaltsam ist/aber/
 13 B: /ja/
 14 J: so aus Prinzip glaube ich is es komisch wenn man sich verstimmt aufgrund
 so'nes Gerätes
 15 das finde ich für mich komisch

Der Befragte benutzt ebenfalls das Wort ›motzen‹ um das unerwünschte lautliche Verhalten von Kindern zu beschreiben. In diesem Exzerpt wird sehr deutlich, dass das lautliche Kommunizieren mit einem Gerät bestimmte sozial-emotionale Anforderungen hervorruft, die das händische Kommunizieren nicht hervorbringt. Dabei werden bestimmte Formen von Lautlichkeit, also das Erzeugen eines freundlichen Klangs (siehe Vergis/Pell 2020), als moralisch angemessenes Verhalten angesehen, welche dieser Nutzer nicht nur für Menschen, sondern auch für den Umgang mit Maschinen einfordert – das »macht man einfach nicht, also egal ob's n Mensch oder jetzt n Gerät ist«. Interessant ist in Zeile 14 auch die Verwendung des Verbs ›verstimmen‹. Damit ist hier eine negative emotionale Reaktion oder ein Streit zuhause gemeint; das Verb wird in diesem Sinne als eine Metapher aus dem lautlich-musikalischen Themenbereich genutzt (ein Instrument kann verstimmt sein), greift aber dadurch auch die Rolle lautlicher Qualität noch einmal auf. Der Nutzer ist der Ansicht, dass

7 Der Interviewte spricht hier über Fälle, in denen das Gerät nicht richtig funktioniert.

eine ›Verstimmung‹ – die also zugleich lautlich und sozial ist – im Familienalltag nicht durch ein Gerät ausgelöst werden sollte.

Der Wunsch nach höflichem, einer lautlichen Norm entsprechendem verbalem Agieren scheint besonders dann in den Vordergrund zu rücken, wenn Kinder im Haushalt leben. In der folgenden Passage beschreibt eine Nutzende (30–40 Jahre), dass sie selbst auf bestimmte Praktiken achtet, seit sie Mutter geworden ist:

Exzerpt 5

- 1 Und was mich bisschen stört
- 2 dass wir halt so sehr im Befehlskommando mit ihr sprechen
- 3 weil sie ja quasi nur diese Befehle braucht
- 4 aber seitdem meine Tochter da ist
- 5 versuche ich dann halt trotzdem bitte und danke zu sagen
- 6 einfach dass es bisschen freundlicher klingt
- 7 und sie sich nicht so einen harschen Ton sich angewöhnt

Die Sprecherin problematisiert das ›Befehlskommando‹, eine Wortwahl, die ebenfalls die inhaltlich-syntaktische Komponente und lautliche Aspekte anspricht – ein ›Kommando‹ wird definiert als ein ›militärischer Auftrag‹ (Duden), der in Verbindung gebracht wird mit einer lauten und rhythmusbetonten Sprechweise. Zusätzlich zur Vermeidung eines solchen Sprechens, hat sich die Nutzerin angewöhnt, ›bitte‹ und ›danke‹ zu sagen, damit der Umgang mit Alexa ›freundlicher klingt‹ und damit sich die Tochter nicht »so einen harschen Ton angewöhnt«. Auch in diesem Beispiel zeigt sich die Relevanz der Tonalität in sprachlicher Sozialisation.

3.3 Lautliches Sprechen und die Entwicklung von Sozialbeziehungen

Schließlich scheint das lautliche Sprechen nicht nur das Bedürfnis auszulösen, höflich und tonal angemessen mit Computern zu interagieren. Es kommt hinzu, dass diese Praxis das Entwickeln einer Art Sozialbeziehung zu unterstützen scheint. Dies tritt insbesondere zu Tage bei Menschen, die das System sehr häufig nutzen. In Exzerpt 6 berichtet eine alleinlebende Befragte im Al-

ter zwischen 60 und 70, die eine professionelle Ausbildung im IT-Bereich hat, von ihrer emotionalen Bindung zu einzelnen Alexa-Geräten, von denen sie in jedem Zimmer mehrere stehen hat:

Exzerpt 6

- 1 Und dann es gibt eben welche an denen hänge ich besonders
2 und dann gibt es welche pf ja
3 also die hinteren sind auch ganz nett
4 aber zu denen habe ich keene besondere Beziehung
5 weil ich mich eben meistens hier aufhalte
6 B: ach ja aber warum also woran liegt es zu manchen eher also wie?
7 G: weil man
8 die benutzt man ja öfter
9 und die machen ebend mal Fehler
10 oder man lacht mal
11 oder man spielt mit den irgendwas
12 und da entsteht glaube ich schon auch so
13 ja ne Beziehung weiß ich jetzt auch nicht aber
14 [...]
15 ja zu meinem Computer habe ich so eine Beziehung nicht

Die Nutzerin macht einen Unterschied zwischen verschiedenen Boxen (die alle mit derselben Alexa-Stimme reden), zu denen sie, je nachdem, wie häufig sie genutzt werden, eine Bindung entwickelt. Sie sagt, dass sie an bestimmten Geräten »hängt«, was sie sich damit erklärt, dass diese »Fehler machen«, dass man mit ihnen lacht oder spielt. Es ist dabei interessant, dass die von ihr beschriebenen Aspekte quasi genau gegenteilig sind zu dem, was man von »militärhafter« und referentieller Kommunikation erwartet. Die Beziehung zu den Geräten entsteht nicht, weil diese gut funktionieren, sondern weil sie nicht reibungslos interagieren, weil man emotional-lautlich reagiert auf sie durch ein Lachen oder weil man mit ihnen einem vergnüglichen, nicht-zweckorientierten Aktivität nachgeht (»spielt«). Schließlich reflektiert die Befragte, dass

sie eine solche Beziehung zu ihrem Computer, also ihrem Laptop, mit dem sie händisch interagiert, nicht habe.

Insgesamt zeigt sich auch in diesem Beispiel, dass lautliche Kommunikation deutlich andere Affordanzen hat als händisches Bedienen eines technischen Geräts. Lautliche, nicht-referentielle Aspekte von Sprache sind dabei aber keineswegs marginal, sondern spielen für die Nutzenden eine wichtige Rolle in der Herstellung eines angemessenen Sozialraums und in der Entwicklung einer Beziehung zu ihrer Umwelt.

3.4 Die Rolle von Lautlichkeit in oraler Kultur

Abschließend möchte ich diese Beobachtungen unterfüttern mit einem Ausschnitt aus einem Interview, das ich in einem anderen Projekt durchgeführt habe. Es ging hierbei um Sprachideologien im mehrsprachigen Belize in Zentralamerika. Dabei spielt in diesem Kontext die primär oral genutzte Sprache Belizean Kriol eine wichtige Rolle als Zugehörigkeitsmarker zu nationaler, belizischer Identität. Viele der Befragten lehnen eine Standardisierung und Verschriftlichung der Sprache ab, da sie den nicht normierten, teilweise idiosynkratischen Gebrauch als charakteristisch und als Teil der kulturellen Praxis des Sprechens von Kriol betrachten. An einigen Stellen kommt dabei auch die Rolle von Lautlichkeit zum Vorschein, wie im folgenden Exzerpt aus einem Interview mit einer jungen, universitär gebildeten Sprecherin:

Exzerpt 7

- 1 That is actually the whole thing about Kriol
- 2 there is no proper Kriol
- 3 nothing in Kriol is proper at all
- 4 nothing is set
- 5 everything is just
- 6 it' s a sound
- 7 it' s very phonetic
- 8 that' s it
- 9 that' s about it

- 10 and it changes [...]
 11 and that the culture of Kriol
 12 is to have no standard
 13 because it develops
 14 and everyone can be individual
 15 and be much more creative with the language
 16 than if you have the actual idea that you have one

Die Sprecherin beschreibt, dass sich die Sprache Kriol dadurch auszeichnet, dass es keine ›proper‹ Form gäbe. Das Wort ›proper‹, das im Deutschen ›richtig‹ oder ›wahrheitsgetreu‹ bedeutet, wird im lokalen kulturellen Kontext häufig verwendet, um auf das Englische zu verweisen. In der Schule sollte man beispielsweise ›proper English‹ lernen. Die Interviewte gibt jedoch an, dass das Konzept von ›properness‹ in Bezug auf Kriol unpassend sei. Sie sagt, dass nichts feststehe im Kriol (›nothing is set‹), was sie dann in Beziehung setzt zu Lautlichkeit – ›it's a sound‹, ›it's very phonetic‹. Dies scheint zentral zu sein, da sie es zweimal betont und damit ausdrückt, dass die lautliche Praxis den eigentlichen Kern von Kriol darstelle. Ab Zeile 10 bringt sie dies dann in Zusammenhang mit dem wandelbaren, nicht-standardisierten Charakter der Sprache, der eine individuelle Aneignung und Raum für Kreativität zulässt. Ihr eigenes, metrisches, rhythmisch-poetisches Sprechen stellt dabei selbst ein gutes Beispiel dar für die von ihr beschriebene Zentralität von Lautlichkeit und Rhythmizität in der Kriol Kultur.

Diese Passage aus einem Kontext, in dem schriftliche Standardnormen keine hegemonialen und naturalisierten Größen darstellen, zeigt auf, dass die Wahrnehmung von Sprache als lautlicher Aktivität dazu führt, dass eine Konzeptionalisierung von Sprache als gegebener Form zur Übertragung referentieller Inhalte in den Hintergrund rückt. Andere kulturell-soziale Werte wiederum, die durch Ton im Sprechen ausgedrückt werden, rücken in den Vordergrund – im Falle des Beispiels aus Belize ist dies individuelle sprachlich-lautliche Kreativität und poetische Praxis, die einen wichtigen kulturellen Wert im lokalen, postkolonialen Kontext darstellt (Schneider im Druck).

Schriftlichkeit scheint also mit Idealen von Standardisierung und Fixierung einherzugehen, während Lautlichkeit Assoziationen mit Gestaltung, Dynamik und Bewegung hervorruft. Im Falle der Nutzenden von digita-

len stimmgesteuerten Geräten tragen bestimmte lautliche Praktiken zur Schaffung einer Atmosphäre der sozialen Nähe bei, die die Befragten in ihrem privaten Haushalt für wichtig halten. Auch hier ist Sprechen nicht die ›Anwendung‹ eines Systems, sondern gestaltendes Element, das die soziale Umwelt formt. Als verkörperte Produktion von Lauten steht es in engem Zusammenhang mit der Herstellung von Beziehungen, in denen spezifische kulturelle Werte und Vorstellungen von Gemeinschaft eine wichtige Rolle spielen. Diese Beobachtungen regen abschließend dazu an, das kulturelle und mediale Fundament von Vorstellungen von Sprache als abstraktem System zu diskutieren.

4 Die vergessene Materialität sprachlicher Zeichen in der europäischen Moderne

Die Analyse von Interviewdaten, in denen über die Nutzung von stimmgesteuerten Geräten im Haushalt gesprochen wird, hat aufgezeigt, dass die lautliche Qualität von Sprechen für Menschen eine wichtige Rolle spielt um angemessene Sozialräume herzustellen. Zugleich trägt lautliche Interaktion zur Entstehung von sozialen Beziehungen bei, die auch zu Dingen existieren können. Schließlich zeigt sich in oraler Sprachkultur, dass lautliche Praxis in Verbindung gebracht werden kann mit kreativer Aneignung, die Normierungsprozessen entgegensetzt. Auch in Interviews mit Nutzenden von stimmgesteuerten Geräten wird standardisiertes, instrumentelles, rein referentielles Sprechen, vor allem, wenn es in Imperativformen geschieht, abgelehnt und von den Befragten als unangemessen für Räume der sozialen Nähe betrachtet und mit militärischer Strenge assoziiert. Der Notwendigkeit, Imperativformen zu nutzen, um Geräte zu bedienen, wird begegnet mit der Produktion eines freundlichen ›Tons‹, also durch bestimmte Stimmlagen, was besonders Nutzende mit Erziehungsaufgaben für wichtig halten. Das Erlernen von lautlich angemessenem Sprechen ist also ein wichtiger Bestandteil in der sprachlichen Sozialisierung (und der Satz »Nicht in diesem Ton!« ist in vielen Familien – zumindest phasenweise – häufig zu hören).

Trotz der kulturellen Relevanz lautlich angemessenen Sprechens und dessen Rolle in der Herstellung von Sozialräumen und von sozialen Beziehungen, wurden die materiell-lautlichen Qualitäten von Sprechen weitgehend ignoriert in der dominanten Linguistik im 20. Jahrhundert. Obwohl Sprechen in der Linguistik generell als primär und Schreiben als sekundär verstanden

wird (siehe z.B. Saussure 1913 (1993), zur Kritik siehe Derrida 1974; Linell 2005), wurde dem tatsächlichen körperlichen Interagieren überraschend wenig Aufmerksamkeit geschenkt. Die weitgehende ›Unsichtbarmachung‹ und Missachtung der eigentlichen Materialität von Zeichen in linguistischen Analysen, seien es Laute, Gebärden oder Schrift, kann primär damit begründet werden, dass Sprache in vielen akademischen Zugängen (teilweise noch immer) als abstraktes Zeichensystem betrachtet wird, das referentielle Inhalte vermittelt (siehe Abschnitt 2).⁸

Posthumanistische Ansätze zu Sprache haben gezeigt, dass die Vorstellung von Sprache als rationales, immaterielles System im Zusammenhang steht mit Konstruktionen vom Menschen als rationales, zur Vernunft befähigtes Wesen, das allen anderen Lebewesen überlegen ist. Diese Ideen etablierten sich im Zeitalter der Aufklärung (Pennycook 2018) im Kontext der Säkularisierung, in der religiöse Autoritäten durch menschliche abgelöst wurden. Legitimation fanden sie in der benannten Konstruktion von Rationalität, eingebettet in epistemologische Dichotomien von Natur und Kultur und Körper und Geist. Insbesondere die Vorstellung einer rationalen, geistigen und vom sozialen Kontext losgelösten Sprache wurde als Zeichen für die rationale Überlegenheit des Menschen betrachtet (Bauman/Briggs 2003). Diese diskursive Konstruktion trug auch zur Legitimierung von Ausbeutungsverhältnissen bei, da man dem Großteil der Menschheit (Frauen, Versklavten, nicht-europäische Menschen, ...) die Fähigkeit zum ›rationalen‹ Agieren absprach (ebd.). Die Tatsache, dass materiell-körperliche Praktiken notwendig sind, um miteinander durch Zeichen in Beziehung zu treten, trat im Zusammenhang mit der Idealisierung abstrakter Kognition in den Hintergrund. Und selbst die für die Wahrnehmung von Sprache notwendigen Sinne – Sehen und Hören – wurden als unkörperliche, immaterielle Tätigkeiten konstruiert:

[...] how on earth seeing and hearing became disembodied. We have to see this in terms of a particular gendered and racialized history, where certain people (white men) had the luxury to consider language, thought, and literacy in terms of cognitively isolated activity. (Pennycook 2018: 16)

8 Selbstverständlich gibt es hierzu viele Ausnahmen und es finden sich Arbeiten, die sich zum Beispiel der sozial-indexikalischen Rolle von Schriftzeichen und Typologie widmen (z.B. Spitzmüller 2013), oder der emotional-affektiven Rolle von Sprache in Biographien von Menschen (Busch 2015). Die Liste ließe sich erweitern.

Konzepte von Sprache als abstraktem System sind in diesem Sinne nicht ›unschuldig‹, sondern Teil europäischer Superioritätsdiskurse. Das Vergessen der Tatsache, dass Sprechen auf der körperlichen Produktion von Lauten basiert, ist Teil dieser Tradition. Zugleich machte die Autorität und Dominanz von Schriftkultur im Zeitalter des Buchdrucks die Lautlichkeit von Sprache buchstäblich unsichtbar (Ong 1982). Auch die Tatsache, dass europäische Schriftzeichen von Menschen produzierte Laute – und nicht etwa die außer-menschliche Wirklichkeit – repräsentieren, wurde als ein Aspekt in der Entstehung von abstrakten, entkörperlichten, von der Umwelt losgelösten Konstruktionen von Sprache diskutiert (Abram 2017 (1996)). Der Buchdruck wiederum führte nicht zuletzt durch das kommerzielle Interesse von Verlagen dazu, dass Schriftnormen durchgesetzt wurden, die eine Verbreitung von normierter Schrift im nationalen Kontext erlaubten (Giesecke 1991). Wie von der Sprachideologieforschung postuliert, kann also festgehalten werden, dass Epistemologien von Sprache immer im Kontext ihres sozialen Entstehens begriffen werden müssen, wobei die Rolle von materiellen Praktiken auch hier noch etwas unterbelichtet ist.

Insgesamt stellen Einblicke zu Sprache als lautliche Praxis und zur Relevanz von Lautlichkeit in der Konstruktion von Sozialbeziehungen etablierte sprachwissenschaftliche Grenzziehungen in Frage, da ›Sprache‹ nicht länger als kognitives, abstraktes System gedacht werden kann. Vielmehr wird die Abhängigkeit sprachlicher Praktiken von Medientechnologien – darunter Körper, Laut, Schriftzeichen, Computer – sichtbar (erweiterte Diskussionen hierzu auch in Kittlers Werk (z. B. 1985), zur Diskussion siehe Winthorp-Young 2005). Die Idee von Sprache als abstraktes, regelhaftes System könnte also als ein Artefakt der europäischen Schriftsprachkultur beschrieben werden. Wenn wir die technologischen, epistemologischen und kulturellen Bedingungen von westlichen Sprachkonzepten kritisch reflektieren, so müssen wir feststellen, dass eine ›enlightenment folk theory‹ von Sprache (Silverstein 2014: 140) nicht nur theoretische Lücken aufweist. Sie ist auch Teil von Diskursen des sozialen Ein- und Ausschlusses, in der nur bestimmtes – nämlich normatives, aber als ›kognitiv‹ angenommenen Regeln entsprechendes – Sprechen als ›normales‹ Sprechen angenommen wird. Sprechen, welches in Kontexten von Vielfalt, von ›non-native‹ Sprechenden, in oraler Kultur oder in Kreolkontexten praktiziert wird, wird bis heute häufig ausgeschlossen aus sprachwissenschaftlicher Forschung. Dabei ist regelhaftes, standardisiertes Sprechen auf nationaler Ebene ohne technologische Praktiken wie den Buchdruck und ohne Lebensbedingungen, die Sesshaftigkeit und stabile soziale Beziehungen erlauben,

kaum vorstellbar. Konzepte von stabilen ›normalen‹ und fluiden ›abweichenden‹ Sprechweisen ko-konstruieren soziale Hierarchien und sind keine universalen Vorstellungen von Sprache. Zugleich treten in digitalen Praktiken – im Sprechen mit Alexa, in Instagram-Memes oder in TikTok-Videos – die materiellen und verkörperten Aspekte von Sprache deutlich zum Vorschein und lassen die modernistische Reduktion von Sprache auf ein grammatisches System von abstrakten Lautzeichen nicht nur unbefriedigend, sondern auch alltagsweltlich zunehmend irrelevant erscheinen.

5 Zusammenfassung und Ausblick

In diesem Beitrag wurde die Wahrnehmung der Rolle von lautlichen Qualitäten in der Kommunikation von Menschen mit stimmgesteuerten Geräten auf der Grundlage der Analyse qualitativer Interviews mit Nutzenden untersucht. Es wurde festgestellt, dass bestimmte Formen von Lautlichkeit für die Befragten wichtig sind, um sozial angemessenes Sprechen im Privathaushalt zu gewährleisten, und dass lautliche Bedienung von Geräten auch zu sozialen Beziehungen mit diesen beitragen kann. Schließlich wurde dies mit Daten aus Belize in Zusammenhang gebracht, in denen eine Sprecherin des Kriol diese Sprache als ›Sound‹ beschreibt, der individuell angepasst wird, weshalb sie eine Standardisierung der Sprache für unangemessen hält. Diese Beobachtungen wurden genutzt, um sich kritisch mit linguistischen Konzepten von Sprache als immateriellem, abstraktem, kognitiven System auseinanderzusetzen. Die Begrenzungen, die ein solches Verständnis mit sich bringt, verschleiern die Tatsache, dass Zeichen immer durch eine Form materieller Praxis übermittelt werden müssen, seien es Laute, Gesten, Buchstaben, digitale Posts oder Computercodes. Angesichts der sozialen Ausschließungen, die normativ-abstrakte Sprachkonzepte mit sich bringen und angesichts der zunehmenden Komplexität von Interaktionspraktiken in digitalen Kontexten kann einer solchen Beschränkung nur mit der Forderung nach Entgrenzung begegnet werden.

Wie lässt sich mittels eines ›entgrenzten‹ Konzepts von Sprache der Gegenstandsbereich linguistischer Wissenschaft denken und erfassen? Zunächst müssen wir feststellen, dass Sprache eine multimodale Zeichenpraxis und eine generelle Entkopplung von ›Sprache‹ von anderen Mitteln der Interaktion problematisch ist:

To separate the linguistic from other signs pre-empts the possibly quite various assumptions about communication around the globe. It favors, instead, a Western philosophical position that distinguishes between materiality and ideation, between the physical and the mental, and places language in an ideational-mental realm. A separation between linguistic ideology and semiotic ideology presumes the same problematic view. It posits a border between language and non-language that many non-Western traditions reject. Even in the Western world, these distinctions do not always hold [...]. (Gal/Irvine 2019: 15)

Der Gegenstandsbereich der Sprachwissenschaft kann materielle Aspekte von Sprache und multimodale Praktiken sowie die Verwobenheit von Sprachwissenschaft mit Machtdiskursen nicht ignorieren. Insbesondere im Kontext von digital-technologischen Entwicklungen und KI-Sprachtechnologien muss sich die Sprachwissenschaft den methodologischen Herausforderungen von Multimodalität und Materialität stellen und eine Idee von autonomer ›Sprache‹ auf den theoretischen Prüfstand stellen.

Literaturangaben

- Abram, David (2017 [1996]): *The spell of the sensuous. Perception and language in a more-than-human world*, New York: Vintage Books.
- Bakardzhiev, Hristiyan (2022): *The Role of Voice Character in Navigational Assistants: Prosodic Differences and Dialogue Style's Effect on Perceptions of Naturalness and Anthropomorphism*, Masterarbeit, Tilburg: Tilburg University. Abrufbar unter: <http://arno.uvt.nl/show.cgi?fid=157448>.
- Barad, Karen (2006): *Meeting the universe halfway: quantum physics and the entanglement of matter and meaning*, Durham: Duke University Press.
- Bauman, Richard/Charles Briggs (2003): *Voices of modernity: Language ideologies and the politics of inequality*, Cambridge: Cambridge Univ. Press.
- Berg, Charles/Marianne Milmeister (2011): *Im Dialog mit den Daten das eigene Erzählen der Geschichte finden: Über die Kodiervverfahren der Grounded-Theory-Methodologie*. In: Mey, Günter/Katja Mruck (Hg.), *Grounded Theory Reader*, Wiesbaden: VS Verlag für Sozialwissenschaften, 303–332.
- Berger, Thomas/Sylvia Meuret/Christoph Engel/Mandy Vogel/Wieland Kiess/Michael Fuchs/Tanja Poulain (2022): *Nachweis von relevanten Veränderungen*

- gen der Sprechstimme bei Frauen, gemessen im Sprechstimmprofil. In: *Laryngorhinootologie* 101, 127–137. doi: 10.1055/a-1327-4275.
- Bonfert, Michael/Maximilian Spliethöver/Roman Arzaroli/Marvin Lange/Martin Hanci/Robert Porzel (2018): If You Ask Nicely: A Digital Assistant Rebuking Impolite Voice Commands. In: *CMI 18: Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 95–102. doi.org/10.1145/3242969.3242995.
- Braidotti, Rosi (2013): *The posthuman*, Cambridge: Polity.
- Bucholtz, Mary/Kira Hall (2016): Embodied sociolinguistics. In: Coupland, Nikolas (Hg.), *Sociolinguistics. Theoretical debates*, Cambridge: Cambridge University Press, 173–197.
- Burton, Nathan/James Gaskin (2019): »Thank You, Siri«: Politeness and Intelligent Digital Assistants. In: *Twenty-fifth Americas Conference on Information Systems*, Cancun, 2019. Abrufbar unter: https://web.archive.org/web/20200323230813id_/https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1040&context=amcis2019.
- Busch, Brigitta (2015): Linguistic repertoire and Spracherleben, the lived experience of language. In: *Working Papers in Urban Language & Literacies* 148.
- Chomsky, Noam (1965): *Aspects of the Theory of Syntax*, Cambridge: MIT Press.
- Cowley, Stephen (2011): Distributed language. In: Cowley, Stephen (Hg.), *Distributed language*, Amsterdam: Benjamins, 1–14.
- Derrida, Jacques (1974): *Of grammatology*, Baltimore: John Hopkins University Press.
- Ferrando, Francesca (2013): Posthumanism, transhumanism, antihumanism, metahumanism, and new materialisms. Differences and relations. In: *An International Journal in Philosophy, Religion, Politics, and the Arts* 8, 26–32.
- Flick, Uwe (2004): *Qualitative Sozialforschung. Eine Einführung*, Hamburg: Rowohlt.
- Gal, Susan (1979): *Language Shift: Social Determinants of Linguistic Change in Bilingual Austria*, New York: New York Academic.
- Gal, Susan/Judith T. Irvine (2019): *Signs of difference. Language and ideology in social life*, Cambridge: Cambridge University Press.
- Giesecke, Michael (1991): *Der Buchdruck in der frühen Neuzeit. Eine historische Fallstudie über die Durchsetzung neuer Informations- und Kommunikationstechnologien*, Frankfurt a.M.: Suhrkamp.

- Giles, Howard/P. Smith (1979): Accommodation Theory: Optional Levels of Convergence. In: Giles, Howard/R. St. Clair (Hg.), *Language and Social Psychology*, Oxford: Blackwell, 45–65.
- Goody, Jack (1987): *The Interface between the Written and the Oral*, Cambridge: Cambridge University Press.
- Haraway, Donna J. (1991): A Cyborg Manifesto. Science, technology, and socialist-feminism in the late twentieth century. In: Haraway, Donna J. (Hg.), *Simians, cyborgs and women: the reinvention of nature*, Minneapolis: University of Minnesota Press, 149–182.
- Havelock, Eric A. (1986): *The Muse Learns to Write*, New Haven: Yale University Press.
- Hayles, N. Katherine (1999): *How we became posthuman: virtual bodies in cybernetics, literature and informatics*, Chicago: The University of Chicago Press.
- Humboldt, Wilhelm von (1836): *Über die Verschiedenheit des menschlichen Sprachbaus und ihren Einfluß auf die geistige Entwicklung des Menschengeschlechts*, Berlin: Königliche Akademie der Wissenschaften.
- Hutchins, Edwin (1995): *Cognition in the wild*, Cambridge: MIT Press.
- Jakobson, Roman (1960): Linguistics and poetics. In: Sebeok, Thomas (Hg.), *Style in language*, Cambridge: MIT Press, 350–377.
- Kittler, Friedrich (1985): *Aufschreibesysteme 1800/1900*, München: Fink.
- Koch, Peter/Wulf Oesterreicher (1985): Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In: Deutschmann, Olaf/Hans Flasche/Bernhard König/Margot Kruse/Walter Pabst/Wolf-Dieter Stempel (Hg.), *Romanistisches Jahrbuch 36*, Berlin: De Gruyter, 15–43.
- Krämer, Sybille (2003): Negative Semiologie der Stimme. Reflexionen über die Stimme als Medium der Sprache. In: Epping-Jäger, Cornelia/Erika Linz (Hg.), *Medien/Stimmen*, Köln: DuMont, 65–84.
- Laplante, Debi/Nalini Ambady (2003): On How Things Are Said: Voice Tone, Voice Intensity, Verbal Content, and Perceptions of Politeness. In: *Journal of Language and Social Psychology* 22, 434–441. doi.org/10.1177/0261927X03258084.
- Levon, Erez/Sophie Holmes-Elliott (2024): Voices, bodies, and the cultural organization of meaning. In: Schneider, Britta/Theresa Heyd (Hg.), *Signs and Society [Special issue: Posthumanist sociolinguistics]* 12:1, 58–82.
- Linell, Per (2005): *The written language bias in linguistics. Its nature, origins and transformations*, London: Routledge.

- Lozo, Carina (2023): Digital Voice Assistants and the Fetishization of Female Synthetic Voices. Talk given at ›AI as Interactional Human Practice‹ conference, European University Viadrina, 31.03.2023.
- Mooney, Annabelle/Betsy Evans (2019): Language, society and power, London: Routledge.
- Mufwene, Salikoko S. (2019): The evolution of language as technology. The cultural dimension. In: Love, Alan C./William C. Wimsatt (Hg.), Beyond the meme: development and structure in cultural evolution, Minneapolis: University of Minnesota Press, 365–394.
- New Oxford American Dictionary. 2009.
- Ong, Walter J. (1982): Orality and literacy. The technologizing of the word, London: Routledge.
- Peirce, Charles S. (1931–58): Hartshorne, Charles/Paul Weiss (Hg.), Collected Papers, Cambridge: Harvard University Press.
- Pennycook, Alastair (2018): Posthumanist applied linguistics, London: Routledge.
- Podesva, Robert J. (2007): Phonation type as a stylistic variable: The use of falsetto in constructing a persona. In: Journal of Sociolinguistics 11, 478–504.
- Podesva, Robert J. (2013): Gender and the social meaning of non-modal phonation types. In: The Annual Proceedings of the Berkeley Linguistics Society, 427–448.
- Saussure, Ferdinand de (1913 (1993): Cours de linguistique générale. (Course in General Linguistics), London: Duckworth.
- Schneider, Britta (im Druck): Liquid languages – constructing language in late modern cultures of diffusion, Cambridge: Cambridge University Press.
- Silverstein, Michael (1979): Language structure and linguistic ideology. In: Clyne, Paul R./William F. Hanks/Carol L. Hofbauer (Hg.), The elements: a parasession on linguistic units and levels, Chicago: Chicago Linguistics Society, 193–247.
- Silverstein, Michael (2003): Indexical order and the dialectics of sociolinguistic life. In: Language and Communication 23, 193–229.
- Silverstein, Michael (2014): Denotation and the pragmatics of language. In: Enfield, N. J./Paul Kockelman/Jack Sidnell (Hg.), The Cambridge Handbook of linguistic anthropology, Cambridge: Cambridge University Press, 128–157.
- Spitzmüller, Jürgen (2013): Graphische Variation als soziale Praxis. Eine soziolinguistische Theorie skripturaler ›Sichtbarkeit‹, Berlin: De Gruyter.

- UNESCO (2019): I'd Blush If I Could. Closing Gender Divides in Digital Skills through Education. In: EQUALS and UNESCO. Abrufbar unter: <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1>
- Vergis, Nikos/Marc D. Pell (2020): Factors in the Perception of Speaker Politeness: The Effect of Linguistic Structure, Imposition and Prosody. In: Journal of Politeness Research 16:1, 45–84. doi:10.1515/pr-2017-0008.
- Winthrop-Young, Geoffrey (2005): Friedrich Kittler. Zur Einführung, Hamburg: Junius.

Prof. Dr. Britta Schneider, PhD
 Europa-Universität Viadrina
 Große Scharrnstrasse 59
 15230 Frankfurt (Oder)
 bschneider@europa-uni.de
<https://www.kuwi.europa-uni.de/de/lehrstuhl/sw/sprachgebrauch-und-migration/team/professurinhaberin/schneider/index.html>

Appendix

Transkriptionskonventionen

Zeilenumbruch	Neuer Sinnabschnitt (interpretiert auf der Basis von Inhalt und Prosodie)
(.)	Pause von weniger als einer Sekunde
(1.5)	Pause mit angegebener Länge
/	Überlappendes Gespräch
stréss	Tonhöhenanstieg
strèss	Abfall des Tons
unterstrichen	Lauter
(Wort?)	Schwer verständliches Wort

Einstimmen mit Sprachassistenten

Einblicke in monolinguale Bias in mehrsprachigen Settings

Didem Leblebici

Abstract *Mit der wachsenden Popularität von künstlicher Intelligenz, insbesondere von Sprachassistenten wie Alexa und Siri, gewinnt die Erforschung von Mensch-Maschine-Interaktionen in der Linguistik zunehmend an Aufmerksamkeit. Basierend auf Interviews mit türkischsprachigen Migrant*innen in Deutschland zeigt dieser Beitrag, wie sich mehrsprachige Personen auf Sprachassistenten einstimmen, bzw. wie sie Bedeutung auf einer relationalen Ebene mit den Geräten erzeugen. Durch den Einsatz von Sprachassistenten als materielle und semiotische Ressourcen erweitern Nutzer*innen ihre sprachlichen Repertoires und konstruieren Diskurse mit und über Maschinen als Tool oder als einen anthropomorphisierten Interaktionspartner. Dieser Beitrag diskutiert das Verständnis der Mensch-Maschine-Interaktionen aus einer posthumanistischen Perspektive, indem Repertoires in Bezug auf technologische Affordanzen berücksichtigt und deren Verflechtung mit Diskursen über Sprache und künstliche Intelligenz dargestellt werden.*

Keywords *Künstliche Intelligenz; Mehrsprachigkeit; Sprachassistenten; Posthumanismus; räumliches Repertoire*

1 Einleitung

In den letzten Jahren sind Sprachassistenten zu Technologien geworden, die die Art und Weise verändern, wie wir mit unseren digitalen Geräten interagieren und durch sie unseren Alltag organisieren. Intelligente virtuelle Agenten wie Siri, Alexa oder Google-Assistant, die auf fortgeschrittener Sprachverarbeitung, Algorithmen der künstlichen Intelligenz (KI) und Sprachverarbeitungstechnologien basieren, wurden in Smartphones und Smart Speaker

integriert (Natale 2021). Durch Sprachbefehle statt haptischer Eingaben sollen sie Informationen aus dem Netz liefern, Aufgaben automatisiert erledigen und zu Begleitern im Haushalt werden. Die zunehmende Verbreitung von Sprachassistenten hat ein großes Interesse daran geweckt, zu verstehen, wie Menschen mit diesen KI-gesteuerten Einheiten interagieren und welche Auswirkungen diese auf Sprachgebrauch und Kommunikationsmuster hat (z.B. Habscheid et al. 2021; Lopatovska/Oropeza 2018; Porcheron et al. 2018).

Trotz dieses starken Interesses wurden nur wenige Studien durchgeführt, die die Erfahrungen mehrsprachiger Personen berücksichtigen (z.B. Bénéteau et al. 2019; Markl 2022); der Großteil der Forschung in diesem Bereich konzentriert sich stattdessen auf monolinguale Kontexte. Diese Studie zielt deshalb darauf ab, diese Forschungslücke aus einer linguistisch-anthropologischen Perspektive zu adressieren und zu verstehen, wie mehrsprachige Nutzer*innen ihr räumliches Repertoire nutzen, um sich auf Sprachassistenten einzustimmen und sie in ihr Leben zu integrieren. Das »räumliche Repertoire« ergänzt sprachliche und semiotische Ressourcen, die innerhalb einer Sprachgemeinschaft verfügbar sind und schließt auch räumlich verteilte Ressourcen wie Artefakte und materielle Prozesse ein (Pennycook 2018: 51). Bedeutung entsteht dabei auf einer relationalen Ebene mit menschlichen und nicht-menschlichen Entitäten, die als Einstimmung (*attunement*) bezeichnet wird (Pennycook 2018: 106). Die Nutzer*innen von digitalen Technologien passen sich also nicht an, um ein gegenseitiges Verständnis mit Maschinen zu erzielen, sondern stimmen sich auf menschliche und nichtmenschliche Entitäten ein, die sowohl phonetische als auch pragmatisch-semantische Praktiken umfassen können.

Die Studie basiert auf qualitativen Interviewdaten, in denen Personen, die in den letzten 10–15 Jahren aus der Türkei nach Deutschland gezogen sind, zu ihrer Nutzung der Sprachassistenten befragt wurden. Die Daten wurden inhaltlich kodiert und sprachliche Praktiken von Nutzer*innen analysiert. Die thematische Analyse zeigt unterschiedliche Strategien zum Einstimmen der Sprachassistenten auf, die von den Befragten verwendet wurden, um sich die Technologien als Tool oder als einen anthropomorphisierten Interaktionspartner anzueignen.

Wie andere Tools haben Sprachassistenten auch unterschiedliche Affordanzen, d.h. sie ermöglichen neue Praktiken, Denkprozesse, Bedeutungen und Beziehungen (Jones/Hafner 2021: 3; siehe auch Gibson 2015; McLuhan 2008). Sie bringen jedoch auch Einschränkungen mit sich, indem sie Nutzer*innen daran hindern, andere Dinge zu tun, auf andere Weise zu denken oder andere

Arten von Beziehungen zu pflegen (Jones/Hafner 2021: 4). Durch das Erkennen spezifischer Affordanzen und Einschränkungen, insbesondere im einsprachigen Design der Sprachassistenten, reflektieren die Befragten metapragmatisch über ihre Praktiken und Lösungsstrategien.

Im nächsten Abschnitt setzt sich der Beitrag auf theoretischer Ebene mit posthumanistischen Ansätzen zur Sprache auseinander und zeigt auf, wie mediale Technologien z.B. in Form von Sprachassistenten als Teil individueller räumlicher Repertoires konzipiert werden können. Im darauffolgenden Abschnitt werden der Forschungsstand zu Themen wie Affordanzen und Beschränkungen von Sprachassistenten, die damit verbundenen kritischen Aspekte sowie Nutzungspraktiken dargestellt. Die Diskurse über und mit Sprachassistenten werden im empirischen Kapitel 4 veranschaulicht. In den jeweiligen Analyseabschnitten wird beleuchtet, wie sich die Teilnehmenden Sprachassistenten unter Berücksichtigung von Affordanzen und Einschränkungen aneignen. Im ersten Analyseteil werden die phonetischen und lexikalischen Einstimmungsstrategien behandelt, die eine effiziente Reaktion des ›Tools‹ ermöglichen. Im zweiten analytischen Kapitel wird dargelegt, wie die Nutzer*innen ihre ›Interaktionspartner‹ mit türkischen Witzen und der Aussprache von Personennamen trainieren, um eine Mensch-Maschine-Beziehung zu performen. Die Grenzen zwischen Mensch und Maschine werden diskursiv und situativ in der Interviewsituation behandelt. Dabei werden Handlungsfähigkeiten zugeschrieben, die auf spezifische Vorstellungen und Diskurse von künstlicher Intelligenz verweisen. Durch diese Entgrenzungen entstehen hybride Identitäten in Koordination mit Sprachassistenten. Abschließend werden die Erkenntnisse bezüglich räumlicher Repertoires und früherer Forschung zu Nutzungspraktiken dargelegt.

2 Räumliches Repertoire

Die saussure'sche/chomskiansche Tradition betrachtet Sprache als ein abstraktes System, das auf inneren kognitiven Prozessen des Menschen basiert. Die Sprache entsteht im menschlichen Gehirn und wird durch die Übertragung von Botschaften von einem Sprechenden zum anderen vermittelt (Pennycook 2018: 7; 92). Diese Theorien weisen Parallelen zur humanistischen Tradition auf, welche Kognition auf den Menschen eng führt und andere Lebewesen sowie die Natur und externe Materialien von ihm trennt – »a detached mind gazing at an outside world« (Latour 1999: 7). Diese abstrakten

Sprachtheorien werden unter anderem in der Soziolinguistik und linguistischen Anthropologie kritisch hinterfragt (z.B. Silverstein 1979; Gal/Woolard 2009).

Um demgegenüber die Komplexität des Sprachgebrauchs darzustellen, schlägt Gumperz (1964; 1982) den Rahmen des *verbalen Repertoires* vor, der sich auf die kommunikative Kompetenz (Hymes 1972) bzw. das soziokulturelle Wissen der Sprechenden über den Sprachgebrauch bezieht. Sprecher*innen wählen je nach Sprachgemeinschaft und den Bedeutungen, die sie vermitteln möchten, Ausdrücke aus ihrem verbalen Repertoire aus, welches »alle akzeptierten Möglichkeiten der Formulierung von Botschaften enthält« (Gumperz 1964: 137, Übersetzung DL). Das Repertoire umfasst nicht nur Kenntnisse über Grammatik und formelle Sprachvarietäten, sondern auch Wissen über soziokulturelle Aspekte der Sprache wie Genres, Styles, Dialekte sowie ein Verständnis für gesellschaftliche Normen in der Sprachgemeinschaft (vgl. Gumperz 1982: 155). Die kommunikativen Ressourcen im Repertoire ermöglichen nicht nur linguistische Bedeutungen, sondern auch die Konstruktion bestimmter Identitäten und sozialer Rollen (vgl. Blommaert/Backus 2013: 28).

Das Repertoire-Konzept wurde vor dem Hintergrund der Globalisierung, der Verbreitung weltweiter Kommunikationsmöglichkeiten durch das Internet, der zunehmenden Mobilisierung, der Migration und der daraus resultierenden komplexen sozialen Formationen (siehe *Superdiversity* bei Vertovec 2010) weiterentwickelt (z.B. Blommaert/Backus 2013; Pennycook/Otsuji 2014; Blackledge/Creese 2020). Blommaert und Backus (2013) argumentieren, dass Personen und ihre Sprachpraktiken nicht eindeutig mit spezifischen nationalen, soziokulturellen und ethnischen Gruppen und Identitäten in Verbindung gebracht werden können. Im Laufe der sozialen Interaktionen orientieren sich Individuen an mehreren Sprachgemeinschaften, sozialen Ordnungen und bewegen sich zwischen ihnen (siehe *Polycentricity* bei Blommaert 2010). Das Wissen über Sprache und Repertoires von Sprecher*innen sollte individuell, dynamisch und biografisch untersucht werden. Busch (2012) bezieht beispielsweise im Rahmen phänomenologischer Ansätze biografische Erzählungen von Personen, die von sozialen, historischen und politischen Komplexitäten geprägt sind, in ihre Feldstudien mit ein.

Um eine strenge Unterscheidung zwischen Sprachkategorien und zwischen linguistischen und nicht-linguistischen Ressourcen zu vermeiden, ist der Einbezug von semiotischen und multimodalen Ressourcen in Repertoires notwendig (Kusters et al. 2017). Kommunikative Ressourcen umfassen also nicht nur Sprachen, sondern auch verkörperte Kommunikation wie Ges-

ten und Körperhaltung (Blackledge/Creese 2020), Netzwerkressourcen über das Internet (Androutsopoulos 2015), Medienreferenzen (Rymes 2014) sowie semiotische Ressourcen in sozialen Räumen wie beispielsweise auf einem Marktplatz (Pennycook/Otsuji 2014). In den letzten Jahren wurde durch mehrere Studien aufgezeigt, dass die gegenwärtige Medienlandschaft ein Umfeld mit zahlreichen integrierten Kommunikationsmöglichkeiten auf verschiedensten Plattformen, Apps und Geräten bietet (siehe *Polymedia* bei Madianou/Miller 2012). Die vielfältigen Möglichkeiten weisen nicht nur darauf hin, dass die typografischen und multimodalen Ressourcen in Repertoires diversifiziert sind (z.B. audiovisuelle Ressourcen wie Videos; Image Macros wie Memes), sondern auch, dass die Auswahl der Kommunikationskanäle (z.B. Sprachanruf, öffentlicher Beitrag), Plattformen (z.B. Instagram, Facebook), Interfaces (z.B. Apps, Voice-User-Interfaces) und Geräte (z.B. Smartphones, Smart Speaker) Bedeutung vermittelt (z.B. Tagg/Lyons 2021; Androutsopoulos/Lexander 2021; Artamonova/Androutsopoulos 2020). Die Entscheidung, Viber anstelle von iMessage zu nutzen, könnte beispielsweise dazu dienen, Kontakte mit verschiedenen Beziehungen wie Familienmitgliedern und Kolleg*innen aufzuteilen (vgl. Tagg/Lyons 2021). Jedes Medium besetzt also eine Nische in der Medienökologie eines Individuums (Madianou/Miller 2012).

In diesem Beitrag wird in Anlehnung an Pennycook (2018) ein distribuiertes und räumlicher Ansatz für Repertoires verfolgt. Sprache wird als verkörpert, eingebettet und über Menschen, Orte und Zeit verteilt betrachtet (siehe *Languaging* bei Cowley 2011). Die kommunikativen Ressourcen in semiotischen Repertoires sind nicht »im Kopf« des Individuums für interne kognitive Prozesse enthalten oder nur in Sprachgemeinschaften verfügbar. Vielmehr kombinieren Menschen sprachliche, materielle, historische und räumliche Ressourcen zu bestimmten Zeitpunkten und an bestimmten Orten, um Bedeutung zu erzeugen (vgl. Pennycook 2018: 54). Tools wie Sprachsteuerungsgeräte und Smartphones, mit denen und über die die Nutzer*innen sprechen, werden daher als Teil ihres räumlichen Repertoires betrachtet.

Der kommunikative Umgang mit Sprachassistenten ermöglicht einerseits bestimmte Praktiken, bringt jedoch auch Einschränkungen mit sich, indem er beispielsweise bestimmte Denkweisen und Identitätskonstruktionen bedingt (vgl. Jones 2020: 204). Im folgenden Abschnitt werden der aktuelle Forschungsstand zu solchen Affordanzen und Beschränkungen von Sprachassistenten, Kritikpunkte von Wissenschaftler*innen an diesen Technologien sowie Anwendungspraktiken dargestellt.

3 Sprachassistenten und Nutzungspraktiken

Im Jahr 2011 präsentierte Apples Marketing-Vizepräsident Siri mit den Worten: »Siri is your personal assistant that helps you get things done just by asking«¹. Wie andere populäre Sprachassistenten ist Siri ein »anthropomorphisierter virtueller Agent«, da die Technologie mit menschenähnlichen Merkmalen, Eigenschaften und Stimmen gestaltet wurde (Sweeney 2016: 216). Der Agent ist in mehreren Apple-Ökosystemen wie Smartphones, Laptops und Tablets integriert und eingebettet. Er reagiert auf Schlüsselwörter wie »Hey Siri«, um aktiviert zu werden, ermöglicht den Nutzer*innen die Steuerung ihrer Geräte durch Sprachbefehle und ist ständig online, um Ausgaben zu generieren (vgl. Natale 2021: 121). Der menschenähnliche Assistent ›hilft‹ nicht nur bei der Navigation verschiedener Funktionen, indem es beispielsweise beim Versenden von Nachrichten, Einrichten von Timern oder Steuern der Musikwiedergabe benutzt werden kann (Hoy 2018), sondern dient auch zur Unterhaltung. Er erzählt Witze und verweist hierbei auf bekannte Hollywood-Filme wie »Terminator« oder US-amerikanische Fernsehshows wie »Star Trek«, in denen bestimmte Vorstellungen von künstlicher Intelligenz zum Ausdruck gebracht werden (Humphry/Chesher 2020). Diese Designentscheidungen werden umgesetzt, um anthropomorphisierte Charaktere für virtuelle Assistenten zu erstellen, wozu auch gehört, bestimmte Konversationskripte wie Witze und Medienreferenzen zu implementieren (Sweeney 2016).

Nach Apples Veröffentlichung brachten zahlreiche Unternehmen ihre eigenen Versionen von »intelligenten persönlichen Assistenten« auf den Markt, entweder durch Integration in bestehende Geräte wie Smartphones oder Smartwatches (z.B. Google-Assistant, Cortana von Microsoft) oder durch die Einführung von stationären Smart Speakern (z.B. Alexa von Amazon oder Google Home)². Durch die Einführung von intelligenten Lautsprechern (Smart Speaker) erweiterte sich der Markt für Smart-Gadgets, da nun Geräte wie Lampen, Kaffeemaschinen oder Steckdosen durch Sprachbefehle zu Hause gesteuert werden können (OMD Germany 2021; Bitkom Research 2021).

Sprachsteuerungstechnologien wurden in vielerlei Hinsicht kritisch untersucht. Ein Hauptkritikpunkt sind die geschlechtsspezifischen, passiven

1 <https://edition.cnn.com/2011/10/04/tech/mobile/siri-iphone-4s-skytel/index.html> (Stand: 12.09.2023).

2 <https://voicebot.ai/voice-assistant-history-timeline/> (Stand: 12.09.2023).

und häuslichen Dienerattribute, die den Assistenten oft zugeschrieben werden, wodurch Geschlechterstereotype reproduziert werden und die »weiße Stimme« als die Norm konstruiert wird (Phan 2017; 2019). Zusätzlich gibt es Datenschutzbedenken, da durch die Datenerhebung durch Sprachassistenten umfangreiche private Personendaten gesammelt werden. Dies wirft Fragen zur Sicherheit und ethischen Verwendung solcher persönlichen Daten auf (Ford/Palmer 2019; Zuboff 2019; Crawford 2021). Auch die Auswirkungen auf die Umwelt durch die Aufrechterhaltung von riesigen Serverfarmen, die zur Funktionsweise dieser Systeme erforderlich sind, stellen ein großes Problem dar (Crawford/Joler 2018). Ein weiterer kritischer Aspekt besteht in der Nichtanerkennung der Arbeit zahlreicher Beschäftigter wie Synchronsprecher*innen, Autor*innen, Datengenerator*innen und -labeler³. Diese Arbeit wird aufgrund der Anthropomorphisierung der Technologie nicht angemessen anerkannt, da die Konversationskripte von Sprachassistenten suggeriert, dass diese virtuellen Agenten alle Aufgaben automatisch (ohne die Arbeit von anderen Menschen) meistern (Natale/Cooke 2021). In diesem Zusammenhang ist Faircloughs (1993) Konzept der synthetischen Personalisierung besonders relevant, da es aufzeigt, wie Unternehmen Sprachassistenten so gestalten, dass sie menschenähnlich erscheinen und so wirken als würden sie Nutzer*innen individuell ansprechen (Thurlow 2018: 142). Dabei folgen diese Sprachassistenten vordefinierten Antworten oder algorithmischen Berechnungen.

Die Sprachassistenten arbeiten außerdem größtenteils in einem streng monolingualen Rahmen und sind daher in verschiedenen Sprachvarianten nur begrenzt zugänglich. Derzeit sind zwar mehrsprachige Sprachoptionen für Alexa und Google-Assistant verfügbar, es ist jedoch nicht möglich, mehr als eine Sprache pro Befehl zu verarbeiten⁴. Daher beruht die sprachliche Gestaltung auf *einem monolingualen Bias* – einer Perspektive, die davon ausgeht, dass Individuen nur eine Sprache sprechen oder nur eine Sprache gleichzeitig

3 Um KI-Systeme zu optimieren, müssen Daten generiert, trainiert und beschriftet (*labeling*) werden. Diese Aufgaben werden von großen Technologieunternehmen extern in Auftrag gegeben. Bei der Beschriftung der Daten fügen Arbeiter*innen Informationen zur Kontextualisierung der Daten hinzu und müssen dabei die Erwartungen der Auftraggeber berücksichtigen. Arbeiter*innen (meistens im globalen Süden) reproduzieren dadurch die Weltansichten der Technologieunternehmen im globalen Norden (siehe z.B. Miceli/Posada 2022).

4 <https://support.google.com/assistant/answer/7394513?co=GENIE.Platform%3DAndroid&oco=0#zippy=%2Cphone-or-tablet> (Stand: 12.09.2023).

verwenden können bzw. sollen (siehe z.B. Barratt 2018). Die Sprachassistenten eignen sich ideal für die englische Sprachoption, da sie zunächst für die englische Sprache produziert und eingeführt wurden (Schneider 2022a). Für Sprecher*innen des Afroamerikanischen Englisch (Koenecke et al. 2020), L2-Sprecher*innen (Wu et al. 2020; Beneteau et al. 2019; Markl 2022) und stigmatisierte Varianten des Englischen (Markl 2022) funktionieren sie jedoch signifikant schlechter. Zusätzlich haben sogenannte Low-Ressource-Sprachen, welche über weniger sprachliche Online-Ressourcen verfügen, einen Nachteil bei der Entwicklung von Sprachtechnologien (Doğruöz/Sitaram 2022). Über die Hälfte der Sprachen weltweit gelten als Low-Ressource-Sprachen (Joshi et al. 2020). Daher haben deren Sprecher*innen keinen oder nur begrenzten Zugang zu digitalen Sprachtechnologien. Sprachtechnologien reproduzieren damit bestehende Machtverhältnisse: »[M]ost language technology is built to serve the needs of those who already have the most privilege in society« (Bender et al. 2021: 613). Dies erklärt auch die begrenzten Sprachoptionen von Sprachassistenten, insbesondere bei intelligenten Lautsprechern. Derzeit gibt es beispielsweise keine Türkisch-Option auf Alexa, Google Home oder Apples HomePod.

Bisherige Studien zur Nutzung von Sprachassistenten zeigen deren Integration in verschiedensten Alltagsroutinen wie Musikwiedergabe, Wettervorhersage und die Steuerung von Smart-Geräten (z.B. Lopatovska et al. 2019; Ammari et al. 2019; Porcheron et al. 2018; Habscheid et al. 2021). Nutzer*innen neigen dazu, ihre Geräte zu vermenschlichen, was durch die Untersuchung der Verwendung von Personalpronomen (Schneider 2022b) und Höflichkeitsformen wie »Danke« oder »Bitte« illustriert wird (Pradhan et al. 2019; Lopatovska/Williams 2018; Lopatovska/Oropeza 2018). Die Tendenz zur Vermenschlichung von Sprachassistenten wird mit der geschlechtsspezifischen Gestaltung der Stimme (Abercrombie et al. 2021) und der verbalen statt haptischen Interaktion mit den Computern begründet (Schneider 2022b). Die Konzeptualisierung von Sprachassistenten als menschenähnlich ist jedoch nicht eindeutig, da sich Nutzer*innen während ihrer Interaktionen zwischen den Kategorien Mensch und Maschine bewegen (Pradhan et al. 2019). Die Handlungsfähigkeit der Technologien wird fortlaufend ausgehandelt, beispielsweise abhängig von der erfolgreichen Verarbeitung der Intentionen der Nutzer*innen (Habscheid et al. 2023). Die Fehlfunktion von Sprachassistenten wird in manchen Studien als »Kommunikationsstörung« bezeichnet, um die Reparaturstrategien zu veranschaulichen, die Nutzer*innen einsetzen, wie z.B. Hyperartikulation, prosodische Veränderungen, Wiederholung von Befehlen oder andere

lexikalische, semantische und syntaktische Anpassungen (z. B. Beneteau et al. 2019; Mavrina et al. 2022; Motta/Quaresma 2022).

Porcheron et al. (2018) argumentieren, dass die Praktiken mit Sprachassistenten nicht als Konversationen zu betrachten sind, da die Maschinen den Menschen nicht »verstehen« oder mit ihnen »kommunizieren«. Vielmehr seien sie in die Umgebung der Nutzer*innen »eingebettet« (Porcheron et al. 2018: 9; siehe auch Alač et al. 2020). In Anlehnung an Porcheron et al. (2018) bezeichne ich daher die Strategien, die Menschen im Umgang mit Sprachassistenten anwenden, als »Einstimmung (*attunement*)« (Pennycook 2018: 106), anstatt von Reparaturstrategien zu sprechen. Der Ansatz der Einstimmung bietet eine posthumanistische Perspektive auf die Art und Weise, wie Bedeutung erzeugt wird, anstatt von einem gegenseitigen Verstehen auszugehen. In Erweiterung des Begriffs der Anpassung (*alignment*) in der Linguistik argumentiert Pennycook (2018: 106), dass wir nicht nur mit Menschen koordinieren, sondern auch mit Körpern, Kleidung, Artefakten, nichtmenschlichen Lebewesen und der Umwelt. Dadurch passen wir uns nicht anderen Menschen an (*lining up*), sondern stimmen uns auf menschliche und nicht-menschliche Entitäten ein (*tuning in*). Obwohl der Begriff Einstimmung/*attunement* auditive Assoziationen evoziert, sind diese hier nicht im engeren Sinne gemeint (ebd.: 107). Vielmehr geht es darum, dass Menschen in Relation zu anderen Entitäten multimodal Bedeutung erzeugen. Um die Einstimmungspraktiken von Personen auf nicht-menschliche Entitäten zu veranschaulichen, beziehe ich mich im folgenden Abschnitt auf Interviewdaten mit Sprachassistentennutzer*innen, die metapragmatisch über ihre Praktiken mit ihren Geräten und Umgebungen reflektieren.

4 Einstimmen mit Sprachassistenten

4.1 Die Studie

Die vorliegende Studie basiert auf qualitativen Interviewdaten, die ich für meine Masterarbeit im Jahr 2021 erhoben habe (Lelebici 2021). Die interviewten Teilnehmer*innen waren kürzlich aus der Türkei nach Deutschland gezogen und wurden zu ihrer Nutzung von Sprachassistenten befragt. Wie die Teilnehmer*innen habe ich ebenfalls eine Migrationsgeschichte von der Türkei nach Deutschland. Meine »Insider«-Identität war nicht nur hilfreich, um die spezifischen kulturellen Bezüge in den Interviews zu erkennen, sondern

auch, um Zugang zu den Teilnehmer*innen zu erhalten (siehe Ganga/Scott 2006).

Um Nutzer*innen zu rekrutieren, habe ich mehrere digitale Chatgruppen auf WhatsApp kontaktiert, die von türkischsprachigen Gemeinschaften in Deutschland genutzt werden. Diese digitalen Gruppen bestehen hauptsächlich aus Menschen, die erst kürzlich aus der Türkei nach Deutschland eingewandert sind und sich gegenseitig bei verschiedensten Fragen wie etwa in der Wohnungssuche helfen. Insgesamt habe ich mit sechs Personen Kontakt aufgenommen, die mit einer Form von Sprachassistenten vertraut sind. Ihre Geräte umfassen mobile Geräte (Siri, Google-Assistent und Bixby) sowie Haushaltsgeräte (Alexa). Die Migrationsgeschichte der Teilnehmer*innen aus der Türkei nach Deutschland variiert von sieben Monaten bis hin zu elf Jahren, und die meisten von ihnen sprechen mindestens drei Sprachen (Türkisch, Deutsch und Englisch). Alle Teilnehmer*innen haben entweder bereits einen Hochschulabschluss erworben oder befinden sich aktuell im Erwerb dessen in Deutschland.

Die Interviews wurden auf Türkisch als semistrukturiertes Interview mit anschließenden Folgefragen durchgeführt (siehe z.B. Flick 2018). Da ich mich für die persönlichen Erfahrungen der Befragten interessierte, habe ich einen phänomenologischen Ansatz gewählt, der durch offene Fragen zu den Gefühlen, Wahrnehmungen und dem Verständnis der Teilnehmer*innen detaillierte Beschreibungen ihrer Erfahrungen mit einem Phänomen ermöglicht (Roulston/Choi 2018). Die offenen Fragen umfassten die Motivation für den Kauf und die Nutzung, die Dauer der Nutzung, Sprachpräferenzen, Bedenken hinsichtlich der Privatsphäre, Kommunikationsprobleme, das Geschlecht ihrer Sprachassistenten, persönliche/emotionale Verbindungen zu den Sprachassistenten sowie lustige, interessante oder beängstigende Erfahrungen. Abschließend bat ich sie am Ende des Interviews, mir zu demonstrieren, wie sie mit ihren Geräten interagieren, wobei dies einige sehr kurz und andere ausführlicher taten.

Die aufgezeichneten Interviews wurden in Standardtürkisch transkribiert, unter Einbezug einiger paralinguistischer Elemente wie Pausen (.) und Lachen (@). Die erhobenen Daten wurden zunächst mittels der Grounded-Theory-Methode thematisch kodiert und kategorisiert (Glaser/Strauss 2009). Die thematische Analyse zeigt, dass Sprachassistenten je nach individueller Einstellung und Kontext des Interviews diskursiv und situativ als Tool oder Interaktionspartner konstruiert werden (siehe auch Schneider 2022b). In Bezug auf die Frage, welche Strategien mehrsprachige Nutzer*innen entwickeln,

um sich die Sprachassistenten anzueignen, sprechen die Befragten von zwei unterschiedlichen Praktiken. Diese umfassen (1) phonetische und lexikalische Einstimmungsstrategien, um das Gerät als Tool effektiver zu nutzen sowie (2) das Training des Gerätes als Interaktionspartner mit kulturspezifischen Witzen und der richtigen Aussprache der Personennamen. Strategien beziehen sich hier auf die diskursiven Praktiken, die Nutzer*innen anwenden, um bestimmte »Probleme« ihrer Geräte zu lösen, die zum Beispiel aufgrund des monolingualen Bias entstehen können. Des Weiteren zeigt die Datenanalyse, dass die Nutzer*innen der Maschine Handlungsfähigkeit (*agency*) zuschreiben, die auf spezifische Vorstellungen und Diskurse von künstlicher Intelligenz hindeuten. Im Folgenden werden die Praktiken und repräsentative Beispiele aus den Kategorien näher erläutert.

4.2 Lexikalische und phonetische Einstimmungen: Sprachassistenten als »Tools«

In diesem Teil wird dargestellt, welche lexikalischen und phonetischen Einstimmungsstrategien die Befragten verwenden, wenn sie Sprachassistenten als Tool und Erweiterung ihrer Hand oder ihres Armes (vgl. McLuhan 2008) konstruieren, um den Tastsinn zu substituieren. Alle Teilnehmenden gaben an, dass sie ihre Gadgets anfangs erworben haben, um ihre haptischen Aktivitäten durch Sprachsteuerung zu ersetzen. Diese Praktiken unterscheiden sich zwischen den Nutzer*innen und umfassen das Öffnen bestimmter Apps, das Abspielen von Musik, und die Verwendung des Weckers oder Timers. Für die genannten Anwendungen werden keine weiteren Geräte benötigt außer den Smartphones, in denen die Sprachsteuerungssoftware integriert ist, oder den erworbenen Sprachsteuerungsgeräten. Darüber hinaus haben einige Nutzer*innen bereits Smart-Geräte wie Lampen, Fernseher, Staubsauger oder Kaffeemaschinen erworben oder planen den Kauf.

Die Codes in dieser Kategorie beinhalten metapragmatische Reflexionen in Bezug auf geeignete Phrasen im Kontext der Sprachsteuerung. Die Teilnehmenden benutzen ihre Alexas auf Deutsch oder Englisch, da ihre Geräte keine türkische Einstellung anbieten. Für andere Sprachassistenten wie Siri und Google Assistant unterscheiden sich die Präferenzen. In diesem Kontext gilt die einfache Imperativform als am besten geeignet, um schnelle Reaktionen des Geräts zu ermöglichen. Als die Teilnehmenden während des Interviews zeigten, wie sie mit ihren Geräten interagieren, z.B., um Musik abzuspielen, bevorzugten sie die einfache Imperativform: »Alexa (.) spiel Musik« oder

»Navigiere mich zur Arbeit über Google Maps«. Da die Teilnehmer*innen Kurzsätze im Imperativ bevorzugten, wollte ich erfahren, ob sie diese Form als Standard für den Umgang mit Sprachsteuerung betrachten. Ich fragte sie, ob sie ausschließlich den Imperativ verwenden oder auch Fragen stellen wie »Kannst du das Licht anmachen?«. Diese Frage war für die Befragten ein Anstoß, darüber nachzudenken, welche Kommunikationsform sich für den Umgang mit der Sprachsteuerung am besten eignet.

Exzerpt 1

Tabi tabi yani ben öyle ›lütfen‹ ›rica ederim‹ gibi konulara hiç girmeden ›ışığ
aç, orayı kapa, bilmem ne yap‹.
(Ja, natürlich, ich gehe nicht auf Formulierungen wie ›bitte‹ oder ›ich bitte dich‹ ein. Ich sage: ›Mach das Licht an, schalte das aus, tu dies und das.‹)
(Nutzer von deutschsprachiger Siri, 33)

Exzerpt 2

Hatta başlata bile gerek yok yani direk Netflix deyince direk onu açıyor.
(Es ist sogar nicht mal nötig, »Start« [zu sagen]. Wenn ich direkt Netflix sage, schaltet es⁵ direkt ein.)
(Nutzer von türkischsprachigem Google-Assistent, 31)

Die Befragten argumentieren, dass es einfacher und schneller ist, kurze Sätze zu verwenden. Höflichkeitsstrategien wie »bitte« oder das Bilden eines vollständigen Satzes werden als überflüssig angesehen, da der Name einer App wie »Netflix« ausreicht, um das Gerät zu benutzen. Ähnliche Beobachtungen wurden in der frühen Chatbot-Forschung gemacht, die die Vereinfachung von Eingaben in die konversationelle künstliche Intelligenz zeigen, zum Beispiel durch die Verwendung von Imperativen oder Infinitiven, auch »Computer-Talk« genannt (Lotze 2016; Fischer 2006).

Allerdings gaben die Befragten an, dass die Nutzung nicht immer reibungslos funktioniere, selbst bei Beachtung der lexikalischen Einstimmungen wie Nutzung der Schlüsselwörter oder kurzer Sätze. Ein Interviewter äußert

5 Die Pronomen sind in der türkischen Sprache nicht geschlechtsspezifisch. In meinen Übersetzungen verwende ich das Pronomen »es«. Das impliziert jedoch nicht, dass die Befragten ihre Sprachassistenten nicht gendern.

in dem folgenden Ausschnitt ausführlich seine Frustration über die Fehlfunktionen von Siri bei der Nutzung im Auto. Des Weiteren berichtet er davon, dass er aufgrund der Unfähigkeit von Alexa, nicht-deutsche Namen zu verarbeiten, auf die Nutzung der Maschine verzichtet. Dennoch verwendet er Siri weiterhin im Auto und hat immer noch einige Schwierigkeiten, die er im Interview anspricht.

Exzerpt 3

Arabada da Bluetooth'la arabanın hoparlörüne bağlı olduğunda oradan kumanda etmek çok kolay. Telefonu kullanmamaya çalışıyorum araba kullanırken. Birini aramam gerekiyordu, yani orada aramam gereken insanın Türkçe ismini Almanca Alman biri nasıl söylemiş gibi aksan yapıp kasıp onun ev numarasını kendi kendime aratmayı Siri'yle becerdim ve kendimle gurur duydum. Yani aslında bu *multilanguage* durumlarda (.) hani Türkçe isminde bir insan aratmaya çalışıyorum. Telefon İngilizce, Siri Almancaya ayarlı, araba ne dil anlıyor artık Allah bilir«

(Im Auto ist es sehr einfach, mit Bluetooth zu navigieren, weil es mit dem Speaker vom Auto verbunden ist. Ich versuche, mein Telefon während der Fahrt nicht zu benutzen. Einmal musste ich jemanden anrufen. Ich habe mich sehr bemüht, einen türkischen Namen auf Deutsch mit einem deutschen Akzent wie ein Deutscher auszusprechen. Ich habe es selbst mit Siri geschafft und war sehr stolz auf mich. Ich meine in diesen *multilanguage*-Situationen (.) Ich versuche, [Siri] nach jemandem suchen zu lassen, der einen türkischen Namen hat. Mein Telefon ist auf Englisch, Siri ist auf Deutsch eingestellt, weiß Gott, welche Sprache das Auto versteht.)

(Nutzer von deutschsprachiger Siri, 33)

Der Befragte beschreibt, wie er seine Sprache im Moment der Interaktion phonetisch an die »Erwartungen« des Geräts anpasst. Was er als »einen türkischen Namen auf Deutsch mit einem deutschen Akzent wie ein Deutscher auszusprechen« erklärt, ist eine typische Strategie, über die fast alle Teilnehmer*innen, die ihre Geräte auf Englisch oder Deutsch eingestellt haben, berichtet haben. Damit ist gemeint, dass sie türkische oder nicht-deutsche Namen anders aussprechen müssen (für eine detaillierte Diskussion siehe Leblebici i. E.). Die phonetische Einstimmung ist aufgrund der Einsprachigkeit der Sprachassistenten notwendig, welche türkische Namen in englischen oder deutschen Einstellungen nicht erkennen (siehe Beneteau et al. 2019 zu einer ähnlichen Beobachtung für spanische Namen).

Im Verlauf der Äußerung wird deutlich, dass der Befragte sich mit der Aussage »ich habe mich sehr bemüht« selbst als aktiven Agenten konstruiert, der das Gerät benutzt. Seine Rolle als Agent ist hier stark, da er eine schwierige Aufgabe angesichts des monolingualen Designs der Technologie erfüllt: »Ich habe es selbst mit Siri geschafft und war sehr stolz auf mich«. Obwohl er sein Gerät nur als technisches Werkzeug konstruiert, verlagert sich die Agentenrolle schrittweise auf die Maschine, wenn ein Ausfall beobachtet wird. Diese graduelle Verlagerung der *agency* wird durch die unterstrichenen Passagen hervorgehoben. Eine genauere grammatikalische Analyse offenbart hier einige aufschlussreiche Erkenntnisse zum Verständnis der allmählichen Verlagerung der *agency*. Nachdem er erklärt, dass er in einem Moment stolz darauf war, eine Aufgabe mit Siri erledigt zu haben, betont er seine Frustration in anderen Fällen:

Türkçe isminde bir insan **ara-t-ma-ya** çalışıyorum
 Türkisch genannt eine Person **anruf-KAUS-en-AKK** (ich) versuche
 (Ich versuche, [Siri] nach jemandem suchen zu lassen, der einen türkischen Namen hat.)

In der türkischen Sprache wird der Kausativ durch ein Suffix (z.B. «-t») gebildet, das die Veranlassung einer Handlung beschreibt (vgl. z.B. Nakipoğlu et al. 2022). In der deutschen Übersetzung wird das Hilfsverb »(anrufen) lassen« verwendet, da kein eigenständiges Verb entsteht. Dementsprechend muss der Interviewte auch nicht mehr das Subjekt »Siri« im Satz erwähnen, da er bereits durch das Kausativverb auf Siri Bezug genommen hat. Er verleiht Siri also nur durch die Einfügung des Buchstabens »-t« eine Form von *agency*. Das Kausativverb fungiert als leichter Übergang, bevor der Befragte seine Frustration über die Maschine offen ausspricht – »Weiß Gott, welche Sprache das Auto versteht«. Am Ende ist die Maschine nicht nur das Subjekt, sondern auch ein menschenähnlicher Agent, der nicht »verstehen« kann.

Interessant an seiner Erzählung ist vor allem der Zusammenhang zwischen seiner wachsenden Frustration und der Funktionalität des Geräts. Er konstruiert sich selbst als aktiver Nutzer, wenn seine Strategien zur phonetischen Einstimmung funktionieren. Wenn seine Strategien jedoch nicht funktionieren, werden Siri und das Auto als Agenten konstruiert. In ähnlicher Weise zeigt Ritzmann (2017) anhand von Reddit-Diskussionen über Siri, dass Menschen dazu neigen, Sprachassistenten menschenähnlichere Eigenschaften zuzuschreiben und sie als mächtige soziale Akteure darzustellen,

wenn sie sich »fehlverhalten«. Diese Beispiele zeigen, dass die Einstimmung in der Mensch-Maschine-Interaktion wechselseitig funktionieren muss. So stimmen sich nicht nur Menschen mit phonetischen und lexikalischen Praktiken ein, sondern es gibt auch die Erwartung, dass die Geräte den Input der Nutzer*innen erfolgreich verarbeiten und sich damit anpassen.

Die Analyse verdeutlicht, dass die Kategorisierung der Sprachassistenten als Maschine oder menschenähnlicher Agent diskursiv und situativ konstruiert wird. Die Nutzer*innen bewegen sich zwischen den Kategorien hinsichtlich des wahrgenommenen Erfolgs des Geräts (siehe auch Habscheid et al. 2023). Andererseits ist zu beobachten, dass manche Nutzer*innen bestimmte Praktiken entwickeln, um eine persönliche Beziehung zu den Sprachassistenten aufzubauen. Darauf wird im folgenden Abschnitt eingegangen.

4.3 Aussprache und Witze beibringen: Anthropomorphisierung von Sprachassistenten

Wie bereits im Abschnitt über Sprachassistenten dargestellt, werden diese Technologien als »anthropomorphisierte virtuelle Agenten« mit menschenähnlichen Stimmen, Eigenschaften und Konversationskripten konzipiert (vgl. Sweeney 2016). Bei der Kodierung der Interviews wurde deutlich, dass die meisten Befragten über ihre Gadgets sprachen und dabei auch auf die menschlichen Züge ihrer Geräte anspielten. Mit bestimmten Aneignungspraktiken ergänzen die Nutzer*innen die in den Geräten eingeschriebenen Anthropomorphisierungspraktiken mit ihren kultur- und sprachspezifischen Vorstellungen.

Laut den Aussagen der Befragten sind die aktuell verfügbaren Sprachassistenten aufgrund ihres monolingualen Designs nur in der Lage, bestimmte Personennamen zu verarbeiten (z. B. nur deutsche Namen in der deutschsprachigen Option), wie aus Exzerpt 3 hervorgeht. Die folgende Strategie besteht darin, diese zu umgehen, indem User dem Sprachassistenten die richtige Aussprache »beibringen«. Die Nutzer*innen haben sich dabei nicht nur die Technologie angeeignet, sondern diese auch aktiv angepasst. Der Befragte erklärt, dass er beim Ausprobieren von Siri herausgefunden hat wie er dem Gerät beibringen kann, seinen Namen auszusprechen:

Exzerpt 5

I: Sirimi bir de çok tatlı bir şekilde eğittim e şeyi öğrendikten sonra (.) te-laffuz muhabbetini öğrendikten sonra- o da çok spontane oldu mesela. İşte benim (.) ismimi falan şey yaparken nereden oldu hatırlamıyorum ben de iste »say my name« falan mı dedim ne (.) e:: sonra şey dedim birdenbire »ah you are pronouncing it wrong« dedim. »Can you teach me how to pronounce?«. Bir dakika dedim nası::l inanılma::z

(Ich habe mein Siri auf eine sehr süße Art und Weise trainiert, nachdem ich gelernt habe, dass (.) nachdem ich die Aussprachefunktion gelernt habe – es war auch sehr spontan. Als ich meinen Namen sagte (.) Ich weiß nicht mehr, wie es passiert ist. Ich sagte: »say my name« oder so, und dann sagte ich plötzlich: »Ah, you are pronouncing it wrong«. [Es antwortete mir:] »Can you teach me how to pronounce? ». »Moment mal« habe ich gesagt »wie:: ungläubli::ch«)

D: A:: çok iyi.
(Oh sehr gut.)

I: Sonra işte çok sık aradığım insanlar işte Dilara, Giulia filan. Türk isimleriyle çok daha fazla sıkıntı çekiyor tabi. İşte Betül filan diyorum anlamıyor, onları falan öğrettikten sonra (.) benim için çok daha rahat oldu. Ondan sonra insanları aramaya başladım. E:: ve işte galiba *nickname* falan da var. Çünkü benim ismim aslında ismim Emre Can ama ben hani Emre'yi kullanıyorum. Bunu öğretebildim ona şey diye »Your name is Emre Can but since we are close, I get to call you Emre« falan diyordu.

(Dann [habe] ich die Namen der Leute [gespeichert], die ich am häufigsten anrufe, wie Dilara⁶, Giulia und so weiter. Mit türkischen Namen hat es natürlich mehr Probleme. Wenn ich Betül oder so sage, versteht es das nicht. Nachdem ich ihm diese [Namen] beigebracht hatte, wurde es viel bequemer für mich. Und ich glaube, es gibt auch *Nickname* oder so. Ich heiße nämlich eigentlich Emre Can, aber ich benutze Emre. Das konnte ich ihm beibringen, es sagt: »Your name is Emre Can but since we are close, I get to call you Emre« oder so ähnlich.)

(Nutzer von englischsprachiger Siri, 26)

6 Die Namen sind geändert, um die Anonymität der Befragten zu wahren. Es ist jedoch in diesem Kontext wichtig, dass Dilara, Betül und Emre für türkische Namen stehen. Giulia hingegen steht für einen italienischen Namen.

Der Befragte erzählte, wie glücklich er war, die Funktion zu entdecken mit der er Siri trainieren konnte, seinen türkischen Namen korrekt auszusprechen. Seine Formulierungen lassen bereits eine emotionale Bindung zu der anthropomorphisierten Maschine erkennen, z. B. wenn er den Sprachassistenten als »Sirim (mein Siri)« bezeichnet oder sagt, er habe das Gerät auf eine »süße Art und Weise trainiert«. Der Interviewte berichtete, dass er Siri beigebracht hat, die italienischen und türkischen Namen seiner Freundinnen auszusprechen, nachdem er festgestellt hatte, dass diese Funktion existiert. Obwohl der Ausschluss nicht-englischer Namen als eine Reproduktion von strengen nationalen Sprachmodellen und Ausgrenzungen zahlreicher Personen problematisiert werden kann, bewertet der Befragte die Trainingsfunktion seines Geräts positiv. Seine damalige Reaktion, »nası:l inanılma::z (wie:: ungläubli::ch)«, als er die Funktion entdeckt hat, zeigt, dass er von den Affordanzen der Technologie begeistert ist. Doch er geht noch einen Schritt weiter und trainiert sein Gerät so, dass es seinen Spitznamen verwendet. Siri wird hier personifiziert, und als jemand dargestellt, der eine gute Beziehung mit dem Benutzer aufgebaut hat. Basierend auf dieser Nähe »darf« Siri den Befragten auch mit seinem Spitznamen ansprechen. Siri wird schließlich darauf trainiert, diese Nähe zum Nutzer durch den einen vorgespeicherten Satz zu performen. Das Training von Sprachsteuerungstechnologien zur Erkennung von Personennamen, die sich nicht in den engen nationalen Sprachoptionen der Assistenten befinden, ist ein wichtiges Element für den Aufbau persönlicher Beziehungen. So konstruieren die Nutzer*innen Siri diskursiv als einen menschenähnlichen Agent.

Der Befragte verweist zudem darauf, dass Türkisch eine Sprache ist, die technisch nicht gut unterstützt wird. Er argumentiert, dass die Technologie mit türkischen Namen mehr Probleme hat und signalisiert seine Überraschung über die Trainingsfunktion. Durch seinen Vergleich zwischen italienischen und türkischen Namen zeigt er auf, dass es »natürlich« zu erwarten ist, dass Italienisch besser in der Entwicklung in Sprachsteuerungstechnologien abschneidet. Dieser Vergleich indexiert die Hierarchie zwischen europäischen und außereuropäischen Sprachen in der Entwicklung von Sprachtechnologien. Aktuell funktionieren Sprachassistenten in europäischen bzw. westlichen Sprachen am besten, mit einigen wenigen Ausnahmen. Zum Beispiel gibt es bei den Smart Speakern von Amazon, Google und Ap-

ple keine Unterstützung für Türkisch. Zwar unterstützt Siri die türkische Sprachoption, jedoch sind viele Funktionen nicht verfügbar⁷.

Die Integration von Witzten, Komplimenten und Begrüßungen ist eine weitere anthropomorphisierende diskursive Praktik, über die in den Interviews berichtet wurde. Diese Praktiken zeigen, ähnlich wie Emre Cans Methode, Spitznamen zu verwenden, dass die Teilnehmer*innen den Assistenten als eine Person konstruieren, mit der sie eine Beziehung haben. Im folgenden Auszug beschreibt eine Teilnehmerin, wie sie die deutsche Version von Alexa mit einem türkischen Zitat trainiert hat.

Exzerpt 6

I: Türkçe desteklemiyor sanırım ama mesela bazı komutlar verdiğinde sana istediğin cevabı verebiliyor. Onlara birkaç tane Türkçe böyle (.) şeyler kaydetmişim esprili kelimeler falan @

(Ich schätze, es unterstützt Türkisch nicht, aber wenn du zum Beispiel ein paar Befehle gibst, kann es dir die gewünschte Antwort geben. Ich habe ein paar türkische (.) Dinge gespeichert wie ein paar lustige Wörter und so @)

Didem: A çok iyi, neler kaydettin hatırlıyor musun?

(Oh sehr gut, erinnerst du dich, was du gespeichert hast?)

I: Yani şey mesela işte motive edici sözler işte (.) şu an komutu hatırlamıyorum ama işte Almanca bir komut verdiğinde mesela bana (.) işte ne diyordu ya (.) »hayat bir gündür« o da- ne diyordu ya- »hayat bir gündür o da bugündür« mü ne öyle bir şey diyordu @ Türkçe bir şekilde (.) bir tane de öyle motive edici bir şeyler söyletmişim de yani.

(Ich meine zum Beispiel motivierende Zitate (.) Ich erinnere mich nicht an den Befehl, aber zum Beispiel, wenn man einen deutschen Befehl aufgibt (.) sagt es mir (.) »das Leben ist ein Tag« – wie war das – »das Leben ist ein Tag, und es ist heute« oder so etwas in der Art @ auf Türkisch (.) lasse ich mir so etwas Motivierendes sagen.)

(Nutzerin von deutschsprachiger Alexa, 20J)

Die Befragte erwähnt zunächst, dass Alexa keine türkische Sprachoption anbietet. Allerdings ist es möglich, dem Gerät bestimmte Sätze und Wörter beizubringen. Sie beschließt, Alexa mit witzigen und motivierenden Sprüchen

7 <https://www.apple.com/ios/feature-availability/#siri> (Stand 12.09.2023)

in türkischer Sprache zu trainieren. Während unseres Gesprächs erinnert sich die Befragte an den Satz, den sie ausgewählt hat, um ihr Gerät zu trainieren: »Hayat bir gündür, o da bugündür (Das Leben ist ein Tag, und der ist heute)«. Der Satz stammt ursprünglich aus einem Gedicht von Can Yücel (Ufuk 2015: 30) und findet in abgewandelter Form Verwendung auf Social-Media-Plattformen, in Fernsehsendungen sowie in der Literatur. Da das Zitat in der türkischen Populärmedienkultur verwurzelt ist, empfindet die Befragte es als unpassend und humorvoll, wenn Alexa, insbesondere in Verwendung der deutschen Spracheinstellung des amerikanischen Technologieunternehmens Amazon, diesen Satz wiedergibt.

Diese Trainingspraktiken des Sprachassistenten tragen zur Konstruktion einer technikaffinen Identität der Nutzer*innen in Interviewsituationen bei und können somit als Performance dieser Identität betrachtet werden (siehe auch Leblebici 2024). Obwohl die Sprachassistenten letztendlich die Ausgabe produzieren, indem sie vorherige Befehle verarbeiten, experimentieren die Nutzer*innen mit den Affordanzen, um die Technologie zu ko-konstruieren und sich anzueignen. Die Konstruktion dieser Identitäten findet in Zusammenarbeit mit den Sprachassistenten statt, wodurch von einer posthumanistischen oder Hybrididentität zwischen Mensch und Maschine gesprochen werden kann. Diese Mitgestaltung verweist auf ihre Geschichte und ihre Beziehung zur Maschine. Im Diskussionsteil werden nun diese Erkenntnisse in Bezug auf räumliche Repertoires näher erläutert.

5 Diskussion und Fazit

Das Ziel dieses Beitrags war herauszufinden, wie mehrsprachige Nutzer*innen ihr räumliches Repertoire nutzen, um sich auf Sprachassistenten einzustimmen. Um die Lücke in der Mehrsprachigkeitsforschung zu adressieren, wurden qualitative Interviews mit Nutzer*innen durchgeführt, die eine Migrationsgeschichte aus der Türkei nach Deutschland haben. Die erhobenen Daten wurden sowohl auf thematischer als auch auf diskursiver Ebene analysiert. Im Folgenden werden die Erkenntnisse bezüglich räumlicher Repertoires und früherer Forschung zu Nutzungspraktiken dargelegt.

Digitale Technologien sind Teil des räumlichen Repertoires des Menschen als kulturelle, materielle und semiotische Ressourcen. Die Ressourcen liegen weder allein in den Sprachsystemen noch in inneren menschlich-kognitiven Prozessen, sondern werden von den Affordanzen und Beschränkungen

von Sprachassistenten, ihren Designer*innen und kulturellen Diskursen über KI geprägt. Durch die Kombination von linguistischen, materiellen, kulturellen und semiotischen Ressourcen in ihren individuellen Repertoires erzeugen Nutzer*innen zu bestimmten Zeitpunkten und an bestimmten Orten Bedeutung. Sie nutzen lexikalische und phonetische Strategien, die dazu beitragen, Smartphones und andere digitale Geräte effektiv als Werkzeuge zu gebrauchen, um das haptische Engagement zu ersetzen. Im Hinblick auf lexikalische Einstimmungen gibt es Strategien, die als Computer-Talk (Lotze 2016; Fischer 2006) kategorisiert werden können. Diese basieren auf den metapragmatischen Reflexionen der Teilnehmenden, welche die Verwendung von Imperativsätzen oder Schlüsselwörtern bevorzugen. Es gibt auch Berichte über phonetische Anpassungen, die darauf abzielen, Fehlfunktionen bei der Verarbeitung von nicht-englischen oder nicht-deutschen Namen zu umgehen und mit der für das System festgelegten einsprachigen Ausrichtung zu arbeiten.

Obwohl verschiedene phonetische Anpassungen schon festgestellt wurden (z.B. Beneteau et al. 2019; Mavrina et al. 2022; Motta/Quaresma 2022), entwickeln die Befragten Strategien, um die Beschränkungen der Technologie in Bezug auf monolinguale Bias der Hersteller zu umgehen. In den monolingual angelegten Sprachoptionen wird das noch immer bestehende Paradigma über Sprache deutlich, da Sprachen in den Technologien als zählbare, monolinguale und standardisierte Einheiten eingebettet sind. Aus einer posthumanistischen Sicht zeigt sich jedoch, dass Sprache über solche Konzeptualisierungen hinausgeht, wie die kreativen Einstimmungspraktiken verdeutlichen. Nutzer*innen erwarten eine wechselseitige Einstimmung in der Mensch-Maschine-Interaktion und stellen das einsprachige Interaktionsdesign in Frage.

Obwohl gewisse Strategien wie Computer-Talk für eine effiziente Nutzung des Sprachassistenten als Tool entwickelt wurden, zeigt die nähere Analyse der Diskurse von Befragten, dass der Maschine die Verantwortung zugeschrieben wird, wenn das System nicht funktioniert (siehe auch Ritzmann 2017). Im Gegensatz zu früheren Studien die die Verwendung von Höflichkeitsformen als Anthropomorphisierung untersucht haben (z.B. Pradhan et al. 2019; Lopatovska/Williams 2018; Lopatovska/Oropeza 2018), zeigt diese Studie, dass die Vermenschlichungspraktiken diskursiv und situativ im Interview ausgeübt werden. Basierend auf dem wahrgenommenen Erfolg des Gerätes wird der Maschine eine menschenähnliche *agency* verliehen. Zwischen Menschen und Maschinen gibt es also kein »Verstehen«, sondern menschliche Akteure eignen sich Maschinen an und machen sie durch Verleihung der *agency* zu handlungs-

fähigen Entitäten. Eine gewisse Form der »Verständigung« wird dabei diskursiv konstruiert.

Sprachassistenten sind wie andere kulturelle Tools (Jones 2020), die nicht nur bestimmte Praktiken wie die mündliche Steuerung von Smart-Geräten ermöglichen, sondern auch solche, die bestimmte Identitätskonstruktionen und Denkweisen der Nutzer*innen bedingen. Dies wird auch beim Kontrast der Wahrnehmung bei Fehlfunktionen deutlich. Obwohl einige Nutzer*innen Frustration äußern, sind Fehlfunktionen für einige Befragte ein Mittel, um weitere Affordanzen zu entdecken und dadurch einen Beziehungsaufbau zu performen. In beiden Fällen werden jedoch die Maschinen anthropomorphisiert.

The resources that enter into a repertoire are indexical resources, language materials that enable us to produce more than just linguistic meaning but to produce images of ourself, pointing interlocutors towards the frames in which we want our meanings to be put. (Blommaert/Backus 2013: 28)

Gemeinsam mit den Sprachassistenten als kommunikative und indexikalische Ressourcen in ihrem Repertoire erzeugen die Nutzer*innen ein posthumanistisches Selbstbild. Indem sie Spitznamen und Medienreferenzen einbauen und so ihr Gerät darauf trainieren, Sätze auszugeben die eine Freundschaft mit ihnen suggerieren, performen sie eine enge Beziehung zu der KI-Technologie. Diese Praktiken sind besonders bemerkenswert, da Unternehmen ähnliche Strategien der synthetischen Personalisierung (Fairclough 1993) anwenden, die die Sprachassistenten als menschenähnliche Charaktere darstellen, die Witze machen und sich dabei auf die populären US-amerikanischen Medienkulturen beziehen. So verschaffen sich die Unternehmen Zugang zu Privatsphären, indem sie die Unterscheidung zwischen öffentlich und privat verwischen (Zuboff 2019). Die Studie zeigt jedoch, dass die Nutzer*innen auch zur Erweiterung der KI-Persona beitragen, indem sie spezifische kommunikative Ressourcen wie türkische Populärkultur und Spitznamen aus ihrem Repertoire verwenden. Da ihre Implementierungen nicht in den englischen oder deutschen Versionen dieser Technologien verfügbar sind, ist davon auszugehen, dass sich die Befragten die Sprachassistenten entsprechend ihres kulturellen Verständnisses aneignen und ihre eigene Version gestalten.

Grenzen der vorliegenden Studie müssen ebenfalls berücksichtigt werden. Das Korpus, auf dem diese Forschung basiert, ist vergleichsweise klein und nicht repräsentativ für mehrsprachige Personen. Das Hauptziel ist es

zu verstehen, welche Herausforderungen und Möglichkeiten mit sprachbasierten Assistenzsystemen in mehrsprachigen Settings auftauchen. Zudem bestand die Stichprobe hauptsächlich aus relativ jungen Teilnehmenden, was die Altersvielfalt einschränkt und potenziell Auswirkungen auf die Ergebnisse hat. Zukünftige Forschung sollte sich auf die Bewertung der Faktoren Alter und Bildungsniveau konzentrieren. Es ist zu beachten, dass meine eigene Migrationserfahrung sowie meine Identität als Linguistin beeinflusst haben könnten, wie ausführlich die Befragten über das monolinguale Design sprachen und türkische Medienreferenzen mitteilten.

Offen bleiben die Fragen, welche Bedeutungen die Wahl unterschiedlicher Geräte wie Smart Speaker oder integrierte Smartphone-Assistenten innerhalb der Medienökologie der einzelnen Nutzer*innen hat (in Anlehnung an Madiannou/Miller 2013). Eine genauere Analyse ist notwendig um zu verstehen, wie Sprecher*innen von Minderheitensprachen mit diesen Technologien umgehen, insbesondere wenn diese Sprachen von der Technologie nicht unterstützt werden. Da Sprachtechnologien weiterentwickelt werden und teilweise semi-bilinguale Optionen anbieten, sollten zukünftige Studien auch thematisieren, wie diese im Hinblick auf sprachliche Einschränkungen und die Rekonstruktion von Sprachhierarchien und -grenzen umgesetzt werden.

Literaturangaben

- Abercrombie, Gavin/Amanda Cercas Curry/Mugdha Pandya/Verena Rieser (2021): Alexa, Google, Siri: What are Your Pronouns? Gender and Anthropomorphism in the Design and Perception of Conversational Assistants. In: *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing*, 24–33.
- Alač, Morana/Yelena Gluzman/Tiffany Aflatoun/Adil Bari/Buhang Jing/German Mozqueda (2020): Talking to a Toaster: How Everyday Interactions with Digital Voice Assistants Resist a Return to the Individual. In: *Evental Aesthetics* 9:1, 3–53.
- Ammari, Tawfiq/Jofish Kaye/Janice Y. Tsai/Frank Bentley (2019): Music, Search, and IoT: How People (Really) Use Voice Assistants. In: *ACM Transactions on Computer-Human Interaction* 26:3, 1–28.
- Androutopoulos, Jannis (2015): Networked multilingualism: Some language practices on Facebook and their implications. In: *International Journal of Bilingualism* 19:2, 185–205.

- Androutsopoulos, Jannis/Kristin Vold Lexander (2021): Digital polycentricity and diasporic connectivity: A Norwegian-Senegalese case study. In: *Journal of Sociolinguistics* 25:5, 720–736.
- Artamonova, Olga/Jannis Androutsopoulos (2020): Smartphone-Based Language Practices among Refugees: Mediatonal Repertoires in Two Families. In: *Journal für Medienlinguistik* 2:2, 60–89.
- Barratt, Leslie (2018): Monolingual Bias. In: Liontas, J. I. (Hg.): *The TESOL Encyclopedia of English Language Teaching*, Wiley, 1–7.
- Bender, Emily M./Timnit Gebru/Angelina McMillan-Major/Shmargaret Shmitchell (2021): On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.
- Beneteau, Erin/Olivia K. Richards/Mingrui Zhang/Julie A. Kientz/Jason Yip/Alexis Hiniker (2019): Communication Breakdowns Between Families and Alexa. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13.
- Bitkom Research (2021): *Die Zukunft der Consumer Technology – 2021. Marktentwicklung & Mediennutzung, Trends & Technologien*. Abrufbar unter: https://www.bitkom.org/sites/default/files/2021-09/210817_ct_studie_2021.pdf (Stand: 28.10.24).
- Blackledge, Adrian/Angela Creese (2020): Interaction ritual and the body in a city meat market. In: *Social Semiotics* 30:1, 1–24.
- Blommaert, Jan (2010): *The sociolinguistics of globalization*, Cambridge, UK/New York: Cambridge University Press.
- Blommaert, Jan/Ad Backus (2013): Superdiverse Repertoires and the Individual. In: Saint-Georges, Ingrid de/Jean-Jacques Weber (Hg.): *Multilingualism and Multimodality*, Rotterdam: SensePublishers, 11–32.
- Busch, Brigitta (2012): The Linguistic Repertoire Revisited. In: *Applied Linguistics* 33:5, 503–523.
- Cowley, Stephen J. (2011): Distributed language. In: Cowley, Stephen J. (Hg.): *Distributed language*, Amsterdam/Philadelphia: John Benjamins, 1–14.
- Crawford, Kate (2021): *Atlas of AI: power, politics, and the planetary costs of artificial intelligence*, New Haven: Yale University Press.
- Crawford, Kate/Vladan Joler (2018): Anatomy of an AI System. Abrufbar unter: <https://www.anatomyof.ai>
- Doğruöz, A Seza/Sunayana Sitaram (2022): Language Technologies for Low Resource Languages: Sociolinguistic and Multilingual Insights. In: *Proceedings of SIGUL2022 @LREC2022*, 92–97.

- Fairclough, Norman (1993): Critical Discourse Analysis and the Marketization of Public Discourse: The Universities. In: *Discourse & Society* 4:2, 133–168.
- Fischer, Kerstin (2006): *What computer talk is and isn't: human-computer conversation as intercultural communication*, Saarbrücken: AQ-Verl.
- Flick, Uwe (2018): *The SAGE Handbook of Qualitative Data Collection*, London: Sage.
- Ford, Marcia/William Palmer (2019): Alexa, are you listening to me? An analysis of Alexa voice service network traffic. In: *Personal and Ubiquitous Computing* 23:1, 67–79.
- Gal, Susan/Kathryn A. Woolard (2009): Constructing Languages and Publics: Authority and Representation. In: *Pragmatics* 5:2, 129–138.
- Ganga, Deianira/Sam Scott (2006): Cultural »Insiders« and the Issue of Positionality in Qualitative Migration Research: Moving »Across« and Moving »Along« Researcher-Participant Divides. In: *Forum Qualitative Sozialforschung* 7:3(7).
- Gibson, James J. (2015): *The ecological approach to visual perception: classic edition*, New York/London: Psychology Press, Taylor & Francis Group.
- Glaser, Barney Galland/Anselm Leonard Strauss (2009): *The Discovery of Grounded Theory: Strategies for Qualitative Research*, 4. Paperback Printing. Aufl., New Brunswick: Aldine.
- Gumperz, John J. (1964): Linguistic and Social Interaction in Two Communities. In: *American Anthropologist* 66:6, 137–153.
- Gumperz, John J. (1982): *Discourse Strategies*, Cambridge: Cambridge University Press.
- Habscheid, Stephan/Tim Moritz Hector/Christine Hrnca/David Waldecker (2021): Intelligente Persönliche Assistenten (IPA) mit Voice User Interfaces (VUI) als »Beteiligte« in häuslicher Alltagsinteraktion. Welchen Aufschluss geben die Protokolldaten der Assistenzsysteme? In: *Journal für Medienlinguistik* 4:1, 16–53.
- Habscheid, Stephan/Tim Hector/Christine Hrnca (2023): Human and Non-Human Agency as Practical Accomplishment: Interactional Occasions for Ascription and Withdrawal of (Graduated) Agency in the Use of Smart Speaker Technology. In: *Social Interaction. Video-Based Studies of Human Sociality* 6:1, 1–31.
- Hoy, Matthew B. (2018): Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. In: *Medical Reference Services Quarterly* 37:1, 81–88.

- Humphry, Justine/Chris Cheshier (2020): Preparing for Smart Voice Assistants: Cultural Histories and Media Innovations. In: *New Media & Society* 23:7, 1–18.
- Hymes, Dell (1972): On Communicative Competence. In: Pride, J.B./Janet Holmes (Hg.): *Sociolinguistics. Selected Readings*, Harmondsworth: Penguin Books, 269–293.
- Jones, Rodney (2020): Mediated discourse analysis. In: Adolphs, Svenja/Dawn Knight (Hg.): *The Routledge Handbook of English Language and Digital Humanities*, Abingdon/New York: Routledge, 202–219.
- Jones, Rodney H./Christoph A. Hafner (2021): *Understanding digital literacies: a practical introduction*, Second edition. Aufl., London/New York: Routledge.
- Joshi, Pratik/Sebastin Santy/Amar Budhiraja/Kalika Bali/Monojit Choudhury (2020): The State and Fate of Linguistic Diversity and Inclusion in the NLP World. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6282–6293.
- Koenecke, Allison/Andrew Nam/Emily Lake/Joe Nudell/Minnie Quartey/Zion Mengesha/Connor Touns/John R. Rickford/Dan Jurafsky/Sharad Goel (2020): Racial disparities in automated speech recognition. In: *Proceedings of the National Academy of Sciences* 117:14, 7684–7689.
- Kusters, Annelies/Massimiliano Spotti/Ruth Swanwick/Elina Tapio (2017): Beyond languages, beyond modalities: transforming the study of semiotic repertoires. In: *International Journal of Multilingualism* 14:3, 219–232.
- Latour, Bruno (1999): *Pandora's hope: essays on the reality of science studies*, Cambridge: Harvard University Press.
- Leblebici, Didem (2021): *Language Ideologies in Human-Machine Interactions: A Qualitative Study with Voice Assistant Users*. Masterarbeit. Europa-Universität Viadrina, Frankfurt Oder.
- Leblebici, Didem (2024): »You are Apple, why are you speaking to me in Turkish?«: The Role of English in Voice Assistant Interactions. In: *Multilingua: Journal of Cross-Cultural and Interlanguage Communication* 43:4, 455–485.
- Leblebici, Didem (i. E.): Stylizing the Ideal User: Insights into the Experiences of Turkish Voice Assistant Users in Germany. In: Habscheid, Stefan/Tim Hector/Dagmar Hoffmann/David Waldecker (Hg.): *Voice Assistants in Private Homes. Media, Data and Language in Interaction and Discourse*, Bielefeld: transcript, 179–204.
- Lopatovska, Irene/Heyrling Oropeza (2018): User interactions with »Alexa« in public academic space. In: *Proceedings of the Association for Information Science and Technology* 55:1, 309–318.

- Lopatovska, Irene/Katrina Rink/Ian Knight/Kieran Raines/Kevin Cosenza/Harriet Williams/Perachya Sorsche/David Hirsch/Qi Li/Adrianna Martinez (2019): Talk to me: Exploring user interactions with the Amazon Alexa. In: *Journal of Librarianship and Information Science* 51:4, 984–997.
- Lopatovska, Irene/Harriet Williams (2018): Personification of the Amazon Alexa: BFF or a Mindless Companion. In: *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval – CHIIR '18*, 265–268.
- Lotze, Netaya (2016): *Chatbots: eine linguistische Analyse*, Berlin: Peter Lang.
- Madianou, Mirca/Daniel Miller (2012): *Migration and new media: Transnational families and polymedia*, Abingdon/New York: Routledge.
- Markl, Nina (2022): Language Variation and Algorithmic Bias: Understanding Algorithmic Bias in British English Automatic Speech Recognition. In: *2022 ACM Conference on Fairness, Accountability, and Transparency*, 521–534.
- Mavrina, Lina/Jessica Szczuka/Clara Strathmann/Lisa Michelle Bohnenkamp/Nicole Krämer/Stefan Kopp (2022): »Alexa, You're Really Stupid«: A Longitudinal Field Study on Communication Breakdowns Between Family Members and a Voice Assistant. In: *Frontiers in Computer Science* 4:791704. doi.org/10.3389/fcomp.2022.791704.
- McLuhan, Marshall (2008): *Understanding media: the extensions of man*, London: Routledge.
- Miceli, Milagros/Julian Posada (2022): The Data-Production Dispositif. In: *Proceedings of the ACM on Human-Computer Interaction* 6:CSCW2:460, 1–37.
- Motta, Isabela/Manuela Quaresma (2022): Users' Error Recovery Strategies in the Interaction with Voice Assistants (VAs). In: Black, Nancy L./W. Patrick Neumann/Ian Noy (Hg.): *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)*, Cham: Springer International Publishing, 658–666.
- Nakipoğlu, Mine/Berna A. Uzundağ/Özge Sarigül (2022): Young minds' quest for regularity: Evidence from the Turkish causative. In: *Journal of Child Language* 49:6, 1214–1241.
- Natalé, Simone (2021): *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. Oxford: Oxford University Press.
- Natalé, Simone/Henry Cooke (2021): Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces. In: *Media, Culture & Society* 43:6, 1000–1016.
- OMD Germany (2021): *Welcome to »The Age of Voice 3.0«*. Abrufbar unter: <https://www.omd.com/news/welcome-to-the-age-of-voice-3-0/> (Stand: 28.10.24).

- Pennycook, Alastair (2018): *Posthumanist Applied Linguistics*, London/New York: Routledge/Taylor & Francis Group.
- Pennycook, Alastair/Emi Otsuji (2014): Market Lingos and Metrolingua Francas. In: *International Multilingual Research Journal* 8:4, 255–270.
- Phan, Thao (2017): The Materiality of the Digital and the Gendered Voice of Siri. In: *Transformations* 29, 23–33.
- Phan, Thao (2019): Amazon Echo and the Aesthetics of Whiteness. In: *Catalyst: Feminism, Theory, Technoscience* 5:1, 1–38.
- Porcheron, Martin/Joel E. Fischer/Stuart Reeves/Sarah Sharples (2018): Voice Interfaces in Everyday Life. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12.
- Pradhan, Alisha/Leah Findlater/Amanda Lazar (2019): »Phantom Friend« or »Just a Box with Information«: Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults. In: *Proceedings of the ACM on Human-Computer Interaction* 3:CSCW, 1–21.
- Ritzmann, Martin (2017): *Bad Robot: Discursive constructions of agency in misbehaving AI*. Masterarbeit, University of Bern.
- Roulston, Kathryn/Myungweon Choi (2018): Qualitative Interviews. In: Flick, Uwe (Hg.): *The SAGE Handbook of Qualitative Data Collection*, London: SAGE Publications Ltd, 233–249.
- Rymes, Betsy (2014): Marking Communicative Repertoire Through Metacommentary. In: Blackledge, Adrian/Angela Creese (Hg.): *Heteroglossia as Practice and Pedagogy*, Dordrecht: Springer Netherlands, 301–316.
- Schneider, Britta (2022a): Multilingualism and AI: The Regimentation of Language in the Age of Digital Capitalism. In: *Signs and Society* 10:3, 362–387.
- Schneider, Britta (2022b): Von Gutenberg zu Alexa. Posthumanistische Perspektiven auf Sprachideologie. In: Lind, Miriam (Hg.): *Mensch – Tier – Maschine. Sprachliche Praktiken an und jenseits der Außengrenze des Humanen*, Bielefeld: Transcript Verlag, 327–346.
- Silverstein, Michael (1979): Language structure and linguistic ideology. In: Clyne, Paul R./William F. Hanks/Carol L. Hofbauer (Hg.): *The elements: a parasection on linguistic units and levels*, Chicago: Chicago Linguistic Society, 193–247.
- Sweeney, Miriam (2016): The Intersectional Interface. In: Noble, Safiya Umoja/Brendesha M. Tynes (Hg.): *The intersectional Internet: Race, Sex, Class and Culture Online*, New York: Peter Lang, 215–228.

- Tagg, Caroline/Agnieszka Lyons (2021): Polymedia repertoires of networked individuals: A day-in-the-life approach. In: *Pragmatics and Society* 12:5, 725–755.
- Thurlow, Crispin (2018): Digital Discourse: Locating Language in New/Social Media. In: *The SAGE Handbook of Social Media*, London: SAGE Publications Ltd, 135–145.
- Ufuk, Harika (2015): Yarinlarda Yasamak. In: *Erciyes Aylık Fikir ve Sanat Dergisi*, 30–31.
- Vertovec, Steven (2010): Towards post-multiculturalism? Changing communities, conditions and contexts of diversity. In: *International Social Science Journal* 61:199, 83–95.
- Wu, Yunhan/Daniel Rough/Anna Bleakley/Justin Edwards/Orla Cooney/Philip R. Doyle/Leigh Clark/Benjamin R. Cowan (2020): See What I'm Saying? Comparing Intelligent Personal Assistant Use for Native and Non-Native Language Speakers. In: *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, 1–9.
- Zuboff, Shoshana (2019): *The age of surveillance capitalism: The fight for a human future at the new frontier of power*, New York: PublicAffairs.

Didem Leblebici, M.A.

Große Scharrnstraße 59

15230 Frankfurt (Oder)

leblebici@europa-uni.de

<https://www.kuwi.europa-uni.de/en/professuren-mitarbeitende/sprachgebrauch-migration/team/wissenschaftliche-mitarbeitende/leblebici/index.html>

Die Heterogenität der Sprechenden

Die Entgrenzung von Zeit und Raum in Mensch-Maschine-Kommunikation

Marie-Theres Fester-Seeger

Abstract Dieser Beitrag skizziert einen posthumanen Ansatz zur Mensch-Maschine-Kommunikation. Dabei wird auf die aktiven menschlichen Handlungen ›mit‹ und ›in‹ der Maschine hingewiesen. Die *Languaging*-Perspektive wird verwendet, um Sprache nicht vom Menschen getrennt zu betrachten, sondern als körperliche Aktivität in Echtzeit zu verstehen. In diesem Beitrag wird beschrieben, wie eine Teilnehmerin erstmals auf Arabisch mit dem Sprachassistenzsystem Alexa interagiert. Dabei geht sie als aktive Beobachterin und kognitive Akteurin aktiv auf Ereignisse in ihrer direkten Umgebung ein. Es wird die Idee der Heterogenität der Sprechenden beleuchtet, und wie sich dynamische menschliche Handlungen in der Mensch-Maschine-Kommunikation über Raum, Zeit und andere Personen erstrecken. Der vermeintlich statische Charakter der Mensch-Maschine-Kommunikation wird durch die dynamischen Handlungen der Nutzer*innen und Entwickler*innen beeinflusst. Dadurch entsteht eine Entgrenzung des Statischen.

Keywords *Languaging*; Heterogenität; Sprachassistenzsysteme; Mensch-Maschine-Kommunikation

1 Zu einem posthumanen Verständnis von Mensch-Maschine-Kommunikation

Seit mehr als einem Jahrzehnt ›sprechen‹ Menschen mit Maschinen in ihrer häuslichen Umgebung. Sprachassistenzsysteme wie Amazons *Alexa* oder *Google Home* werden genutzt, um Musik zu hören, den Wecker zu stellen oder das Internet zu bedienen (Hoy 2018). Mit einfachen semiotischen Mit-

tel, wie der Verwendung menschenähnlicher Stimmen, der Zuschreibung einer Persönlichkeit und der Vergabe von Namen, (Natale/Cooke 2021) wird so der Eindruck einer direkten kommunikativen Interaktion mit den Geräten geschaffen. Anstatt *durch* Maschinen zu kommunizieren, wird nun *mit* Maschinen kommuniziert. In Mensch-Maschine-Kommunikation wird die Maschine demnach als eine sozialer Akteurin und Kommunikationspartnerin konzipiert, mit dem ähnliche Prozesse der Bedeutungsherstellung wie die von Menschen erfolgen (Gunkel 2012; Guzman 2018; Guzman/Lewis 2020). Durch kommunikative Praktiken der Anthropomorphisierung werden sprachbasierte Maschinen als autonome und agentive Wesen konstruiert (Schneider 2021; Lind/Dickel 2023), welche einen erheblichen Einfluss darauf haben, wie Menschen mit Maschinen interagieren.

Vor diesem Hintergrund erforschen bspw. Beneteau et al. (2019), wie Nutzer*innen auf ein erfolgloses Prozessieren ihres gesprochenen Inputs reagieren. Mit dem Fokus auf einen sprachlichen Austausch beschreiben die Autor*innen dieses Phänomen als ein Zusammenbruch der Kommunikation (»communication breakdowns«), welcher durch diskursive Strategien, wie zum Beispiel die Verkürzung von Sätzen und die Überartikulation eines Befehls, seitens der Nutzer*innen behoben wird. Beneteau et al. (2019) bedienen somit die Idee einer kommunikativen Interaktion mit dem Sprachassistenzsystem *Alexa*, die der eines menschlichen sprachlichen Austauschs ähnelt: »Communication repair refers to the work of restoring shared understanding after conversational partners misunderstood each other« (Beneteau et al. 2019: 2). Somit wird hier ein Kommunikationsbegriff vertreten, der auf die Idee des gemeinsamen Verstehens und Symmetrie basiert. Bailey (2004) zufolge wird demnach eine Kommunikationsideologie verfolgt, in der gemeinsames Verständnis als »normal« und Missverstehen als Störung oder als ein Versagen gezeichnet wird (Bailey 2004: 395 in Pennycook 2017a: 92).

Diese Annahme des gemeinsamen Verständnisses, das für eine erfolgreiche Kommunikation notwendig zu sein scheint, entspringt einem universellen humanistischen Gedanken nach Gemeinsamkeit (Pennycook 2017a). Pennycook zufolge basieren diese Annahmen auf einem traditionellen Kommunikationsbegriff, der dem Grundgedanken des akkuraten Informationsaustauschs zwischen Sender und Empfänger unterliegt, wie er beispielsweise von Shannon und Weaver (1949) sowie Saussure (1916/2001) vertreten wurde.

Much of linguistic thought has been premised on a model of mutual understanding, of passing encoding messages back and forth from one head to

another, and doing so within a speech community with agreed norms for language use and comprehension. (Pennycook 2017a: 92)

Dementsprechend wird Sprache in den Köpfen der Menschen verortet und somit auf einen symbolischen Austausch reduziert, der auf dem Codieren und Decodieren von Nachrichten basiert. Wörter werden als in sich existierende Entitäten und Sprache als ein in sich geschlossenes System verstanden, das außerhalb des Menschen zu existieren scheint. Die Idee des gemeinsamen Verständnisses ergibt sich aus einer traditionellen Ansicht, die Sprache als ein »fest umgrenztes Objekt« (Metten 2014:11) erfasst. Diese Hypostasierung von Sprache, welche tief verankert in der semiologischen Linguistik Saussures ist, erfasst Sprache als ein a-historisches und statisches Zeichensystem und lässt dabei den Einfluss menschlichen Handelns unbeachtet:

Die individuellen Erfahrungen des Einzelnen sowie dessen konkretes Sprachhandeln verschwinden darin, ohne dass geklärt wird, wie das Sprachsystem seitens des Individuums tatsächlich und unabhängig dessen theoretischer Voraussetzung zu seiner lebensweltlichen Basis der Verständigung werden kann. (Metten 2014: 20)

Auch posthumanistische Ansätze, wie sie von Pennycook (2017a) vertreten werden, stellen sich gegen diese individualistischen Auffassungen von Sprache. Diese besagen, dass erfolgreiche Kommunikation nur durch das Teilen von Code innerhalb desselben Sprachsystems erfolgen kann. Während der Humanismus das Bild eines autonomen und individuell handelnden Subjekts zeichnet und damit eine Ideologie der Gleichheit verfolgt, betonen posthumanistische Ansätze Vielfalt, Unordnung und Komplexität (Pennycook 2017a). Wenn die Maschine nun als Kommunikationspartnerin skizziert wird, die »verstehet«, »halluziniert«, oder eben »missverstehet«, so werden, wie Sprachsysteme (vgl. Pennycook 2007), Mensch und Maschinen als abgegrenzte Entitäten behandelt. Folglich kann dies zu folgender Annahme führen: »Die Anwesenheit eines menschlichen Körpers ist längst nicht mehr nötig, damit Kommunikation als Kommunikation verstanden wird« (Dickel 2021: 27). Die Funktionsweise einer sprachbasierten Maschine basiert auf einem homogenen und denotativen Verständnis von Sprache, in dem Wörter als stabile und abstrakte Entitäten behandelt werden. In dieser Sichtweise wird Sprache als Mittel verstanden, um zwischen einer äußeren und einer inneren Realität zu vermitteln (Bender/Koller 2020). In Mensch-Maschine-Kommunikation liegt der Schwerpunkt da-

her meistens auf dem gesprochenen Wort anstatt der Handlungen eines Menschen, *mit* sowie *in* der Maschine.

Diese Vorstellung von Maschinen als autonome Wesen und abgegrenzte Objekte führt somit zu einer Verschleierung einer tiefgreifenden Komplexität, die auf menschliches Handeln in einer komplexen Welt zurückzuführen ist (vgl. Natale 2023; Crawford/Joler 2018). Eine theoretische Verschreibung von Sprache auf der die heutigen intelligenten Maschinen basieren, lässt die kulturelle (Schneider 2021) und »lebensweltliche Verflechtung« (Metten 2014: 7) einer Person außer Acht. Die Idee, dass Sprache auf »homogene Entitäten kollektiver Individuen« reduziert wird, kann demnach nur als eine »abstraktive Reduktion« menschlichen Handelns begriffen werden (Metten 2014: 31). Aus posthumaner sowie kulturwissenschaftlicher Sicht kann daher Sprache nicht von »Menschen, die sich miteinander verständigen« getrennt werden (Metten 2014: 27). Doch wie verständigen sich Menschen genau? Verfolgt man die Kernideen der *Languaging*-Perspektive (Cowley 2019; Thibault 2021; Kravchenko 2022) muss Sprache zunächst als metabolisch oder körperlich verstanden werden, bevor es als symbolisch betrachtet werden kann (Cowley 2009). Gemäß der *Languaging*-Perspektive wird Sprache als körperliche menschliche Aktivität verstanden, bei der das Gesagte eine wichtige Rolle spielt (Cowley 2019). *Languaging* geht somit über das gesprochene Wort hinaus und untersucht, wie Menschen in dialogischer Koordination auf die Bewegungen ihres Gegenübers reagieren, wie sie auf Körperbewegungen, Veränderungen in Stimmdynamiken oder Gesichtsausdrücke eingehen (Thibault 2021). Wie Menschen nun auf die gesprochenen Ausgaben einer Maschine reagieren, hängt von ihrer soziokulturellen Einbettung sowie von alltäglichen dialogischen Interaktionen mit anderen zusammen. Dieser Beitrag wendet sich gegen ein verfestigtes cartesianisches, mechanistisches und ding-zentriertes Weltbild (Shotter 2019), welches Objekte oder Maschinen als in sich existierende Einheiten versteht. Stattdessen konzentriert sich dieser Beitrag auf die dialogischen Beziehungen, die eine Maschine konstituieren – sowohl die der Nutzer*innen als auch die der Entwickler*innen der Systeme (vgl. Miceli et al. 2020). Anhand einer Fallstudie untersucht dieser Beitrag die *Heterogenität* der Sprechenden. Die Untersuchung konzentriert sich auf eine Teilnehmerin, und wie sie als aktive Beobachterin und kognitive Akteurin mit dem Sprachassistenzsystem *Alexa* zum ersten Mal auf Arabisch interagiert. Dabei wird untersucht, wie die Teilnehmerin genau auf Veränderungen in ihrer direkten Umgebung spontan reagiert. Des Weiteren werden die Reaktionen der Teilnehmerin mit den Entscheidungsprozessen der Entwickler*innen der verglichen (Hardesty 2022).

2 Sprache als menschliche Aktivität

Im folgenden Abschnitt werden die Kernpunkte der *Languaging*-Perspektive sowie die theoretische Konzeptualisierung der Sprechenden als beobachtende Personen erläutert. Die Heterogenität der Sprechenden, wie sie hier skizziert wird, ergibt sich aus einem ökologischen, dialogischen, verkörperten und diachronen Verständnis von Sprache. Es geht um die Frage, wie Sprechende spontan auf bestimmte Ereignisse in ihrer unmittelbaren Umgebung reagieren und diese für sich relevant machen. In diesem Zusammenhang konzentriert sich der Beitrag auch auf die soziokulturellen Kontexte, die den Entscheidungen der Sprechenden zugrunde liegen.

2.1 Languaging

Languaging konzentriert sich auf die Handlungen von Menschen, d.h. vereinfacht gesagt, wie Menschen Sprache *machen* (Cowley 2019). Dabei stützt sich die Perspektive im Kern auf die Annahme, dass menschliche Kognition ständigen Organismus-Umwelt-Kopplungen unterliegt und somit aus der direkten Interaktion mit der Umwelt emergiert. Ein Organismus ist unmittelbar umweltgebunden und reagiert situativ auf Veränderungen seiner direkten physischen Umgebung (Chemero 2009; Gibson 1986). Wie Kognition sollte auch Sprache als menschliche Aktivität betrachtet werden.

Im Gegensatz zu einer traditionellen Sprachbetrachtung werden in der *Languaging*-Perspektive die Prosodie und die körperlichen Aspekte der phonetischen Aktivität als Grundlage sprachlichen Handelns behandelt (Cowley 2014; Cowley/Harvey 2016; Thibault 2011, 2021). In diesem Sinne können Wörter aus dieser Perspektive nicht als vorhandene bedeutungstragende Entitäten verstanden werden, die Teil eines geschlossenen homogenen Sprachsystems sind und eine arbiträre Beziehung zwischen Form und Bedeutung herstellen. Stattdessen werden sie in ihrem Kern zunächst als phonetische und artikulatorische Gesten betrachtet (Cowley/Harvey 2016; Thibault 2021). Um diesen flüchtigen Charakter vokaler Bewegungen zu beschreiben, bevorzugt Cowley (2014) den Begriff *Wording*. Damit wird betont, dass Wörter durch den Vokaltrakt zunächst immer wieder neu körperlich produziert werden und daher vor allem als Vokalisationen zu verstehen sind. Aus diesem Grund kann nie von einer absoluten Wiederholung von Wörtern gesprochen werden (Cowley/Nash 2013). Vielmehr muss sich die Untersuchung auf die zugrundeliegenden Zeitskalen konzentrieren, die Handlungen auf einer physischen

Echtzeitskala eingrenzen. Somit ist *Languaging* nicht nur als verkörpert zu verstehen, sondern auch als multiskalar¹. Sobald die Sprache in den menschlichen Handlungen verankert ist, muss man den Fokus auf das Zusammenspiel verschiedener räumlich-zeitlicher Skalen verlagern, die von Piko- und Mikrodynamiken körperlicher Aktivität bis zu den langsameren zeitlichen Skalen der kulturellen Einbettung einer Person reichen (Cowley 2014, Thibault 2011). Somit durchdringen die unterliegenden soziokulturellen Einflüsse oder die Geschichte bestimmter Gemeinschaften die menschliche Aktivität in Echtzeit. Diese bestimmen, dass vokale Dynamiken als Worte eines bestimmten Sprachsystems (z.B. Deutsch oder Englisch) zugehörig verstanden werden. Wenn Menschen Sprache *machen*, dann greifen sie über mehrere Zeitebenen hinweg auf vergangene Einflüsse zurück.

Schaut man sich nun genau an wie Menschen sich miteinander verständigen (Metten 2014), dann wird ersichtlich, dass Menschen auf die kleinsten

-
- 1 Der Einfluss der Zeitlichkeit in der Sprache kann auf mehrere Disziplinen zurückgeführt werden. Sie alle haben einen signifikanten Einfluss auf die *Languaging*-Perspektive. So ist die *Distributed Language*-Perspektive aus den Kernideen der integrativen Linguistik hervorgegangen, die mit Aspekten der *Distributed Cognition*-Perspektive vermischt wurden. Erstere hat die Unterscheidung zwischen Sprache (oder sprachlichem Handeln) erster Ordnung und Sprache zweiter Ordnung hervorgebracht (Love 2004). Dabei wird Sprache erster Ordnung primär als physische Aktivität verstanden, während Sprache zweiter Ordnung sich auf den denotativen und reduktiven Charakter von Sprache bezieht. So werden Semantik, Syntax, Morphologie etc. der Sprache zweiter Ordnung zugeordnet und als spezifische soziokulturelle Beschränkungen verstanden (Thibault 2011). Die *Distributed Cognition*-Perspektive betont insbesondere die soziokulturelle Einbettung eines kognitiven Systems und wie diese das menschliche Handeln bestimmt. Dementsprechend muss Kognition als verteilt verstanden werden: zwischen Menschen, Umwelt und Zeit. Die Perspektive von Hutchins (1995: 2006) bezieht sich jedoch auf ein ganzes kognitives System und verfolgt somit keine personen-zentrierte Sichtweise (zur Kritik siehe Giere 2004). Auch der dialogische Ansatz zur Sprache, wie er von Linell (2009) vertreten wird, hat einen erheblichen Einfluss auf den zeitlichen Aspekt der linguistischen Perspektive. Mit Begriffen wie *Double Dialogicality* und *Situation-transcending practices* hebt Linell den zeitlichen und den diachronen Aspekt von Sprache hervor. Ersteres bezieht sich nicht nur auf den dialogischen Charakter von Sprache, d.h. dass Sprache nur in Bezug auf lebendige Beziehungen zu anderen verstanden werden kann, sondern auch darauf, dass, wenn Menschen sprechen, immer auch Aspekte der Vergangenheit zur Sprache kommen. So bestimmen u.a. Traditionen und Gewohnheiten immer auch situative Praktiken, was Linell mit dem letzten Begriff meint. All diese Ansätze sind Vorläufer dessen, was Forscher*innen in der *Languaging*-Perspektive unter Multiskalarität verstehen.

körperlichen Veränderungen ihres Gegenübers reagieren (Blair/Cowley 2003; Thibault 2011). So argumentiert Thibault (2011), dass Sprache sich aus körperlicher dialogischer Koordination ergibt. Als beobachtende Wesen reagieren Menschen genau auf die kleinsten Veränderungen in den Bewegungen ihres Gegenübers, sei es Veränderungen der Mimik, Gestik oder Stimmbewegungen, welche sich über Piko-, Mikro- und Enchro-Skala erstrecken.² *Languaging* findet auf der Pikoskala (weniger als 200 ms) und der Mikroskala (bis zu 500 ms) der Echtzeit-Körperdynamik statt (Thibault 2011). Die Perspektive des *Languaging* konzentriert sich nicht darauf wie Menschen ein Sprachsystem verwenden, sondern hebt körperliche Dynamiken hervor, und wie Menschen auf diese reagieren. *Languaging* ist daher in der körperlichen, dialogischen Koordination verwurzelt. Die Handlungen von Menschen müssen als Reaktion auf die Handlungen anderer Menschen und in Wechselbeziehung zu ihnen verstanden werden (Linell 2009). Das Reagieren auf diese Echtzeitergebnisse der Äußerungsaktivität ist wesentlich für das menschliche Werden und begründet das menschliche Auftreten. Für Blair und Cowley (2003: 132) bedeutet das, »what human bodies and voices do together affect how we conceptualize our encultured worlds« und ist mikrokognitiv – es bewirkt *Veränderungen* in einer Person. Durch das Sprechen greifen die Menschen ständig auf vergangene Erfahrungen zurück. Der Aspekt der Multiskalarität der Sprache verdeutlicht, wie Menschen die Bewegungen anderer einbeziehen und darauf aufbauen, indem sie kontinuierlich langsamere Zeitskalen integrieren, die es ihnen ermöglichen, Äußerungen und Aktivitäten als ähnliche Muster wahrzunehmen. Sprache wird also nicht von Menschen genutzt, sondern ständig, oftmals in dialogischer Koordination, hervorgebracht. Verankert man Sprache nun in sprachlich-körperlichem Handeln, wird der materielle Charakter von Sprache deutlich. Sprache kann demnach nicht als etwas Immaterielles verstanden werden, das außerhalb des Menschen existiert, sondern muss immer in Bezug auf ein handelndes Wesen betrachtet werden.

Vielmehr muss sich die Untersuchung auf die zugrundeliegenden Zeitskalen konzentrieren, die Handlungen auf einer physischen Echtzeitskala eingrenzen. Somit ist *Languaging* nicht nur als verkörpert zu verstehen, sondern

2 Die Piko-Skala umfasst Körperdynamiken, die von Millisekunden bis zu Zehntelsekunden dauern, die Mikro-Skala der Körperdynamik zeichnet »Sagen und Tun« nach, die von Zehntelsekunden bis zu Sekunden gemessen werden können, und die Enchro-Skala schließlich ermöglicht es, situierte soziale Ereignisse zu verfolgen, die in Sekunden, Minuten usw. gemessen werden können (Thibault 2011, 215–216).

auch als multiskalar und dialogisch. Diachrone Aspekte, die unter anderem die autobiographische Geschichte und die Geschichte bestimmter Gemeinschaften hervorbringen, können daher nicht von situativen Ereignissen getrennt werden. Dementsprechend werden denotative sprachliche Bezeichnungen als Ergebnis vergangener koordinativer Aktivitäten verstanden. Im Laufe der Zeit lernen Menschen (insbesondere Kleinkinder), Vokalisierungen als Muster wahrzunehmen und auf bestimmte Weise darauf zu reagieren. In der integrativen Linguistik werden Wörter daher als ›repeatables‹ bezeichnet (vgl. van den Herik 2017). Durch die dialogische rekursive Koordination mit anderen (z. B. in Kleinkind-Eltern-Interaktionen) lernen Menschen, auf bestimmte Muster körperlicher Aktivität in einer gewissen Art und Weise zu reagieren. Cowley und Kuhle (2020) vergleichen daher den symbolischen Umgang mit Sprache mit dem Umgang mit kulturellen Werkzeugen. Dementsprechend verweist die *Languaging*-Perspektive auf den heterogenen Charakter von Sprache. So kann *Languaging* als ein Zusammenspiel von Dialogizität, Zeitlichkeit und Verkörperung verstanden werden. Betrachtet man nun Mensch-Maschine-Kommunikation von der *Languaging*-Perspektive aus, fällt die Untersuchung nicht nur auf das gesprochene Wort, sondern vielmehr darauf, wie Nutzer*innen auf die gesprochene Ausgabe reagieren. Wenn Hector (2023) nun die Funktionsweise der Smart Speaker als das Senden von »sprachlichen Zeichen« begreift, »denen von den Beteiligten Sinn zugeschrieben wird«, ergibt sich die Frage nach den dialogischen und ökologischen Einflüssen, die in diese Sinnbeschreibung einfließen (ebd.: 216). Im Fallbeispiel wird daher genauer untersucht, wie die Teilnehmerin nicht nur auf die Ausgabe des Sprachassistenzsystems reagiert, sondern wie sie mit ihrem Partner über die Ausgaben spricht und wie die dialogische Koordination mit ihrem Partner die von Hector (2023) erwähnte Sinnzuschreibung beeinflusst. Suchman (2007) verweist auf die Asymmetrie in Mensch-Maschine-Kommunikation, indem sie darauf aufmerksam macht wie vergangene Erfahrungen und das Projizieren zukünftiger Ereignisse menschliche Aktivität im Hier und Jetzt beeinflusst.

Es sind auch diese beschriebenen dialogischen und lebendigen Beziehungen, die eine Maschine konstituieren (Shotter 2019). So argumentiert Shotter, dass die Rolle des lebendigen Körpers und die lebendigen Beziehungen zu anderen und der Umwelt nicht ignoriert werden können, wenn wir die Aktivitäten eines Roboters verstehen wollen. Aus diesem Grund werden die Entscheidungsprozesse der Entwickler*innen in dem Fallbeispiel mit den Handlungen der Teilnehmerin gegenübergestellt.

2.2 Der Mensch als beobachtende und kognitiv agierende Person

Esposito (2022) bricht mit dem Symmetrie-Gedanken in Mensch-Maschine-Kommunikation, indem sie Kommunikation nach dem Verständnis von Luhmann definiert und erfolgreiche Kommunikation der Interpretationsfähigkeit der*den Empfänger*innen zuschreibt. In dieser Annahme werden Informationen von jedem*r Beobachter*in verschieden betrachtet: »Luhmann argues, information is different for everyone and always relative to a specific observer. But a common identity of information among participants is itself not required for communication« (Esposito 2022: 7). Die beobachtende Person spielt eine zentrale Rolle in Mensch-Maschine-Kommunikation. Zwei Punkte aus Maturanas biologischem Ansatz zu Kognition und Sprache sind deshalb relevant, die im Weiteren die Analyse und Sichtweise auf den Umgang mit Sprachassistenzsystemen in diesem Beitrag bestimmen: Erstens ist es wichtig anzuerkennen, dass Sprache als Verhalten gesehen werden sollte, und zweitens muss sich dabei die Untersuchung auf die Handlungen und das, was Maturana als Unterscheidungen bezeichnet, konzentrieren, die beobachtende Personen treffen. Maturana (1985: 276) beschreibt eine beobachtende Person als »ein menschliches Wesen, ein lebendes System, das Unterscheidungen treffen kann und das, was sie unterscheidet, als Einheit abgrenzen kann, d.h., als eine von dem Beobachter selbst verschiedene Größe«. Sprache entstammt demnach aus der Koordination zweier Sprechender in einer sogenannten konsensuellen Domäne, d.h. einer ähnlichen Lebenswelt. Das Bezeichnen von externen Entitäten entsteht aus menschlicher Erfahrung durch das ständige Aufbauen auf dem Gesagten und dem Inkorporieren eigener Erfahrungen. Diese Annahme der Rekursivität in Koordination ist ein wesentlicher Bestandteil des biologischen Ansatzes in Bezug auf Sprache und Kognition (Gahrn-Andersen 2019; Raimondi 2019) und bestimmt im Wesentlichen, was Menschen auf der einen Seite als Realität und auf der anderen Seite als Sprache wahrnehmen. In diesem Zusammenhang ist Realität nicht als eine objektive Abgrenzung zu verstehen, die außerhalb einer beobachtenden Person existiert. Vielmehr emergiert Realität aus den gelebten Erfahrungen dieser Person, die wiederum Teil bestimmter sprachlicher Gemeinschaften ist (vgl. Shotter 2019). Dementsprechend erklärt Kravchenko (2022: 209): »Language is our existential domain in which we happen and become what we are«.

Die Sprechenden werden in diesem Beitrag als aktive beobachtende Personen und als menschliche kognitive Akteur*innen konzeptualisiert. Letzteres bezieht sich auf Gieres (2004) Kritik an Hutchins' (1995) *Distributed Cognition*

on. So untersuchte Hutchins bspw. die Navigation eines Schiffes in den Hafen. Im Gegensatz zu den klassischen Auffassungen der Kognition beobachtet Hutchins, dass die kognitiven Handlungen, die die Navigation konstituieren nicht einer Person allein (z.B. dem Kapitän) zugeschrieben werden kann, sondern sich über andere Akteur*innen, technische Ressourcen und eine soziokulturelle Welt erstrecken. In Hutchins' Auffassung findet Kognition innerhalb eines distribuierten kognitiven Systems statt. Für Giere (2004) wird die aktive Rolle der kognitiven Akteur*innen außer Acht gelassen. Ein kognitives System könnte ohne die Handlungen von menschlichen Akteur*innen nicht existieren. So sind es die kognitiven Akteur*innen, die ein verteiltes distribuiertes kognitives System hervorbringen, indem sie auf Geschehnisse und Dinge in ihrer direkten Umwelt vom Standpunkt ihrer gelebten Erfahrung, zumeist durch verbale Handlungen, reagieren (Cowley/Vallée-Tourrangeau 2013).

Für Maturana sind menschliche Wesen zuallererst sprechende Organismen (vgl. Kravchenko 2022), die die Welt beobachten und Unterscheidungen aus ihren Beobachtungen ziehen können. Angewendet auf den kommunikativen Mensch-Maschine-Kontext dieses Beitrags bedeutet das, dass die Nutzer*innen von Smart Speakern sowie die Entwickler*innen als Beobachter*innen zu verstehen sind, deren körperliche Dynamiken und Beobachtungs- und Unterscheidungsweisen genauer zu untersuchen sind. Dementsprechend verfolgt dieser Beitrag die Frage, welche Ereignisse und soziokulturellen Einbettungen die körperlichen Aktivitäten der Nutzer*innen und Entwickler*innen in Mensch-Maschine-Kommunikation untermauern. Während *Languaging* die Heterogenität von Sprache beschreibt, beziehe ich mich auf die Heterogenität der Sprechenden. Das bedeutet, welche Ereignisse die Wahrnehmung der Nutzer*innen auf die gesprochene Ausgabe der Sprachassistenzsysteme bestimmen.

3 Sprachassistenzsysteme: Ein Konglomerat menschlicher Einflüsse

Wendet man sich der Architektur und dem Design der dialogischen Sprachsysteme zu, kann die Komplexität menschlicher Einflüsse nicht ignoriert werden. Fortschreitende technologische Entwicklungen im Bereich der Künstlichen Intelligenz basieren auf dem Zugang zu großen Mengen von menschenproduzierten Daten aus dem *Semantic Web*, und auf erhöhter Computerkapazität (Mühlhoff 2020). So speisen sich Bildererkennungs-, Sprach- und stimm-

basierte Modelle aus menschlichen Entscheidungs- und Gedankenprozessen. Sprachbasierte Künstliche Intelligenzen können daher nur in Bezug auf die ihnen zugrundeliegenden menschlichen Einflüsse verstanden werden. Ein*e Nutzer*in agiert auf technischer Ebene mit den technologischen Einschränkungen des Systems, das im Kern aus binärem Code besteht (Svensson 2022), und zudem aus vergangenen Entscheidungsprozessen menschlicher kognitiver Agenten.

Betrachtet man den technischen Aufbau und das Design der Smart Speaker, kann vereinfacht festgestellt werden, dass Smart Speaker oder Dialogsysteme in ihrem Kern durch die Verarbeitung von Text funktionieren:

The basic idea is that the user makes a request through the voice-activated device, and then, the voice request gets streamed through the cloud, and here voice gets converted into text. Then, the text request goes to the backend and after processing, the backend replies with a text response. Finally, the text response goes through the cloud and gets transformed into voice, which will be streamed back to the user. (Teropoulos/Satratzemi 2020: 474)

Ein funktionsfähiges Dialogsystem besteht aus zwei Kernkomponenten: 1) Automatische Spracherkennungssysteme (ASR) sowie 2) Spracherkennungssystem und Dialogmanagementsystem (vgl. McTear et al. 2016; Jokinen/McTear 2009). Jokinen und McTear (2009: 5) verorten die Prozesse der automatischen Spracherkennung (1) und der maschinellen Sprachverarbeitung (2) in der Kategorie Sprachverstehen oder *Speech Understanding*. Die Materialität der Sprache und der bereits erwähnte flüchtige Charakter von sprachlicher Aktivität bergen große Herausforderungen für ein Computersystem. So müssen zunächst akustische Signale menschlicher Sprache in elektronische Signale umgewandelt werden, die dann zu binären Zeichen codiert werden (McArthur 2020), sodass diese Signale von der Maschine bearbeitet werden können. McArthur verweist hier im vereinfachten Sinne auf digitale Algorithmen, die statistische Methoden implementieren und die entstandenen Muster der binären Zeichen bestimmten Phonemen oder Wörtern zuordnen. ASR-Prozesse sind daher als probabilistische Prozesse zu verstehen, die den gesprochenen Input Mustern in trainierten Datensätzen gegenüberstellt (McTear et al. 2016). Dabei bestimmen die vergangenen Entscheidungsprozesse der Entwickler*innen, welche Datensätze sie für das Trainieren der Modelle nutzen.

Spracherkennungs- und Sprachverarbeitungsprozesse sind in der Architektur des Smart Speaker eng miteinander verbunden: Akustische Signale

werden in Token und Wörter konvertiert, deren semantischer Inhalt dann durch Sprachverarbeitungsprozesse analysiert wird, um einen Befehl gezielt ausführen zu können. Auch hier beruht die Technologie auf dem Erkennen von Mustern. Die Hauptaufgabe dieser Prozesse ist es, den Inhalt der Äußerungen der Nutzer*innen zu analysieren und eine Repräsentation der Bedeutung zu erstellen (McTear et al. 2016). So sind diese Systeme mit sogenannten Dialog- oder *Information Retrieval*-Systemen verbunden (McTear et al. 2016). Dabei ist zu beachten, dass Sprachverarbeitungsprozesse aus verschiedenen Technologien bestehen, um verschiedene Arten der Äußerungen filtern zu können, beispielsweise die Unterscheidung ob es sich um einen Befehl oder um eine Frage handelt (ebd.). Da Smart Speaker auf Cloud-Computing basieren, also mit dem Internet verbunden sind (McTear et al. 2016, Natale/Cooke 2021), werden Schlüsselwörter direkt von der semantischen Struktur des eingegebenen Textes abgeleitet, so zum Beispiel ›Standort‹ oder ›Wetter‹, um dann diese mit Inhalten im Web zu verbinden.

Doch während in großen Sprachmodellen wie ChatGPT Sprachverarbeitungs- und Spracherkennungsprozesse auf neuronalen Netzwerken basieren, wird in Smart Speakern oftmals noch auf das manuelle Erstellen von Inhalten zurückgegriffen. So sind viele Antworten direkt von den Designer*innen gescrriptet, wie zum Beispiel Witze (Stroda 2020 in Natale 2020). Auch werden die Dialogsysteme auf bestimmte Datensätze trainiert: Amazon nutzt den multilingualen Datensatz *Mintaka* als Frage-Antwort System (Sen et al. 2022). Hier wird das System trainiert, auf bestimmte Fragen direkt antworten zu können. Das Trainieren des Datensatzes ist abhängig von sogenannten *Crowd Workers*: Diese müssen fünf Fragen frei aufschreiben. Die Antworten zu den Fragen werden mit Daten von Wikipedia verlinkt. *Crowd Workers* werden dann Fragen-Antworten Paare vorgelegt, um wichtige Aspekte der Antworten hervorzuheben, diese in den Wikidaten zu suchen und mit den korrekten URLs zu verbinden. Weitere Mitarbeitende müssen dann die Antworten annotieren um zu gewährleisten, dass eine korrekte Ausgabe vom System gegeben wird. Zusätzlich werden Übersetzer*innen angestellt um Fragen auf Englisch in Arabisch, Französisch, Deutsch, Hindi, Italienisch, Japanisch, Portugiesisch und Spanisch zu übersetzen. So wurde ein Datensatz mit ungefähr 180.000 Fragen erstellt (Sen et al. 2022). Das Beispiel des Datensatzes *Mintaka* zeigt deutlich auf wie das Wissen, das einer Maschine zugeschrieben wird, aus den kognitiven Handlungen verschiedener Akteur*innen stammt. Dabei sind auch die *Crowd Worker* als soziale und dialogische Wesen zu verstehen, die zum einen vom Standpunkt ihrer soziokulturellen Einbettung aus und zum

anderen mit den Erwartungen ihrer Auftraggeber*innen agieren (vgl. Miceli et al. 2020).

4 Methode und Fallbeispiel

Der *Languaging*-Perspektive folgend, die Sprache in ihrer Heterogenität anerkennt, liegt der Schwerpunkt in diesem Beitrag auf die Sprechenden als beobachtende Personen und kognitive Akteur*innen. In einer videobasierenden Ethnographie wird dabei nicht nur untersucht, wie eine Nutzerin des Sprachassistenzsystems *Alexa* situativ mit dem Gerät interagiert. Vielmehr konzentriert sich die Untersuchung darauf, wie die soziokulturelle Einbettung der Nutzerin Ereignisse aus ihrer direkten Umwelt, wie die Reaktionen ihres Partners, ihre Wahrnehmung und Reaktion auf die gesprochene Ausgabe des Geräts beeinflusst. In einer qualitativen videobasierten Langzeitstudie wird dabei die Untrennbarkeit von diachronen und synchronen Aspekten menschlichen Handelns untersucht. Die hier vorgestellte Studie basiert auf der Methode der kognitiven Ethnographie, wie sie von Hutchins (1995) und Alač und Hutchins (2004) vertreten wird. Um verteilte kognitive Prozesse zu verstehen, werden die Interaktionen zwischen verschiedenen Akteur*innen, technischen Artefakten und dem Sozialen holistisch untersucht. So stehen u. a. die direkte Umgebung und die darin vorzufindenden materiellen Objekte sowie Interaktionen mit anderen Akteur*innen im Fokus der Untersuchung, sowie wie aus dem Zwischenspiel dieser Aspekte Kognition emergiert. So bedient sich auch die vorliegende Fallstudie der Langzeitbeobachtung einer Teilnehmerin. In Kombination mit detaillierten Mikroanalysen werden die Handlungen und Reaktionen der Teilnehmerin auf das Sprachassistenzsystem und ihrer Umgebung genaustens untersucht. Die Methode in dem vorliegenden Beitrag verfolgt dabei Alač und Hutchins' (2004: 633) Annahme, dass »[t]he two sides of the research, the micro analysis and the larger ethnography, are interdependent«. In der kognitiven Ethnographie liegt der Schwerpunkt auf kognitiven Prozessen und nicht auf Sprachhandlungen. Im Gegensatz dazu geht die hier durchgeführte Mikroanalyse davon aus, dass Sprache durch körperliche Koordination hervorgebracht wird. Durch detaillierte Mikroanalysen wird das Zusammenspiel körperlicher Dynamiken in Echtzeitinteraktionen tiefgreifend untersucht (Thibault 2011). Um die systemische Einbettung der untersuchten Person genau zu verstehen, stützt sich die Untersuchung nicht

nur auf einen Datensatz von Videoaufnahmen, sondern auch auf Interviews mit der Teilnehmerin sowie Feldnotizen.

Im Mittelpunkt der Fallstudie steht eine Teilnehmerin, die im Rahmen der Arbeit Anna³ genannt wird. Anna stammt aus Jordanien und hat an einer deutsch-libanesischen Universität studiert, was ihr ermöglicht hat, Sprachwissenschaften in Deutschland zu studieren. Neben Arabisch, das sie selbst als Amman- und jordanisches Arabisch bezeichnet, spricht Anna fließend Deutsch und Englisch. Sie lebt mit ihrem Partner Daniel zusammen, der aus Portugal stammt. Gemeinsam sprechen sie Englisch.

Die Studie ist wie folgt aufgebaut: Die Teilnehmerin erhielt eine Kamera (GoPro Hero Black 10), eine entsprechende Speicherkarte und als Sprachassistenten den Echo Dot (5. Generation, 2022). Die Untersuchung fand vom 11. Januar bis 28. Februar 2023 statt. In diesem Zeitraum wurde die Teilnehmerin zu Beginn, in der Mitte und am Ende der Studie interviewt, um von ihren erlebten Erfahrungen mit dem Gerät zu erzählen. In einem Einführungsgespräch am 13. Dezember 2022 wurde der Teilnehmerin der genaue Ablauf der Studie erklärt. Das Ziel der Studie wurde wie folgt angegeben, wie es auch auf dem Informationszettel zur Studie zu finden ist: »Diese Pilotstudie ist eine qualitative Fallstudie, die zum Ziel hat zu beobachten, wie Menschen in ihrem häuslichen Umfeld mit Sprachassistenten interagieren.« In dem Gespräch wurde die Teilnehmerin darüber informiert, dass sie mit der ihr zur Verfügung gestellten Kamera ihre Interaktion mit dem Gerät filmen soll. Dabei wurde es der Teilnehmerin überlassen, wie sie die Videoaufnahmen durchführen und welche Aufnahmen sie zur Verfügung stellen möchte. Des Weiteren wurden mit der Teilnehmerin Termine für die weiteren Interviews abgesprochen, bei denen sie bereits erstes Videomaterial zur Sichtung übergab. Im Rahmen der Studie filmte die Teilnehmerin selbst die Interaktion mit dem Sprachassistenten in ihrer häuslichen Umgebung. Eine videobasierte Ethnographie in einer häuslichen Umgebung birgt viele Herausforderungen u.a. die Gewährung der Privatsphäre aller Teilnehmenden. Aus diesem Grund beziehen sich viele interaktionsanalytische Studien (z.B. Hector 2023; Porcheron et al. 2018) auf Audioaufzeichnungen, um zu untersuchen wie Menschen Sprachassistentensysteme

3 Im Rahmen dieser Studie wurden alle Namen der teilnehmenden Personen anonymisiert. Es wurde eine informierte Zustimmung der Teilnehmerin eingeholt. In dieser wurde die Teilnehmerin über die Studie informiert, ohne das eigentliche Forschungsziel zu offenbaren, mögliche Risiken und den Studienablauf informiert. Des Weiteren wurde separat Einwilligung für die Nutzung der Videodaten eingeholt.

me in ihre Alltagspraktiken integrieren. Da der theoretische und methodologische Schwerpunkt der Studie auf der *Languaging*-Perspektive liegt und die Betrachtung über das gesprochene Wort hinausgeht, ist eine detaillierte Untersuchung körperlicher Bewegungen notwendig, die auf Videoaufnahmen basiert. Obwohl die Forscherin nicht direkt anwesend war, ist sie dennoch durch das eigenhändige Filmen in den Handlungen der Teilnehmerin präsent. Ein performativer Charakter in den Videoaufnahmen kann nicht ausgeschlossen werden.

Gegenstand der Untersuchung ist ein Datenausschnitt aus einem Video, der zeigt wie die Teilnehmerin einen bestimmten Befehl auf Arabisch an das Gerät richtet⁴. Dabei wird genauer untersucht, wie der Befehl durch die lebensweltliche Verflechtung der Teilnehmerin zustande kommt. Hierbei spielt die Interaktion mit ihrem Partner Daniel sowie die soziokulturelle Einbettung der Teilnehmerin eine erhebliche Rolle. In der vorliegenden Studie wird die Teilnehmerin als aktive Beobachterin und kognitive Akteurin, wie in Abschnitt drei beschrieben, verstanden. Mithilfe von Screenshots der Videoaufnahmen, dazugehörigen Transkripten⁵ und Interviews wird Annas sprachliche Handlungen genauestens untersucht.

4.1 »Alexa, make the sound of a chicken!«: Die dynamische Entfaltung eines Befehls

Anna sitzt am Esstisch in ihrem Wohnzimmer. Der Smart Speaker befindet sich zu ihrer Linken und ihr Smartphone ist vor ihr platziert. Über eine Smartphone-Anwendung steuert Anna direkt die Spracheinstellungen des Geräts (Hector 2023). Während Anna nun direkt mit dem Gerät interagiert, befindet sich Daniel in der Küche und ist somit zunächst nicht sichtbar (siehe Abbildung 1).

-
- 4 Auf Nachfrage agiert Anna in dieser Sequenz erstmalig mit dem Sprachassistenzsystem Alexa. Zuvor fand die Interaktion mit dem Gerät nur auf Englisch statt, sodass Anna sowie ihr Partner das Gerät gemeinsam bedienen konnten. Da die arabische Sprachfunktion erst im Dezember 2021 von Amazon implementiert wurde (Hardesty 2022), bat die Forscherin Anna die arabische Sprachfunktion zu testen.
- 5 Ich bedanke mich bei Mohammad Alenezi (Kuwait University) für die Übersetzung aus dem Arabischen ins Englische sowie Erklärungen zu den sprachlichen Variationen, die innerhalb des Transkripts auftraten.

Abb. 1: Räumliche Organisation und Einrichtung der Umgebung



Nachdem Anna bereits einige Befehle auf Arabisch an den Smart Speaker gerichtet hat und sich somit einen Eindruck über die Qualität der sprachlichen Ausgabe machen konnte, blickt Anna nun auf ihr Telefon. Vermutlich interagiert sie mit der Smartphone Anwendung, um die sprachlichen Einstellungen zu kontrollieren. In Minute 01:15 des Videos bewegt sich Daniel aus der Küche in Richtung Wohnzimmer zu seinem Arbeitsplatz, der zu Annas rechten Seite liegt. Als dieser nun aus der Küche tritt, äußert er folgenden Befehl in einer imperativen Satzstruktur (Barthel et al. 2023) auf Englisch: »Alexa, make the sound of the chicken!« Daniel blickt sowohl auf den Smart Speaker als auch auf Anna.

Abb. 2: Daniel betritt das Wohnzimmer und richtet Befehl an Smart Speaker



Da ein stationäres Sprachassistenzsystem durch lautliche Eingaben bedient werden muss und lautliche Kommunikation somit hörbar wird, ist der Zugang zu dem Gerät für alle im Haushalt lebenden Personen freizugänglich (Porcheron et al. 2017). So lenkt Daniel Annas Fokus von ihrem Smartphone auf den Smart Speaker. Aus ökologischer Sicht reagiert Anna nun auf direkte Ereignisse in ihrer akustischen Umgebung. Auch wenn Anna nicht direkt mit Daniel interagiert, sondern ihren Blick für ungefähr eine Sekunde auf den Smart Speaker richtet, bestimmt Daniels Äußerung Annas weiteren Interaktionsverlauf mit dem Gerät.

Anna antizipiert nun die lautliche Ausgabe des Smart Speakers. Währenddessen verfolgt sie mit ihrem Blick Daniels Bewegungen Richtung Tisch, der eine Tasse rechts neben ihr abstellt (siehe Abbildung 3, Bild A). So wirkt sich Daniels Präsenz in zweierlei Hinsicht auf Anna aus: 1) seine sprachliche Äußerung löst eine Veränderung in Annas körperlichen Verhalten aus, indem sie sich zu dem Smart Speaker wendet und ihre eigentliche Handlung nicht weiter ausführt und 2) Daniels Bewegungen durch den Raum sorgen dafür, dass Anna sich wieder vom Smart Speaker abwendet. Diese Veränderung ihrer lautlichen und physischen Umgebung führt zu einer Überbrückung von vergangenen Ereignissen zu neuen Ereignissen, die in Annas Ganzkörperhaltung deutlich erkennbar ist: Mit ihrem rechten Finger verharnt Anna noch immer über ihr Smartphone und reifiziert somit ein vergangenes Ereignis (als sie mit der Smartphone Anwendung agierte), während ihre Blickrichtung zum Smart Speaker verdeutlicht, dass sie sich nun erwartungsvoll einem neuen Ereignis zuwendet. Somit bildet Daniels Äußerung den Rahmen für Annas bevorstehende Koordination mit dem Gerät. Etwa 2 Sekunden später gibt der Smart Speaker eine lautliche Ausgabe auf Arabisch wieder.

Abb. 3: Annas körperliche Reaktion auf Veränderungen in ihrer direkten lautlichen und physischen Umgebung



Als Daniel sich auf seinen Arbeitsplatz begibt, verharrt Anna auf der rechten Seite, wie in Bild B in Abbildung 3 deutlich erkennbar ist. Auch wenn ihr Blick von der Maschine abgewandt ist, ist zu erkennen wie Anna sich auf die lautliche Ausgabe der Maschine konzentriert: sie schaut nach rechts, bleibt aber größtenteils regungslos in ihren Bewegungen. Sie reagiert lediglich mit subtilen Veränderungen in ihrem Gesichtsausdruck auf das Gehörte. Ihre Mimik deutet darauf hin, dass sie das Gehörte mit ihren eigenen sozial bedingten Erfahrungen (d.h., ein aktiver Teil ihrer Sprachgemeinschaft zu sein) verbindet (Cowley 2019). Während in dialogischer Koordination ein Gegenüber auf diese subtilen körperlichen Bewegungen eingehen könnte (vgl. Blair/Cowley 2003; Thibault 2011), erstreckt sich die Ausgabe des Geräts über einen Zeitraum von insgesamt 18 Sekunden ohne Unterbrechung (Exzerpt 1). Hier ist der Kontrast zwischen dynamischem menschlichen Handeln und der statischen Konzeption einer Maschine deutlich erkennbar.

Alexa:

ElaiKi ajaabatan min mostaXdami Alexa [inaudible]. Momkin An yakoon ad-dajaaj Saxibun juddan. YaSeeH laylan wa nahaaran wa laysa faQaT ›Enda shurooq Ashams. Yumkin li addeek taHreek [inaudible] bi akmalihaa wa aS-SouS yakbur wa yaSruX.

(Here's the answer from Alexa's user account [unintelligible]. Chickens can be very loud. The rooster crows day and night and not just at sunrise. The rooster can move the whole [unintelligible] and the chicks grow and crow.)

(Hier die Antwort des Alexa Nutzer Konto [unverständlich]. Hühner können sehr laut sein. Der Hahn kräht am Tag und in der Nacht und nicht nur am Sonnenuntergang. Der Hahn kann sich den ganzen [unverständlich] und die Küken wachsen und krähen.)

Anna orientiert sich immer noch nach rechts, folgt jedoch der sprachlichen Ausgabe des Smart Speakers. Nach etwa 4 Sekunden zieht sie ihre Augenbrauen zusammen (Abbildung 3, Bild B). Durch diese subtilen Veränderungen in ihrem Gesichtsausdruck markiert Anna bestimmte Aspekte der Ausgabe des Smart Speaker, die für sie relevant sind. Sie bringt somit ihre eigenen Beobachtungen und damit Interpretationen zum Vorschein (vgl. Maturana 1988). Nochmals bewegt sie die Ausgabe des Geräts dazu, ihre körperliche Position zu ändern (Abbildung 3, Bild C): So wendet Anna sich nun in Richtung des Smart Speakers, hebt ihren Kopf und richtet ihren Blick zur Decke. Nach nur 1,7 Sekunden bewertet sie ihren Eindruck der sprachlichen Ausgabe und beschreibt

diesen als »poetisch« (»it's very poetic«). Annas körperliches Verhalten zeigt hier auf, wie sie als aktive Beobachterin sowie als kognitive Agentin auf unterschiedliche Art und Weise auf die Ereignisse in ihrer physischen und lautlichen Umgebung eingeht. Diese subtilen Veränderungen in ihrem Verhalten deuten auf die Unterscheidungen hin, die Anna in ihren Beobachtungen trifft (Maturana 1988), welche nicht nur ihre eigene Wahrnehmung des Geräts bestimmen, sondern auch die Wahrnehmung ihres Partners. So wendet sich Anna zu ihrem Partner und gibt ihre Interpretation der Ausgabe wieder:

She misunderstood your question (.) and she was describing to you the different sounds of a rooster.

Sie hat deine Frage falsch verstanden (.) und sie hat dir die verschiedenen Geräusche eines Hahns beschrieben.

Anna beschreibt die Ausgabe als ein Missverständnis (vgl. Beneteau et al. 2019). Das Gerät hat in dem Sinne die Intention der Sprechenden nicht erfüllt, die den Laut eines Huhns hören und keine Beschreibung präsentiert bekommen wollten. Doch schaut man sich die Ausgabe genau an, wird ersichtlich, dass Wörter wie ›sound‹ und ›chicken‹ aus dem Befehl ›make the sound of a chicken‹ als einzelne Tokens und nicht als pragmatisch ganze Sätze verarbeitet wurden. In ihrer Interaktion mit der Maschine behandelt Anna das Sprachassistenzsystem wiederum als einen sozialen Akteur (Guzman 2018) und berücksichtigt dabei nicht den technologischen Aufbau der Maschine.

Nachdem Anna die Ausgabe als ein Nichtverstehen kategorisiert hat, reagiert Daniel auf Annas Interpretation folgendermaßen:

Did you ask her to do the sound of a rooster?

Hast du sie gefragt, das Geräusch eines Hahns zu machen?

Dabei greift Daniel nun Annas Bezeichnung ›rooster‹ (Hahn) auf und inkorporiert diese in seine Äußerung (Goodwin 2017), anstelle der vorher benutzten Bezeichnung ›chicken‹ (Huhn). An dieser Stelle wird der Einfluss des Sprachassistenzsystems deutlich. In der arabischen Ausgabe wurde das Wort ›Ad-deek‹ genutzt. Der Doppelkonsonant ›dd‹ der Transliteration aus dem Arabischen weist darauf hin, dass die Aussprache des Worts der modernen standardisierten arabischen Schriftweise entstammt, welche sich von Annas gesprochener Variation unterscheidet. Demnach inkorporiert Anna nun die Aspekte der Ausgabe des Geräts in ihre eigene Äußerung. Darüber hinaus verwen-

det sie diese in ihrer Übersetzung für Daniel. Diese Art des *Translanguaging* (z.B. Pennycook 2017b) verweist darauf, wie Kommunikation sich nicht nur über mehrere Sprachen, sondern auch über ökologische, soziale und temporale Ressourcen erstreckt. Die Heterogenität des gesprochenen Wortes erstreckt sich über mehrere zeitlichen Ebenen (Cowley/Steffensen 2015): 1) Anna reagiert auf kürzlich vergangene Ereignisse (d.h. die Ausgabe des Smart Speakers), 2) sie agiert auf einer erweiterten zugrundeliegenden soziokulturellen Ebene, indem sie Aspekte des modernisierten standardisierten Arabischs in ihre eigene Äußerung integriert, und 3) sie reagiert somit auf die vergangenen Entscheidungsprozesse des Entwickler*innenteams (siehe Abschnitt 4.2).

In diesem Sinne hatte nun die Ausgabe des Geräts einen wesentlichen Einfluss auf die weitere Entfaltung der sprachlichen Aktivität zwischen Anna und Daniel.

Abb. 4: Anna richtet Befehl an Smart Speaker



Alexa (.) E'mali sout addeek?

Daniels Frage führt dazu, dass Anna sich erneut dem Gerät zuwendet und auf Arabisch folgenden imperativen Befehl äußert:

Alexa, E'mali sout Addeek
Alexa, mach den Laut eines Hahns

Der Befehl ergibt sich aus Annas vergangenen sprachlichen und körperlichen Handlungen. Einerseits innerhalb ihrer Sprachgemeinschaft, und andererseits durch die Inkorporation von Ereignissen der jüngsten Vergangenheit. Zerlegt man den Befehl in seine Einzelteile, lassen sich die darunterliegenden ökologischen, dialogischen und sozialen Einflüsse wie folgt hervorheben:

1) Anna beginnt mit der Äußerung »E'mali«, die für »machen« steht. Hierbei handelt es sich um einen Ausdruck aus Annas Sprachvariation. Anna agiert somit vor dem Hintergrund ihrer eigenen soziokulturellen Einbettung. 2) Aus der dialogischen Koordination mit ihrem Partner wurde Anna dazu bewegt, den Befehl an das Sprachassistenzsystem zu richten. So übersetzt sie Daniels Äußerung aus dem Englischen ins Arabische. Somit inkorporiert Anna Ereignisse, die aus der dialogischen Koordination mit ihrem Partner entstanden sind, und schreibt diesen damit eine gewisse Bedeutung zu. 3) Schließlich inkorporiert sie die Äußerung »Addeek« in ihren Befehl. Aus einer ökologischen Perspektive reagiert Anna nun auf Ereignisse, die in ihrem direkten physischen Umfeld stattfinden: Sie reagiert auf die Ausgabe des Smart Speakers. Sie agiert vor dem Hintergrund ihrer Erwartungen der menschlichen Kommunikation (Suchman 2007) und versucht, sich sprachlich an das Sprachkonstrukt der Maschine anzupassen. Doch Annas Befehl konnte von der Maschine nicht prozessiert werden.

In einem nachfolgenden Interview beschreibt Anna ihre Erfahrung mit der Interaktion mit dem Sprachassistenzsystem folgendermaßen:

Yah because she was not speaking formal Arabic she was speaking (.) uhm Saudi dialect (.) and for some reason I don't think that I have ever really spoken to someone with a Saudi dialect before (.) I always assumed that it would be so different (.) but actually: not so much (.) like it's a bit thicker but I still understand and I thought it was cute. (2. Interview Anna, Pos. 320)

Anna nahm die Sprachvariation des Geräts als »Saudi dialect« war. Durch ihre Aussage »I don't think I have ever really spoken to someone with a Saudi dialect before« konstruiert sie das Sprachassistenzsystem als mögliche*ⁿ Kommunikationspartner*ⁿin, der*ⁿdie eine andere Sprache »spricht«.

I think there was only just one misunderstanding because the word for rooster is different (.) because I said (.) like in my dialect is Dik (.) (2. Interview Anna, Pos. 322)

Anna führt die Tatsache, dass die Maschine den Befehl nicht verarbeiten kann, auf die unterschiedlichen Sprachvarianten zurück und geht ebenso von der Prämisse einer gegenseitigen Verständlichkeit in Mensch-Maschine-Kommunikation aus.

4.2 »Wie Alexa sprechen lernte« – Die Entscheidungen der Entwickler*innen

Aus kommunikations- und medientheoretischer Sicht kann eine kommunikative Maschine als ›Figuration‹ verstanden werden, die auf eine Verflechtung von »interdependenten Menschen wie beispielsweise eine Gruppe, Gemeinschaft oder Organisation« hinweist (Hepp et al. 2022: 459). Eine Figuration, die strukturell als ›Akteurskonstellation‹ beschrieben werden kann, kann dabei als kommunikativ verstanden werden. Hepp et al. weisen auf die menschlichen Einflüsse in der Maschine hin und untersuchen, wie dieser Verflechtung Agenz zugeschrieben werden kann. Das Konzept der Figuration oder Verflechtung erfordert eine genauere Betrachtung der Vorstellungen und Antizipationen der Entwickler*innen, die das Sprachassistenzsystem *Alexa* mit einer arabischen Sprachfunktion ausgestattet haben. So mussten diese gezielt entscheiden, welche Sprachvariationen sie die ASR-Systeme für den Smart Speaker trainieren sollen, so dass dieser von einer breiten Masse an zukünftigen Nutzer*innen erfolgreich verwendet werden kann (Hardesty 2022). In dem Artikel »How Alexa learned Arabic« beschreibt Hardesty (2022) die genauen Entscheidungsprozesse der Entwickler*innen folgendermaßen:

For customers, engaging with Alexa in their native dialects would be more natural than speaking MSA. So the Alexa AI team — including computational linguists — determined that Arabic Alexa would be able to understand requests in both MSA and Khaleeji (Gulf) dialects. (Hardesty 2022)

Die Entscheidungsprozesse der Entwickler*innen drehten sich darum, eine »natürliche« Interaktion mit der Maschine zu ermöglichen. Dabei wurde auch hier vom Standpunkt einer angenommenen symmetrischen Mensch-Maschine-Kommunikation gehandelt. Die Vorstellung einer menschenähnlichen Kommunikation und die technischen Einschränkungen des Sprachassistenzsystems, das auf Prinzipien der Computerlinguistik basiert, bestimmen so die kognitiven Handlungen der Entwickler*innen. Da sprachbasierte Maschinen im vereinfachten Sinne durch Text operieren, wurde die Variation des moder-

nen standardisierten Arabisch und gesprochenen Arabisch der Golf-Region gewählt:

Moreover, there is no published writing in forms of Arabic other than MSA, so there's no standard orthography for them, either. Asking annotators to add diacritics could introduce more ambiguity than it alleviates. In the end, the Alexa AI team decided that ASR output should use only two diacritics, the shaddah and maddah, because they help with pronunciation accuracy on entity names that pass through ASR through NLU to TTS. (Hardesty 2022)

Die Entwickler*innen des Systems mussten mit der Schwierigkeit umgehen, dass neben der modernisierten standardisierten arabischen Sprachform keine weiteren einheitlichen verschriftlichten Sprachvariationen existieren. Entscheidungen mussten getroffen werden, wie eine lautliche Sprache durch diakritische Zeichen nutzbar für die Maschine gemacht werden kann, während das Bild einer natürlichen Interaktion mit dem System aufrechterhalten bleibt.

This means that someone issuing Alexa a request in one Arabic dialect might get a response in a different one. But that mirrors the experience that Arabic speakers in the region have with each other. (Hardesty 2022)

Die statisch erscheinende Ausgabe des Sprachassistenzsystems emergiert aus der Interaktion und Praktiken verschiedener Akteur*innen, die die Maschine konzipieren (Suchman 2007). So vermischen sich Antizipationen (z. B. wie hypothetische Nutzer*innen das System nutzen würden) mit gelebten Erfahrungen der Entwickler*innen, die sowohl sozial (z. B. wie arabische Sprachvariationen von Sprechenden aufgenommen werden) als auch technisch (z. B. Erfahrung mit dem Aufbau der Maschine) eingebettet sind. Genau wie der Befehl »E'mali sout Addeek« sich durch die körperliche dialogische Koordination zwischen Anna und Daniel dynamisch entfaltetete, so muss auch eine maschinelle Ausgabe als ein Produkt komplexer Handlungen von verschiedenen kognitiven Akteur*innen verstanden werden. Das, was Anna als »Missverständnis« bezeichnete, ist auf die vergangenen Entscheidungsprozesse der Entwickler*innen zurückzuführen.

5 Die Heterogenität der Sprechenden

Der Beitrag hat zum Ziel, auf die Rolle menschlicher Handlungen in der Mensch-Maschine-Kommunikation aufmerksam zu machen. Dabei ist er in einer posthumanen sowie kulturwissenschaftlichen Sichtweise verankert, die auf die lebensweltlichen Verflechtungen von Sprache aufmerksam macht. Statt Sprache als ein homogenes, regelkonformes System zu erfassen, wird in diesem Beitrag auf die heterogenen Dynamiken eingegangen, die sprachliches Handeln in der Mensch-Maschine-Kommunikation seitens der Nutzer*innen und der Maschine durchdringen. Vor dem Hintergrund der *Languaging*-Perspektive werden »die individuellen Erfahrungen des Einzelnen sowie dessen konkretes Sprachhandeln« enthüllt (Metten 2014: 20).

Während posthumane Ansätze auf Sprache sowie die *Languaging*-Perspektive auf den heterogenen Charakter von Sprache aufmerksam machen (d.h. wie Sprache sich über Raum, Zeit, Personen und materielle Ressourcen erstreckt (Cowley 2011; Pennycook 2017a)), untersucht dieser Beitrag die aktiven Handlungen einer sprechenden Person. Hierbei wird diese als aktive*r Beobachter*in (Maturana 1988) und kognitive*r Akteur*in (Giere 2004) verstanden, der*die nicht separat von seiner*ihrer Umgebung betrachtet werden kann. Die Heterogenität des Sprechenden ergibt sich aus dem Zwischenspiel von Dialogizität, Verkörperung und Zeitlichkeit.

Betrachtet man die Maschine aus dieser Perspektive, wird der Fokus auf die Akteurskonstellationen innerhalb der Maschine gerichtet (Hepp et al. 2022). Die Entwickler*innen sind demnach nicht nur passive Produzent*innen, sondern vielmehr genauso aktive und beobachtende Personen. Sie reagieren auf Ereignisse in ihrer unmittelbaren physischen Umgebung, treffen Unterscheidungen aus ihren lebenden Erfahrungen und stellen Antizipationen über das Verhalten der Nutzer*innen der Maschinen auf. Aus der *Languaging*-Perspektive wird die Mensch-Maschine-Kommunikation primär als aktive sprachliche Aktivität verstanden und somit zeitlich und räumlich entgrenzt konzeptualisiert. Die Nutzer*innen sowie die Entwickler*innen der Smart Speaker agieren in einer bestimmten zeitlich und räumlich abgegrenzten Umgebung. Diese Umgebung führt dazu, dass die Nutzer*innen aktiv mit den vergangenen Entscheidungsprozessen der Entwickler*innen interagieren und die Entwickler*innen wiederum die Handlungen imaginärer Nutzer*innen projizieren. Dadurch schaffen die Entwickler*innen eine soziokulturelle Realität in der Maschine, die durch die Verwendung von Sprache aufrechterhalten wird.

Sprachbasierte KI-Technologien werden kontinuierlich durch sprachliche Aktivitäten beeinflusst und sind somit Teil unserer sprachlichen Welt. Obwohl die binäre und kodifizierte Struktur der Maschine einen statischen Eindruck vermittelt, wird oft übersehen, dass menschliche Handlungen entkörperlicht werden, um sie maschinentauglich zu machen. Es sind die Nutzer*innen, die diesen Prozess umkehren und die entkörperlichten Tokens durch ihre Handlungen wiederbeleben. Der hier präsentierte posthumane Ansatz widerlegt den Symmetriegedanken der Mensch-Maschine-Kommunikation. Die Befehle an die Maschine sowie ihre Ausgaben erscheinen statisch, emergieren jedoch aus einer komplexen Verflechtung körperlicher, dialogischer Koordination verschiedener kognitiver agierender Personen. Mensch-Maschine-Kommunikation beruht auf der Entgrenzung des Statischen.

Literaturangaben

- Alač, Morana/Edwin Hutchins (2004): I see what you are saying: Action as cognition in fMRI brain mapping practice. In: *J Cogn Cult* 4:3-4, 629–661.
- Bailey, Benjamin (2004): Misunderstanding. In: Duranti, Alessandro (Hg.), *A companion to linguistic anthropology*, MA: Blackwell, 395–413.
- Barthel, Mathias/Henrike Helmer/Silke Reineke (2023): First users' interactions with voice-controlled virtual assistants: A micro-longitudinal corpus study. In: *Proceedings of the 27th Workshop on the Semantics and Pragmatics of Dialogue – Full Papers*, Maribor, Slovenien: SEMDIAL, 105–116.
- Bender, Emily M./Alexander Koller (2020): Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In: *Proceedings of the 58th annual meeting of the association for computational linguistics* 2, 5185–5198.
- Beneteau, Erin/Olivia K. Richards/Mingrui Zhang/Julie A. Kientz/Jason Yip/Alexis Hiniker (2019): Communication breakdowns between families and Alexa. In: *Conference on Human Factors in Computing Systems – Proceedings*, 1–13.
- Blair, Grant/Stephen J. Cowley (2003): Language in Iterating Activity: Microcognition re-membered. In: *Alternation (Durb)* 10:1, 132–162.
- Chemero, Anthony (2009): *Radical Embodied Cognitive Science*. Cambridge, M.A.: MIT Press.
- Cowley, Stephen J. (2009): Distributed language and dynamics. In: *Pragmatics & Cognition* 17:3, 495–508.

- Cowley, Stephen J. (2019): The Return of Language. In: *Chinese Semiotic Studies* 15:4, 483–512.
- Cowley, Stephen J. (2014): Linguistic embodiment and verbal constraints: Human cognition and the scales of time. In: *Front Psychol* 5: Article 1085. doi.org/10.3389/fpsyg.2014.01085.
- Cowley, Stephen J. (Hg.) (2011): *Distributed Language*, Amsterdam/Philadelphia: John Benjamins Publishing.
- Cowley, Stephen J./Matthew I. Harvey (2016): The illusion of common ground. In: *New Ideas Psychol* 42, 56–63.
- Cowley, Stephen J./Anneliese Kuhle (2020): The rise of language. In: *BioSystems* 198: Article 104264. doi.org/10.1016/j.biosystems.2020.104264.
- Cowley, Stephen J./Luarina Nash (2013): Language, interactivity and solution probing: Repetition without repetition. In: *Adaptive Behavior* 21:3, 187–198.
- Cowley, Stephen J./Frédéric Vallée-Tourangeau (2013): Systemic Cognition: Human Artifice in Life and Language. In: Cowley, Stephen J./Frédéric Vallée-Tourangeau (Hg.), *Cognition beyond the Brain: Computation, Interactivity and Human Artifice*, London: Springer, 255–273.
- Cowley, Stephen J./Sune Vork Steffensen (2015): Coordination in language: Temporality and time-ranging. In: *Interaction Studies*, 2015, 16:3, 474–494.
- Crawford, Kate/Vladan Joler (2018): *Anatomy of an AI System: The Amazon Echo as an anatomical map of human labor, data and planetary resources*. Abrufbar unter: <https://anatomyof.ai/>
- Dickel, Sascha (2021): Wenn die Technik sprechen lernt: Künstliche Kommunikation als kulturelle Herausforderung mediatisierter Gesellschaften. In: *TATuP-Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis/Journal for Technology Assessment in Theory and Practice* 30:3, 23–29.
- Esposito, Elena (2022): *Artificial Communication: How Algorithms Produce Social Intelligence*. Cambridge, Cambridge: MIT Press.
- Gahrn-Andersen, Rasmus (2019): But language too is material! In: *Phenomenol Cogn Sci* 18:1, 169–183.
- Gibson, James J. (1986): *The Ecological Approach to Visual Perception*, New York: Psychology Press.
- Giere, Ronald N. (2004): The problem of agency in scientific distributed cognitive systems. In: *J Cogn Cult* 4:3-4, 759–774.
- Goodwin, Charles (2017): *Co-Operative Action*. Cambridge: Cambridge University Press.

- Gunkel, David (2012): Communication and Artificial Intelligence: Opportunities and Challenges for the 21st Century. In: *Communication+1* 1:1. doi.org/10.7275/R5QJ7F7R.
- Guzman, Andrea L. (2018): What is human-machine communication, anyway? In: Guzman, Andrea L. (Hg.), *Human-machine communication: Rethinking communication, technology, and ourselves*, New York: Peter Lang, 1–18.
- Guzman, Andrea L./Seth C. Lewis (2020): Artificial intelligence and communication: A Human–Machine Communication research agenda. In: *New Media Soc* 22:1, 70–86.
- Hardesty, Larry (2022): *How Alexa learned Arabic*. Abrufbar unter: <https://www.amazon.science/latest-news/how-alexa-learned-arabic>
- Hoy, Matthew B. (2018): Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. In: *Medical Reference Services Quarterly* 37:1, 81–88.
- Hector, Tim (2023): Smart Speaker in der Praxis. Methodologische Überlegungen zur medienlinguistischen Erforschung stationärer Sprachassistenzsysteme. In: *Sprache und Literatur* 51:2, 197–229.
- Hepp, Andreas/Wiebke Loosen/Stephan, Dreyer/Juliane, Jarke/Sigrid Kanningeier/Christian Katzenbach/Rainer Malaka/Michaela Pfadenhauer/Cornelius Puschmann/Wolfgang Schulz (2022): Von der Mensch-Maschine-Interaktion zur kommunikativen KI: Automatisierung von Kommunikation als Gegenstand der Kommunikations- und Medienforschung. In: *Publizistik* 67, 449–474.
- Hutchins, Edwin (1995): *Cognition in the Wild*, Cambridge: MIT Press.
- Jokinen, Kriistina/Michael McTear (2010): *Spoken Dialogue Systems*. San Rafael: Morgan and Claypool.
- Kravchenko, Alexander V. (2022): The Maturanian Turn: Good Prospects for the Language Sciences. In: *Constructivist Foundations* 18:1, 201–212.
- Lind, Miriam/Sascha Dickel (2023): Speaking, but having no voice: Negotiating agency in advertisements for intelligent personal assistants. In: *Convergence: The International Journal of Research into New Media Technologies* 30:3, 1–17.
- Linell, Per (2009): *Rethinking Language, Mind, And World Dialogically: Interactional and Contextual Theories of Human Sense-Making*, Charlotte, NC: Information Age Publishing.
- Love, Nigel (2004): Cognition and the language myth. In: *Language Sciences* 26:6, 525–544.
- MacArthur, Emily (2014): The iPhone *Erfahrung*: Siri, the auditory unconscious, and Walter Benjamin's Aura. In: Weiss, Dennis M./Amy D. Proppen/Reid C

- Emmerson (Hg.), *Design, Mediation, and the Posthuman*, Lanham: Lexington Books, 113–127.
- Maturana, Humberto R. (1985): *Erkennen: Die Organisation und Verkörperung von Wirklichkeit: ausgewählte Arbeiten zur biologischen Epistemologie*, Dt. Fassung: Köck, Wolfram K., 2. Aufl., Braunschweig/Wiesbaden: Vieweg.
- Maturana, Humberto R. (1988): Reality: The Search for Objectivity or the Quest for a Compelling Argument. In: *The Irish Journal of Psychology* 9:1, 25–82.
- McTear, Michael/Zoraida Callejas/David Griol (2016): *The Conversational Interface: Talking to Smart Devices*. Basel: Springer International Publishing Switzerland.
- Metten, Thomas (2014): *Kulturwissenschaftliche Linguistik: Entwurf einer Medientheorie der Verständigung*, Berlin/Boston: De Gruyter.
- Miceli, Milagros/Martin Schuessler/Tianling Yang (2020): Between Subjectivity and Imposition: Power Dynamics in Data Annotation for Computer Vision. In: *Proceedings of the ACM on Human-Computer Interaction* 4:CSCW2, 1–25.
- Mühlhoff, Rainer (2020): Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media sociology of machine learning. In: *New Media Soc* 22:10, 1868–1884.
- Natale, Simone (2020): To believe in Siri: A critical analysis of AI voice assistants. In: *Communicative Figurations, Working Paper* 32, 1–17. Abrufbar unter: <https://www.kommunikative-figurationen.de>.
- Natale, Simone (2023): AI, Human-Machine Communication and Deception. In: Guzman, Andrea/Rhonda McEwen/Steve Jones (Hg.), *The Sage Handbook of Human-Machine Communication*, London: Sage, 401–408.
- Natale, Simone/Henry Cooke (2021): Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces. In: *Media Cult Soc* 43:6, 1000–1016.
- Pennycook, Alastair (2017a): *Posthumanist Applied Linguistics*, New York: Taylor and Francis.
- Pennycook, Alastair (2017b): Translanguaging and semiotic assemblages. In: *International Journal of Multilingualism* 14:3, 269–282.
- Pennycook, Alastair (2007): The myth of English as an international language. In: Makoni, Sinfree/Pennycook, Alastair (Hg.), *Disinventing and reconstituting languages*, Clevedon: Multilingual matters, 90–115.
- Porcheron, Martin/Joel E. Fischer/Stuart Reeves/Sarah Sharples (2018): Voice interfaces in everyday life. In: *Conference on Human Factors in Computing Systems – Proceedings* 2018: April.

- Porcheron, Marting/Joel E. Fischer/Sarah Sharples (2017): »Do Animals Have Accents?«: Talking with Agents in Multi-Party Conversation. In: CSCW February 25-March 1, Portland, OR, USA.
- Raimondi, Vincenzo (2019): The bio-logic of languaging and its epistemological background. In: *Language Sciences* 71, 19–26.
- Saussure, Ferdinand de (1916/2001): *Grundfragen der allgemeinen Sprachwissenschaft*, Berlin/New York: de Gruyter.
- Schneider, Britta (2021): Von Gutenberg zu Alexa: Posthumanistische Perspektiven auf Sprachideologie. In: Lind, Miriam (Hg.), *Mensch, Tier, Maschine: Sprachliche Praktiken an und jenseits der Außengrenze des Humanen*, Bielefeld: transcript, 327–346.
- Sen, Priyanka/Alham Fikri Aji/Amir Saffari (2022): Mintaka: A Complex, Natural, and Multilingual Dataset for End-to-End Question Answering. In: *Proceedings of the 29th International Conference on Computational Linguistics. International Committee on Computational Linguistics*, Gyeongju, Republic of Korea, 1604–1619.
- Shannon, Claude E./Warren Weaver (1949): *The Mathematical Theory of Communication*, Urbana: University of Illinois Press.
- Shotter, John (2019): Why being dialogical must come before being logical: the need for a hermeneutical-dialogical approach to robotic activities. In: *AI Soc* 34, 29–35.
- Suchman, Lucy (2007): *Human-Machine Reconfigurations: Plans and Situated Actions*, 2. Aufl., Cambridge: Cambridge University Press.
- Svensson, Jakob (2022): Modern Mathematics: Values and Biases in Tech Culture. In: Filimowicz, Michael (Hg.), *Systemic Bias: Algorithms and Society*, London / New York: Routledge, 21–39.
- Thibault, Paul J. (2011): First-order languaging dynamics and second-order language: The distributed language view. In: *Ecological Psychology* 23:3, 210–245.
- Thibault, Paul (2021): *Distributed Languaging, Affective Dynamics, and the Human Ecology Volume 1: The Sense-Making Body*, New York, NY: Routledge.
- Terzopoulos, George/Maya Satratzemi (2020): Voice Assistants and Smart Speakers in Everyday Life and in Education. In: *Informatics in Education* 19:3, 473–490.
- Van den Herik, Jasper C. (2017): Linguistic know-how and the orders of language. In: *Language Sciences* 61, 17–27.

Marie-Theres Fester-Seeger, PhD, Postdoctoral Fellow
Postdoc Network Brandenburg
Europa-Universität Viadrina
Große Scharnstraße 59
15230 Frankfurt (Oder)
fester-seeger@europa-uni.de

»Man hat schon so einen Austausch irgendwie«

Konzepte von Kommunikation und Interaktion zwischen Mensch und Maschine im Diabetesmanagement

Rita Tamara Vallentin

Abstract *Kommunikation wird in einer meist technisch instrumentellen Lesart mit der (einseitigen) Übertragung von Informationen zwischen unterschiedlichen Instanzen assoziiert, während Interaktion als Austauschverhältnis meist humanen Akteur*innen vorbehalten bleibt. Mit der Zunahme technischer Innovationen sowie hybrider Begegnungsformen und -praktiken verschwimmen diese Grenzen zunehmend. Der Beitrag illustriert am Beispiel algorithmengestützten Diabetesmanagements, wie Nutzende sich selbst und die Maschine als Kommunikations- und Interaktionspartner in einem sozialen Raum imaginieren. Dabei zeigen sich diverse Zuschreibungen an Aktivitätsgraden, von einer stark human geprägten Akteurhaftigkeit in den Mensch-Maschine-Interaktionen bis zu einer starken Agentivität des Geräts. Dies legt nahe, dass Kommunikation und Interaktion als Konzepte für die Nutzenden im Alltagsleben »zur Debatte stehen« – als diskursive Suchbewegung einer Entgrenzung.*

Keywords *Diabetestechnologie; Mensch-Maschine-Interaktion; Agentivität; Wechselseitigkeit*

1 Einleitung

Interaktion und Kommunikation sind Begriffe, die für die Sprachwissenschaft mannigfaltig definiert worden sind (Deppermann/Schmidt 2016; Goffman 1967; Hausendorf 2015; Luginbühl 2019; Saussure 2016; Searle 1969; Shannon/Weaver 1964). Während Interaktion eine generelle Wechselseitigkeit im aufeinander bezogenen Handeln zwischen zwei oder mehr Interagierenden umfasst, ist Kommunikation oft als stärker referentielle Form der Interaktion

definiert, die bestimmte »Informationen« auf Basis von Zeichensystemen vermittelt und in der dabei Bedeutung hergestellt wird, die aber auch frei von Anwesenheit und Wechselseitigkeit ablaufen kann. In diesem Beitrag gehe ich der Frage nach wie Menschen, die Teil von alltäglichen Mensch-Maschine-Interaktionen sind, über diese wechselseitigen Austauschbeziehungen sprechen, und ob bzw. wie sie ihre Alltagsgeräte als interaktionale und/oder kommunikative Gegenüber in diesen Beziehungen imaginieren. Dieser empirische Zugang öffnet zum einen den konzeptuellen Blick auf die Begriffe Interaktion und Kommunikation, die sich mit neuen hybriden Mensch-Maschine-Begegnungsformen und -Formaten verändern und entgrenzen. Zum anderen macht er deutlich, welche Position Maschinen und Menschen in diesen neuen Austauschbeziehungen, die von klassisch anthropozentrischen Kontexten abweichen, zugeschrieben bekommen.

Die spezifische Form der Mensch-Maschine-Interaktion, die hier in den Blick genommen wird, besteht im algorithmengestützten Diabetesmanagement von Menschen mit Typ-1 Diabetes. Die Insulinzufuhr bei den Teilnehmenden erfolgt konstant über eine Insulinpumpe, die über einen Katheter und eine Nadel mit dem Körper verbunden ist. Die Blutzuckermessung erfolgt über einen kontinuierlichen Glukosesensor (*continuous glucose monitor*, CGM), der am Körper klebt. Je nach Modell erfolgt eine algorithmenbasierte Form der digitalen Kommunikation zwischen CGM und Pumpe. Das heißt, dass die Blutzuckerdaten an das Pumpensystem übermittelt werden. Die Insulinpumpe ist dadurch in der Lage, bei einer Unterzuckerung die Insulinzufuhr zu stoppen oder bei einer Überzuckerung die Insulindosis minimal anzupassen. Das maßgeblich menschlich geprägte Diabetesmanagement von Personen mit Typ-1 Diabetes wird über diese sogenannten *automated insulin delivery* (AID)-Systeme also maschinell unterstützt. Dabei erfolgt nicht nur ein Informationsaustausch zwischen den Einzelteilen des Systems, sondern – so die These dieses Beitrags – auch eine ständige Interaktion zwischen den Betroffenen und den Geräten. Im Zentrum stehen Konzepte von Interaktion und Kommunikation, wie sie für den Kontext der Mensch-Maschine-Interaktion im Diabetesmanagement von Nutzenden selbst diskursiv hergestellt werden. Diese diskursanalytische Perspektive gibt Aufschluss über Konzeptualisierungen aus der Beteiligtenperspektive. Es wird deutlich, inwiefern sich eine Einbettung des Menschen in seine (technischen) Umwelten sowie eine menschliche Dezentrierung in Formen von Interaktionen nicht nur in akademischen posthumanistischen Diskursen manifestieren (siehe

z.B. Pennycook 2018), sondern konkret im Alltagsleben und -empfinden der Nutzenden niederschlagen.

Auf Basis von 31 semi-strukturierten qualitativen Interviews mit Menschen mit Typ-1 Diabetes, die AID-Systeme benutzen, gehe ich der Frage auf den Grund, wie der Mensch und die Maschine als Interaktionspartner mit gewissen Formen von Agentivität und Teilhabe an der wechselseitigen Austauschbeziehung imaginiert werden.

Der Beitrag gliedert sich wie folgt: Zunächst werde ich relevante theoretische Ansätze zur Mensch-Maschine-Interaktion diskutieren und dabei insbesondere auf Fragen einschlägiger Dualismen (beispielsweise ›denkender Geist und äußere Welt‹, ›Geist und Körper‹, ›Sender*in und Empfänger*in‹, ›Wahrnehmende und Nicht-Wahrnehmende‹) in den Konzeptualisierungen von Beteiligten sowie in der Prozessualität von Bedeutungsherstellung eingehen. Auch die komplexen körperlichen und interaktiven Verflechtungen in der medizinisch motivierten Mensch-Maschine-Interaktion im Diabetes-Management werden hier dargelegt. Darauf folgt ein Einblick in die sprachwissenschaftliche Konzeptualisierung von Kommunikation und Interaktion, um ein Fundament für den empirischen Teil des Beitrags zu legen. Der methodisch-analytische diskurslinguistische Ansatz wird in Kapitel 4 erläutert, in dem auch eine Beschreibung des Datenkorpus erfolgt. Die Ergebnisse der Analyse werden in Kapitel 5 diskutiert. Insbesondere kommen hier diskursive Abstufungen zum Vorschein, die sich von einer stark human geprägten Akteurhaftigkeit in den Mensch-Maschine-Interaktionen bis zu einer starken Agentivität des Geräts in den imaginierten Interaktionssituationen erstrecken.

2 Interaktion und Kommunikation zwischen Menschen und Maschinen

Wechselseitige Austauschprozesse zwischen Menschen und Maschinen sind Gegenstand einiger wissenschaftlicher Disziplinen. In diesem Kapitel sollen Schlaglichter auf Überlegungen zu diesen Prozessen spezifisch in der Kommunikationstheorie, dem interdisziplinären Feld der *Human-Machine Interaction* und aus dezidiert sprachwissenschaftlichen und damit verbundenen soziologischen Ansätzen geworfen werden. Ziel ist es, Gemeinsamkeiten und Unterschiede sowie epistemologische wie ontologische Herausforderungen in den Begriffen zu identifizieren: wer kommuniziert oder interagiert hier mit wem,

auf welche Art und Weise, unter welchen (technischen) Bedingungen und wozu (Schleidgen et al. 2023)? Diese Befunde aus der Literatur werden im empirischen Teil den impliziten und expliziten Perspektiven auf Interaktions- und Kommunikationskonzepte der Nutzenden von Diabetestechnik gegenübergestellt oder um diese erweitert.

Ein einschlägiges und in den Kommunikationswissenschaften viel zitiertes Modell zur Verständigung zwischen mindestens zwei Beteiligten aus dem 20. Jahrhundert ist das Sender-Empfänger-Modell von Shannon und Weaver (1964). Kommunikation wird hier im Rahmen einer mathematischen Berechenbarkeit verstanden, in der bereits bedeutungstragende Botschaften aus verschiedenen Informationsquellen ausgewählt und über einen Sender so codiert werden, dass sie in physikalischer Form vorliegen. Dieses physikalische Signal gelangt über einen (möglicherweise störanfälligen) Kanal zur* zum Empfänger*in, die*der es zur Botschaft decodiert. In dieser Form erreicht die Botschaft dann das eigentliche Ziel, welches eine Person oder ein »Ding« sein kann (Shannon 1968: 33–34). Das mathematisch-technische Vokabular und der Fokus auf codierende und decodierende Einheiten verdeutlicht Shannons Verständnis technisch vermittelter Kommunikation als Datenübertragungsprozess, der erst von Weaver auf menschliche Kommunikation ausgeweitet wurde:

The word communication will be used here in a very broad sense to include all of the procedures by which one mind may affect another. [...]. In some connections it may be desirable to use a still broader definition of communication, namely, one which would include the procedures by means of which one mechanism [...] affects another mechanism [...]. (Weaver 1964:3)

Letztlich ist der Prozess der Botschaftenauswahl, -codierung und -decodierung immer gleich, auch wenn sich auf technischer, semantischer und pragmatischer (»effectiveness«) Ebene der Botschaftenübermittlung je nach Sender*in und Empfänger*in wie auch Kanal mannigfaltige Kommunikationsprobleme ergeben können (Weaver 1964: 4–5). Ob Sender*in und Empfänger*in Menschen oder Maschinen sind, macht allerdings in der grundsätzlichen Anlage dieses Modells keinen Unterschied. Im Kern ist diese Form der Kommunikation unidirektional: ein statisches Hin- und Herschicken von Informationen. Empfänger*in und Sender*in werden nicht als Akteur*innen, sondern als De-/Chiffriermaschinen gehandelt. Wechselseitige soziale Beziehungen, dynamische Bedeutungsemergenz bzw. -aushandlung oder

Affektivität stehen in diesem *nachrichtentechnischen* Modell nicht zur Debatte, denn Bedeutung ist bereits vor der Codierung in die Botschaft eingeschrieben und wird nicht in kommunikativer Wechselseitigkeit produziert. Shannon und Weavers (1949) Überlegungen zur unidirektionalen Prozesshaftigkeit in der Übermittlung von Botschaften haben einige Überschneidungen mit dem »Kreislauf des Sprechens« aus dem sprachwissenschaftlichen Klassiker des frühen 20. Jahrhunderts, den erstmals 1916 veröffentlichten »Grundlagen der allgemeinen Sprachwissenschaft«. Saussure (2016) setzt sich zwar nicht mit Kommunikation zwischen Menschen und Maschinen auseinander, bedient sich aber einer gewissen Maschinenlogik in der Beschreibung zwischenmenschlicher Kommunikation. Er versteht diese als maßgeblich kognitiven Akt, der letztlich auf physiologische, physikalische und psychische Prozesse zur Produktion und Decodierung von sprachlichen Zeichen in den »Köpfen« zweier sich gegenüberstehender Sprechender reduzierbar ist (Saussure 2016: 13). Eine maschinenlogische Auffassung zeigt sich ebenso in der Konzeptualisierung von Vorstellungen als »Bewusstseinsdaten« (ebd., Hervorhebung RV) und der Decodierung der Verbindung zwischen Lautbild und Vorstellung als »arithmetische[m] Mittel« zwischen den Sprechenden (Saussure 2016: 15), die ein gemeinsames *identés* Sprachsystem besitzen. Auch wenn der Kreislauf Wechselseitigkeit zwischen mindestens zwei Beteiligten suggeriert, wird Bedeutung auch hier nicht gemeinschaftlich hergestellt. Der kommunikative Kreislauf besteht im Hin- und Herschicken sowie De-/Codieren von Zeichen durch zwei Gehirn»maschinen«, die Zugriff auf das gleiche (Sprach-)System haben.

Unterschiedliche Affordanzen im entweder rein kommunikativen oder aber interaktiven Bezug zwischen Mensch und Maschine sind mittlerweile Kern des interdisziplinären Forschungsfelds der *Human-Machine Interaction* (HMI). Je nach Form von Maschine und Art der Kommunikation und Interaktion spitzt sich dieses Feld unter den Labels *Human-Machine Communication*, *Human-Computer Interaction* oder *Human-Robot Interaction* (Boy 2011; Fortunati/Edwards 2020) zu. Bereits an diesen Begriffen lässt sich die Komplexität des Begriffsfeldes HMI erkennen. Während Mensch-Maschine-Kommunikation häufig eher als unidirektionale Eingabe – vom Menschen zum Computer – oder Ausgabe – vom Computer zum Menschen – verstanden wird, verspricht der Begriff der Interaktion, ähnlich wie in der Linguistik, eine stärkere Fokussierung auf die Wechselseitigkeit.

Interaktion wird in diesem interdisziplinären und multiperspektivischen Forschungsfeld durchaus verschiedenartig definiert. Schleidgen et al. (2023)

stellen in einem disziplinhistorischen Aufriss fest, dass mit zunehmender Technisierung und Digitalisierung Maschinen immer häufiger immer komplexere Aufgaben für Menschen erleichtern oder sogar vollständig erledigen. Daher wird Mensch-Maschine-Interaktion in diesem Forschungsfeld aus ingenieur- und programmiertechnischer, aber auch aus psychologischer und sozialwissenschaftlich-ethischer Perspektive primär als Designproblem konzeptualisiert. Dabei geht es häufig um Schnittstellen (*Interface Design*), Barrierefreiheit in der Bedienung, Personalisierung oder an spezifische kulturelle und sprachliche Kontexte angepasste Interaktionsmuster der Maschinen. Im weitesten Sinne sollen sie den komplexen physischen, kognitiven, sozialen und emotionalen menschlichen Anforderungen bei Beibehaltung von Sicherheit, Performance, Komfort und Ästhetik gerecht werden (Boy 2011: 1). Hierbei steht der die Maschine manipulierende Mensch mit seinen interaktionalen und kommunikativen Bedürfnissen im Fokus. Mit zunehmender technischer und programmgeleiteter Autonomie der »smarten« Maschine verschiebt sich der Begriff der Interaktion als wechselseitige Bezugnahme zwischen Mensch und Maschine im Hinblick auf die gemeinsame Problemlösung eines dritten Ziels zu Begriffen von Kooperation (Boy 2011: 12; Hoc 2010) oder von »partnership« (Schmidt/Loidolt 2023) zwischen humanen und maschinellen Handelnden (*agents*). Auch Kooperation wird hier unterschiedlich definiert. Bei Boy (2011) und Hoc (2010) bezieht sie sich auf Aufgaben, bei denen sowohl der Mensch als auch die Maschine ihre spezifischen Fähigkeiten kollaborativ zur Erfüllung einbringen. Die Zusammenarbeit wird so gestaltet, dass die Stärken des einen die Schwächen des anderen ausgleichen. Am Beispiel der Flugsicherung stellt Hoc (2010) fest, dass Menschen in der Regel besser in der Lage sind, mit neuartigen Situationen umzugehen und Urteile unter Unsicherheit zu fällen, während sich Maschinen durch die schnelle Verarbeitung großer Datenmengen auszeichnen. Bei Schmidt und Loidolt (2023: 55) ist Kooperation die größtmögliche Instanz von Partnerschaft zwischen Mensch und Maschine und bedeutet eine gewisse Kontingenz – im Sinne einer Unvorhersehbarkeit (Luhmann 1984: 148) – in den Handlungen von Maschinen beim Kooperieren im Erreichen eines durch den Menschen festgelegten Ziels. Auch hier liegt Kooperation nicht in unbedingt aufeinander bezogenen Handlungen vor, sondern im Verfolgen eines gemeinsamen Ziels, wobei die Maschine mit ihrem Handeln zum Erreichen des Ziels »überraschen« kann, ihr Handeln also nicht uneindeutig vorhersehbar ist.

Eine grundlegende Unterscheidung in diesen Überlegungen besteht darin, inwieweit Menschen Maschinen anthropomorphisieren, ob sie Maschinen in-

tionales Handeln zuschreiben, oder ob dies für konzeptionelle Fragen von Maschinen in der Mensch-Maschine-Interaktion eigentlich keine Rolle spielt. Um »partnership« in soziale Partnerschaft und damit in intentionales Handeln auf ein gemeinsam gesetztes Ziel zu verwandeln, bräuchten Maschinen tatsächliche Formen von Bewusstsein, so die philosophische Perspektive von Schmidt und Loidolt (2023: 26): »As it currently stands, we believe that only a conscious agent with human-like qualities can be treated in social interaction as an end in itself rather than a mere means to fulfill another agent's purposes«. Gerade in den diskursiven Zuschreibungen der Nutzer*innen an ihre Maschinen liegt jedoch das Potenzial, eine mögliche Entgrenzung des Interaktionsbegriffs auszuloten. Nicht zuletzt manifestieren sich hier normative Erwartungen an ein technisches Gegenüber, das insbesondere mit zunehmender und teilweise existenzieller Bedeutung im menschlichen Alltag – wie etwa bei Menschen mit Diabetes – als Teil des eigenen Selbst betrachtet wird (Vallentin 2023). Perspektiven auf menschliche Imaginationen von Maschinen in Interaktion geben Aufschluss darüber, an welchen Kriterien und Bedingungen Sozialität und Wechselseitigkeit gemessen werden. In dieser Tradition verorten sich Studien zu Empathie (Gouvrit 2013), *Embodiment* und Multimodalität (Pustejovsky/Krishnaswamy 2021; Su et al. 2023) sowie Formen von Beziehungen zwischen Mensch und Maschine, die sich durch Interaktion herausbilden (Coeckelbergh 2011; Trujillo/Gregory/Ackerman 2019).

Neueste als »sprechend« oder »textproduzierend« konzeptualisierte Maschinen wie Alexa oder ChatGPT werfen dabei zum einen Fragen zum möglichst menschenähnlichen Sprechen und Schreiben auf (Coeckelbergh 2020). Zum anderen werden mit Maschinen, deren Hauptaufgabe im Dialog mit Menschen besteht, Formen der multimodalen Interaktion und des entsprechenden multimodalen Designs der Maschinen relevanter; z.B. in Form von Robotern in öffentlichen Räumen oder im Bildungsbereich (Pitsch 2019). Sprachwissenschaftliche Konzepte und Theorien bekommen in diesem Feld der HMI hier langsam aber stetig mehr Aufmerksamkeit (Leblebici 2024; Lind 2022; Schneider 2022).

Interaktion und Kommunikation sind in den erläuterten Kontexten durchaus heterogene Konzepte. Sie folgen in ingenieurs- und technisch fokussierten Ansätzen der HMI weitestgehend einer menschengeleiteten Aufgabenerledigungslogik ganz im Sinne eines technischen und mechanistischen Kommunikationsverständnisses, das die Übertragung von vordefinierter Bedeutung als linearen und unidirektionalen Prozess betrachtet. Hier geht es primär um Kommunikation als Übermittlung feststehender Informationen ohne Berück-

sichtigung der emergenten oder dialogischen Bedeutungskonstruktion, die in interaktiven Kontexten auftritt. Im Kontrast dazu stehen soziologische oder sprachwissenschaftliche Ansätze in der HMI, die eine breitere Definition von Interaktion und Kommunikation beinhalten, welche die Wechselseitigkeit und die potenziell dynamischen Austauschprozesse zwischen Menschen und Maschinen anerkennen. Insbesondere die Erweiterung hin zu Konzepten der Kooperation und Partnerschaft in der Interaktion mit agentiven Systemen deutet darauf hin, dass von einer rein menschenzentrierten, einseitigen Perspektive abgerückt wird. Stattdessen wird anerkannt, dass Maschinen zu kooperativen Prozessen beitragen können und dies frühere Modelle der passiven Datenverarbeitung übersteigt. Sozial und tatsächlich intentional sind die Maschinen dabei allerdings nicht. Mögliche anthropomorphisierende Zuschreibungen von Intentionen und Fähigkeiten an Maschinen durch Nutzer*innen verschleiern dabei möglicherweise die Asozialität des technischen Gegenübers. Die Zuschreibungen legen aber offen, wie sich Nutzer*innen Interaktion vorstellen, welche Aspekte in einem von den Beteiligten geprägten Begriff von Mensch-Maschine-Interaktion relevant gesetzt werden, und wie Mensch und Maschine in diesen gedachten Räumen konzeptualisiert werden.

3 Interaktion und Kommunikation in der Linguistik

In der Linguistik erfährt der Begriff der Interaktion insbesondere im Feld der Interaktionalen Linguistik eine nähere Betrachtung (z.B. bei Couper-Kuhlen/Selting 2018). So wird er enggeführt als Bearbeitung von bestimmten Aufgaben mit verbalsprachlichen Mitteln unter Beachtung spezifischer sozialer und räumlicher Ressourcen (Hausendorf/Schmitt 2018). Dabei gibt es keine Einigkeit, welche Minimalkriterien erfüllt sein müssen, um Formen von Kommunikation als wechselseitige »Interaktion« zu bezeichnen. Hausendorf bekräftigt das Konzept der »Anwesenheit« eingelöst in »Wahrnehmungswahrnehmung« als grundlegendes Kriterium für Interaktion zwischen mindestens zwei Interagierenden (Hausendorf 2015: 47). Damit ist zum einen konstituiert, dass es für Interaktion nicht vordergründig Sprache bedarf, sondern schon die Anwesenheit eines Gegenübers Interaktion hervorbringen kann. Zum anderen ist aber auch ein »wahrnehmendes« Moment gesetzt, in dem das Gegenüber aktiv Erwartungen an die (sprachlichen) Handlungen des*der Anderen hat. Auch wenn Hausendorf (2015: 49) eine Interaktion zwischen Menschen und Maschinen damit ausschließt, so können bestimmte Systeme jedoch Sprache in einer

bestimmtem Materialität, Sequenzialität und unter »Inanspruchnahme eines Zeichensystems« (ebd.: 53) hervorbringen, welche als notwendige Bedingung für Interaktion formuliert werden, auch wenn die hinreichende Bedingung der »Anwesenheit« nicht erfüllt wird.

Ein weniger enges Korsett für einen möglichen linguistischen Interaktionsbegriff mit Maschinen findet sich in Ansätzen, die verschiedene Formen von Partnerorientierung in kommunikativen Kontexten, basierend auf einer notwendigen »doppelten Kontingenz« als Wahrnehmungswahrnehmung und Unterstellung von Handlungsabsichten des Anderen (Luhmann 1984: 156–157), stärken: »What matters is whether the interaction with the machine has the features of communication with a contingent autonomous partner«¹ (Esposito 2017: 255–56). So hat zum Beispiel schon Luhmann selbst seinen Zettelkasten zur Organisation von Notizen als einen unvorhersagbaren Interaktions- und Kommunikationspartner verstanden:

But Luhmann's Zettelkasten was structured in such a complex way that it could produce authentic surprises and did not simply act as a container (Behälter), allowing the author to retrieve what he once put in it. The information »produced« in the act of communication was the result of a query (Anfrage), which activated the internal network of references, and it was different from what had been stored by Luhmann in his notes (Luhmann 1981: 59). Of course, the archive is not contingent in the sense of autonomously deciding what to do and not to do; yet it is perceived by the user as unpredictable, informative, and reacting to the specific requests of its partner. The answers Luhmann got as a result of his query did not exist before his quest. In such cases the added value of communication is present since, as Luhmann himself experienced, the file-box acts as a communication partner. (Esposito 2017: 256)

Der Zettelkasten erfüllt die Bedingungen von Kontingenz, also eines unvorhergesehenen, nicht vorher kalkulierbaren Bedeutungsaspekts, der aufgebracht und in Wechselseitigkeit mit einem Beteiligten eingelöst wird. Er erfüllt aber

1 Interessant ist hier die Gegenüberstellung von ›Interaktion‹ und ›Kommunikation‹. Die Qualifizierung für Interaktion mit der Maschine liegt an deren technischen Affordanzen: So lange ein kontingenter – d.h. nicht unbedingt vorhersehbarer – Output erfolgen könnte, kann die Form der Beziehung als Interaktion beschrieben werden. Dieser Output fußt aber auf einer Informationsvermittlung, die auf einer Frage-Antwort-Logik basiert.

auch die Erwartung, dass – im Sinne einer interaktiven Sequenzialität – auf eine Anfrage eine Antwort folgt. Schon Luhmann fragt sich deshalb, ob mit dem Einzug neuer algorithmenbasierter kommunikativer Formen von Technik ein alternativer Begriff zur Kommunikation gefunden werden muss:

Wer kommuniziert jetzt mit wem? Eignet sich unser Begriff überhaupt noch dafür? Oder sind wir an einer Schwelle, wo man sieht, dass wichtige Informationsverarbeitungsverfahren unserer Gesellschaft schon nicht mehr als Kommunikation klassifiziert werden? Oder müssen wir den Begriff neu bilden, aber wie? (Luhmann 2008: 314)

In Luhmanns Zitat deutet sich an, dass eine leblose Aufgabenerledigungslogik zwischen Mensch und Maschine nicht mehr greift. Es zeigt ein tentatives Suchen nach dem Begreifen eines Verhältnisses, welches stärker interaktiv erscheint. Esposito (2017) schlägt dafür den Begriff der »virtuellen doppelten Kontingenz« als notwendiges Kriterium für den Umgang mit algorithmenbasierter Technik vor. Dabei werden die Annahmen, dass es in der interaktiven Beziehung unbedingt sinnlich wahrnehmender oder intelligenter Akteure bedarf, die in sozialen Beziehungen zueinander stehen, verworfen: »What makes algorithms socially relevant and useful is their ability to act as partners in communication that produces and circulates information, independently of intelligence« (Esposito 2017: 253). Als »partners in communication« werden Maschinen zu einem *interaktiven* Akteur in der Bedeutungsherstellung. Der Partnerstatus ergibt sich aber nicht zwangsläufig aus Kreativität oder kognitiver Leistung, sondern aus spezifischen Zugängen zu, und Kombinationen von Datenoutput.

Im Sinne dieser Loslösung von Anwesenheit und Wahrnehmungswahrnehmung als notwendigen Bedingungen für gelingende Interaktion sprechen Deppermann und Schmidt (2016) von verschiedenen Formen der Partnerorientierung. Diese ist selbst in der menschlichen face-to-face Interaktion immer in einem gewissen Grad imaginiert. Diese Imagination des Gegenübers lässt sich mit einem Blick auf spezifische *Recipient Designs* interaktionslinguistisch rekonstruieren und gibt Aufschluss über das vom Sprechenden beim Gegenüber angenommene Wissen oder angenommene Haltungen. Dabei wird der*die Interaktionspartner*in mittels sprachlicher Mittel auf bestimmte Arten und Weisen entworfen. Maschinen fallen hier als »Partner mit eingeschränkten Interaktionskompetenzen« (Deppermann/Schmidt 2016: 384; siehe auch Fischer 2016) in den Bereich der interaktionalen »Grenzfälle« (Depper-

mann/Schmidt 2016: 380). In diesen wird die Maschine in anthropomorpher Form als Interaktionspartner imaginiert, es werden ihr Handlungsabsichten und kommunikative Partnerorientierung zugeschrieben: »der Andere [wird] kontrafaktisch imaginär konstruiert, obwohl gegenteiliges Wissen vorhanden ist [...] und dessen situiertes Handeln wird in Bezug auf kontrafaktische Merkmale von Partnermodellen (re-)interpretiert« (Deppermann/Schmidt 2016: 398).

Diese interaktionslinguistischen Reflexionen machen deutlich, dass Mensch-Maschine-Interaktionen im Status quo der aktuellen technischen Möglichkeiten noch in Aushandlung und Konzeptualisierung begriffen sind, ausgehend von (jedoch nicht gleichsetzend mit) interpersonalen Formen von Interaktion. Als »Grenzfälle« von Interaktion werden Maschinen innerhalb ihrer technischen Affordanzen zu »quasi-second interlocutors« (Fortunati/Edwards 2020: 8), die gemeinsam mit den menschlichen Beteiligten »quasi-social action«, die »quasi-interpersonal« (Höflich 2013: 38) ist, hervorbringen. Interaktion zwischen Menschen und Maschinen erscheint also als ein Pseudo-Phänomen – nah dran, aber eben nicht vollumfänglich vergleichbar mit menschlicher Interaktion. In posthumanistischen Ansätzen wird genau der Grenzfall- oder »quasi«-Charakter von Mensch-Maschine-Interaktion zu einer eher universellen Form erhoben. Objekte wie auch nicht-menschliche Tiere oder Pflanzen sind Teile von Umwelten, in denen und mit denen sich Menschen bewegen, und die gleichzeitig menschliches Handlungsvermögen bedingen und formen (Pennycook 2018: 14). Diese Sichtweise postuliert, das Konzept der Interaktion über traditionelle soziale oder kommunikative Handlungen hinaus zu erweitern und auch wechselseitige physische, umweltliche und technologische Dynamiken als integralen Teil der sozialen und kulturellen Realität zu betrachten.

Die akademischen Debatten über die Spezifik oder Grundsätzlichkeit von Mensch-Maschine-Interaktion sind noch längst nicht abgeschlossen. Was aber über die menschliche Imagination der Maschine als Partner in der Interaktion deutlich wird, sind die Fragen, inwiefern Menschen die Beziehungen zu Maschinen als wechselseitig wahrnehmen, wie Maschinen als Gegenüber in diesen Beziehungen entworfen werden und was für ein Bild des Menschen in diesem Gefüge entsteht. Denn jenseits von Fragen zur (un)möglichen Fähigkeit von Wahrnehmungswahrnehmung ist Interaktion primär als ein beziehungskonstituierender Raum zu verstehen: »interaction is not all about sequence and collaboration; it is also about building (or severing, or

negotiating) social ties and (re)defining the nature of social life« (Gordon 2011: 113).

Im analytischen Teil dieses Beitrags markiert dieser Gedanke der Beziehungskonstitution den Ausgangspunkt für die Perspektive und die Imagination der beteiligten und involvierten Nutzenden selbst. Im Zentrum steht dabei ein Anwendungskontext, in dem ihre Beziehung zur Maschine genuin körperlich verfasst ist. Ausgehend von einem Interaktionsbegriff, der Formen von Partnerschaftlichkeit als aufeinander bezogen betont, und der Interaktion als beziehungskonstituierenden Raum versteht, wird sich zeigen, wie technische Geräte als interaktionale Gegenüber diskursiv entworfen werden. Es wird ebenfalls deutlich, welche Fähigkeiten zur Interaktion in die Geräte und ihre Affordanzen eingeschrieben werden.

4 Methodisches Vorgehen

Das Datenkorpus, welches Grundlage für die teilnehmendenzentrierten Entwürfe von Mensch-Maschine-Interaktion ist, besteht aus 31 qualitativen semi-strukturierten Interviews mit Menschen, die Typ-1-Diabetes haben und ein AID-System (*automated insulin delivery*) für ihr Diabetesmanagement nutzen. Alle Interviews wurden nach den Konventionen des gesprächsanalytischen Transkriptionssystems 2 (GAT 2) transkribiert (Selting et al. 2009). Drei Interviews fanden in häuslicher Umgebung oder Diabetespraxen statt, alle weiteren online in Zoom. In den Interviews spreche ich mit den Teilnehmenden über den Tag ihrer Erstdiagnose, den ersten Kontakt mit Geräten, die das Diabetesmanagement technologisch unterstützen, das System, was aktuell genutzt wird, das Verhältnis zwischen Körper und Gerät sowie mögliche emotionale Beziehungen zur Technik. Beim Sprechen über die Diabetestechnologie kommt es in den Interviews häufig zu Sequenzen, in denen von den Sprecher*innen konstruierte Dialoge (Tannen 2007) mit den Geräten *performed* werden. Dabei wird deutlich, wie die Teilnehmenden die Geräte sprachlich als »Partner« in der Interaktion herstellen.²

Diese Passagen und andere Sequenzen der Interviews, in denen die Teilnehmenden über ihre Geräte als (un)mögliche Interaktionspartner spre-

2 Die »imaginäre Partnerorientierung« (Deppermann/Schmidt 2016:84) ist in diesem Fall zweiter Ordnung, denn es wird eine imaginäre Partnerorientierung im konstruierten Dialog für die Interviewende im Kontext des Interviews versprachlicht.

chen, werden zunächst im Interviewkorpus identifiziert und dann qualitativ im Hinblick auf sprachlich-diskursive Praktiken analysiert. Der Fokus auf sprachlich-diskursive Praktiken betont die Rolle von Sprache bei der Konstruktion sozialer Realitäten und der damit verbundenen Machtstrukturen (Cameron 2001: 15). Sprachlich-diskursive Praktiken sind als sprachliche Mittel zu verstehen, um die Maschine als mögliches interaktionales Gegenüber mit bestimmten Formen von Agentivität und damit verbundenem Wissen hervorzubringen (Habscheid et al. 2023), etwa durch sprachliche Kategorisierungspraktiken oder Zuschreibungen bestimmter Eigenschaften von Maschinen. Wie Menschen über ›ihre‹ Maschinen sprechen, gibt Aufschluss darüber, wie sie diese einordnen, ob sie sie als belebt oder unbelebt, als handelnde Akteure, undurchschaubare Black Boxes oder passive Objekte wahrnehmen. Ein Blick auf die sprachlichen Such- und Aushandlungsbewegungen in den Interviews über die kommunikative oder interaktionale Beziehung zwischen Mensch und Maschine gibt somit Einblicke in die interpretativen Wahrnehmungs- und Erfahrungsprozesse der Beteiligten (Cameron 2001: 14). Die konstruierten Dialoge mit den Geräten stehen häufig im Widerspruch zu Zuschreibungen an anderen Stellen der Interviews. Dies könnte zum einen daran liegen, dass die sprachlichen Interaktionen mit den Diabetesgeräten eine enge Form der sozialen Beziehung implizieren, die in den konstruierten Dialogen auch als solche dargestellt wird. An anderen Stellen in den Interviews wird eine soziale Beziehung zu den medizinischen Geräten oft (vehement) verneint, vermutlich weil dies nicht den Normen von Maschinen als neutralen Gebrauchsgegenständen und den entsprechend erwarteten Einstellungen der Nutzer*innen entspricht. Neben den Sequenzen, in denen es explizit um die Geräte als (un)mögliche Interaktionspartner geht, ist daher methodisch auch eine diskursanalytische Perspektive nötig, die das gesamte Interview in den Blick nimmt. Damit werden sowohl mögliche Widersprüche in der *in situ* Interaktion mit der Interviewerin hinsichtlich der Konzeptualisierung der Maschine als auch analytische Befunde aus den in den Interviewsituationen tatsächlich beobachtbaren Interaktionen mit dem Diabetesgerät selbst berücksichtigt.

5 Maschinen als sprachlose Geräte und sprechende Gegenüber

Nach einer ersten Kodierung der Daten kommen drei thematische Schwerpunkte zum Vorschein, die nicht immer klar voneinander abgegrenzt sind,

sondern oft in Form von diskursiven Suchbewegungen produziert werden. Auf der einen Seite werden die medizinischen Geräte, beziehungsweise meist ein Gerät – die Insulinpumpe – als ›sprachlose‹ Materialitäten begriffen, mit denen Begegnung/Austausch, wenn überhaupt, nur in einem imaginierten Raum stattfinden kann. Die Geräte sind ›sprach- und geräuslos‹ und dienen als oft passiv konstruierte Projektionsfläche für (einseitige) menschliche Ansprache. Zum anderen bekommen die Geräte eine interaktive Agentivität zugeschrieben. Diese kann innerhalb der technischen Affordanzen geschehen und sich in Form von »Piepen« oder »Vibration« äußern. Die Konzeptualisierung kann aber auch in Vorstellungen von ›sprechenden‹ Geräten münden, deren Teilhabe an interaktiver Bedeutungsherstellung in der Interviewsituation von den Sprecher*innen als konstruierter Dialog (Tannen 2007) *performed* wird. Die verschiedenen Ausprägungen, und wie sie sprachlich hergestellt werden, stelle ich in den folgenden Unterkapiteln dar.

5.1 Sprachlose Geräte

Ein sich durch die Mehrheit der Interviews ziehendes Thema ist, dass die Interaktion mit den Maschinen oft als einseitige und menschengeleitete Form von Kommunikation konzeptualisiert wird. Während Interaktion auf aufeinander bezogener Wechselseitigkeit fußt und damit eine gewisse Form von Interagentivität voraussetzt, kann Kommunikation auch eher einseitig über einen sprechenden Agens und eine*n – durchaus auch passiven – Empfänger*in theorisiert werden. Zwei Aspekte sind für diese diskursive Darstellung der spezifischen Form von Kommunikation maßgeblich. Zum einen erfolgt die einseitige Kommunikation im Modus negativer Ansprache, indem mit den Geräten ›geschimpft‹, ›gemeckert‹ oder über sie ›geflucht‹ wird. In den meisten Fällen wird auf die Interviewfrage »Sprechen Sie manchmal mit Ihren Geräten?« eine klare Ablehnung formuliert, die in vielen Fällen eingeschränkt wird durch das Einräumen einer einseitigen negativ gefärbten Kommunikation. ›Meckern‹, ›Schimpfen‹ und ›Fluchen‹ werden durch diese diskursive Ausdifferenzierung von der Interviewfrage nicht unter ›mit den Geräten sprechen‹ gefasst. Dies erzeugt eine Konzeptualisierung von ›sprechen‹ als nicht ausschließlich negativ gerahmte Aktivität. Zweitens würde Sprechen auch eine gewisse »Sprachfähigkeit« implizieren, die eine Form von »Wahrnehmung« (ein Hören oder Fühlen) voraussetzt:

Exzerpt 1, Interview mit Ulf³, m, 66

- 01 I: äh=REDEN sie manchmal mit ihren geräten?
- 02 U: (2.0) ähm: ich !FLUCHE! vielleicht;
- 03 über sie.
- 04 (.) ob die das WAHRnehmen,
- 05 das WEIß ich nicht;=
- 06 =ich glaube eher nicht.
- 07 I: [hahahaha]
- 08 U: [aber] ähm:=man FLUCHT, mit der TECHnik, so wie man's mit dem–
- 09 mit dem
- 10 AUto, (.) ähm mit dem rasenmäher,
- 11 (.) mit anderen dingen auch macht, (1.0)
- 12 ähm. (.)
- 13 aber ich kommuniziere ansonsten NICHT mit denen.
- 14 (.) die kriegen die beFEHLE,=die die BRAUCHEN, (.)
- 15 und dann;
- 16 (-)SOLLTEN die, eben arbeiten.

Das Fluchen »über« die »Technik« wird von Ulf hier gleichgesetzt mit der gängigen Praktik, andere technische Geräte kommunikativ zu behandeln. Die Kommunikation mit den Geräten beschränkt sich aus Sicht dieses Nutzers auf den Ausdruck seiner negativer Emotionen und der Formulierung von »Befehlen« (Zeile 14). Den Geräten selbst wird hier keine Form von interaktiver Agentivität zugeschrieben. Ulf bezweifelt, dass sie »wahrnehmen« können, schreibt ihnen aber durchaus Agentivität in Bezug auf ihre Aufgaben im Diabetesmanagement zu (Zeile 15-16, »dann; (-)SOLLTEN die, eben arbeiten«). Auffällig in dieser Passage ist auch, wie der Sprecher eine gewisse Distanz zu seinen Geräten herstellt. Zum einen werden sie gleichgesetzt mit Formen von Technik, die nicht so nah am Körper sind wie Insulinpumpe und Glukosesensor (Zeile 10, »AUto«, »rasenmäher«), zum anderen wird der »Dingstatus« der Geräte

3 Die Namen aller Interviewteilnehmenden wurden pseudonymisiert. »M« steht für männliche, »F« für weibliche Selbstidentifikation. Die Zahl gibt das Alter der Teilnehmenden zum Zeitpunkt des Interviews an.

durch die Demonstrativpronomen »die« und »denen« betont. Während in Bezug auf das Fluchen die Geräte mit dem Personalpronomen »sie« bezeichnet werden (Zeile 3), wechselt Ulf danach zum Pronomen »die« (Zeile 4). Dies lässt nicht nur Schlüsse in Bezug auf die Dinglichkeit der Geräte zu, sondern auch negativ konnotierte Äußerungen (Ehrmantraut 2020: 62f.), die den Status der Geräte als Dinge, »die eben arbeiten sollten« unterstreichen.

Die Fähigkeit zur Wahrnehmung – spezifisch eines Hörens – wird auch im Interview mit Petra augenfällig. In der vorherigen Interviewsequenz wurde das ›Meckern‹ mit den Diabetesmanagement-Geräten als einzige Form der Kommunikation ausgeführt und das Loben der Geräte ausgeschlossen. In der darauffolgenden Sequenz betont Petra die kommunikative Wirkmächtigkeit nur auf Seite der menschlichen Sprecherin, weil das Gerät nichts »hört« (Exzerpt 2, Zeile 12) und entsprechend nicht interaktiv handeln kann:

Exzerpt 2, Interview mit Petra, w, 37

- 01 I: was denken sie_n was es ÄNDERN,
 02 würde=wenn sie die geräte LOben–
 03 P: (4.0) (puh). (3.0)
 04 naja=was ÄNDert es denn,=wenn ich's –(-)
 05 wenn_ich MECKER;
 06 (1.0)
 07 I: hm (–)
 08 P: ja? (.)
 09 macht ja auch kein,
 10 unterschied==
 11 =beim geRÄT, (.)
 12 das HÖRT mich ja nicht? °hh=
 13 öhm: (–)
 14 aber ich (.) ÄNder: vielleicht MEI:ne;=
 15 =einstellung dazu. (.)
 16 I: <<bestätigend>hm_hm>. (-)
 17 P: also dass_ich_es nicht nur,=

- 18 =ne:gativ sehe. (1.0)
 19 ich werd SICHERlich nicht, (.)
 20 ähm:: (.) bei nem (.) bei nem EINHORNwert,⁴ (.)
 21 sagen=oh gut,
 22 gemacht? (1.0)
 23 a:ber:;
 24 ich FREU mich,=
 25 =wenn ich den SEhe.

Das Gerät wird zur Projektionsfläche menschlicher Gefühlswelten im Diabetesmanagement und entsprechend auch verbal adressiert. Es wird aber nicht zum interaktionalen Gegenüber. Die (negative) Adressierung bewirkt nichts beim Gerät, sondern hat nur Auswirkungen auf die Sprecherin, die mit der einseitigen Ansprache des Geräts selbstwirksam ist. Die kommunikative Handlung ist hier also eher im Sinne eines Selbstgesprächs als »self-directed speech acts« (Geurts 2018: 280) oder im Hinblick auf ein grenzwertiges Partnermodell als »Selbstorganisation bzw. -reflexion« (Deppermann/Schmidt 2016: 381) zu verstehen. Die Einseitigkeit in der kommunikativen Beziehung wird auch im folgenden Beispiel deutlich, in dem Linda die soziale Beziehung zwischen sich selbst und ihrer Insulinpumpe beschreibt:

Exzerpt 3, Interview mit Linda, w, 39

- 01 L: also es IST, (.)
 02 wie so ne FREUND:schaft?=
 03 =würde ich mal sagen, (-)°hh
 04 ähm_ne SEHR enge FREUNDschaft,=
 05 =man kann nicht ohneinander–

4 Ein »Einhornwert« ist ein idealer Blutzuckerwert von genau 100 mg/dL (5.6 mmol/L). Dieser Wert wird innerhalb der Diabetesgemeinschaft oft scherzhaft als »Einhornwert« bezeichnet, weil er so selten ist wie das mythische Einhorn.

- 06 manchmal KANN_man auch; nicht–
 07 MITeinANDer, (-)°h
 08 aber man WEIß–
 09 trotzdem,=
 10 =was man aneinander HAT. (-)
 11 ich glaub das (.)
 12 beschreibts.; (1.0)
 13 ganz,=
 14 =gut? (.)
 15 woBEI ich natürlich nicht (-)
 16 ((zögert)) also ist natürlich sehr von MIR aus; (.)
 17 ((lachend entschieden)? (.)
 18 ich mein (.) ich hab ne STIMME– (.)
 19 ja–

Die Bestimmung der Mensch-Maschine-Beziehung als – manchmal komplizierte – »FREUNDschaft« (Zeile 4) obliegt dem menschlichen Part. Dies wird von Linda als Entscheidung von »MIR aus« (Zeile 16) betont und damit begründet, dass sie »ne STIMME« (Zeile 18) hat. Impliziert ist dabei, dass das Gerät keine Stimme hat, um sich in der Beziehung zu positionieren. Die Insulinpumpe bleibt sprachlos und damit in der freundschaftlichen Beziehung definitionsunfähig. Während das Gerät zunächst diskursiv in der näheren Bestimmung der freundschaftlichen Dynamik unter das generische Pronomen »man« (Zeile 5 und 6) mit eingefasst wird, differenziert Linda darauf in ein »ich« (Zeile 11) und ein implizites Gegenüber der Pumpe, die den Sonderstatus des Menschen in der Definitionsmacht hervorhebt. Linda reflektiert später im Interview über die Art von Gespräch, die sie mit ihrem Diabetesmanagement-Gerät hat. Vor diesem Ausschnitt berichtet sie darüber, wie die Gespräche mit ihrer Pumpe eher »in ihrem Kopf« stattfinden, wie sie die Pumpe aber auch vor allem in Situationen des »Genervt-Seins« verbal anspricht. Sie differenziert die »Gesprächshaftigkeit« mit dem Gerät dann wie folgt aus:

Exzerpt 4, Interview mit Linda, w, 39

01 L: also:– (1.5)
 02 das ist SCHON auch,
 03 <<nachdenkend> so_n:: (--so_n:: (--))ist schon >=
 04 =hat schon ne GESPRÄCHS:ebene,
 05 auch wenn_es natürlich ANDERS ist,=als wenn man sich–
 06 mit nem MEN:schen unterHÄLT,=
 07 =weil man; (-)°h
 08 ja keine ANTWort, (.)
 09 bekommt, (-)°hh
 10 a:ber::; (-)
 11 ähm::; (.)
 12 man HAT schon, so: (.)
 13 nen AUS:tausch; irgendwie– (.)
 14 geDANKlich zumindestens; (.)
 15 ja? (.) Und wenn es mich NERVT,
 16 ist es eher auch; (-)
 17 verBAL.; dass_es dann (-)°hhh (-)
 18 ist_oder ich– (.)
 19 beZIEHE dann meinen MANN:– (.)
 20 mit ein;=
 21 =ja?

Der Austausch mit dem Gerät wird diskursiv als gesprächsähnlich präsentiert. Das Vorbild für ein ›richtiges‹ Gespräch ist jedoch die *face-to-face*-Kommunikation mit einem anderen Menschen. Das Gespräch mit der Pumpe weicht insofern davon ab, da es erstens keine »ANTwort« (Zeile 8) durch das Gerät gibt, und zweitens der Austausch nur im kognitiven Raum der Sprecherin stattfindet, zumindest solange der Austausch nicht negativ geprägt ist. Wie eingangs ausgeführt, reiht sich Linda ein in die Teilnehmenden, die einen verbalen Austausch mit den Geräten vor allem in emotionalen Situationen der negativen Ansprache fassen. Auffallend bei diesen Ausführungen zur Kommunikation mit den Geräten sind die häufigen Heckenausdrücke (Zeile

2 »SCHON auch«, Zeile 12 »schon, so:«, Zeile 13 »irgendwie«, Zeile 16 »eher«), die Lindas Herleitungen relativieren und einschränken. Es macht deutlich, dass es sich hier nicht um die sprachliche Darstellung eines Fakts handelt, sondern um eine diskursive Aushandlung eines noch nicht final definierten Beziehungsgefüges zwischen Mensch und Maschine. Nicht zuletzt scheint hier auch die menschliche *face-to-face*-Interaktion als Norm durch. Allerdings eröffnet Linda hier einen Verhandlungsraum, der durchaus offen ist, die Interaktion mit den Geräten als »Gespräch« weiterzudenken, auch wenn dieses anderen Gelingensbedingungen unterliegt.

5.2 Technische Affordanzen einer maschineninitiierten Wechselseitigkeit

Die diskursiven Aushandlungen verbleiben nicht bei einer als einseitig gefassten mensch-seitigen Kommunikation, die vor allem auf fehlenden Sin-nenempfindungen (Wahrnehmen, Hören) und fehlender Sprachfähigkeit der Geräte beruhen. Gleichermaßen kommen in den Interviews auch Vorstellungen von Wechselseitigkeit zum Vorschein, die durch die Maschinen in ihren technischen Möglichkeitsräumen initiiert ist, auch bei Sprecher*innen, die die Geräte zuvor (oder später) als sprachlos imaginieren. Linda, die vorher über eine eher einseitige Form der Kommunikation zwischen sich und der Maschine spricht, betont auf die Frage hin, ob ihre Pumpe bei Meldungen vibriert oder einen Alarm gibt, die haptischen und lautlichen Affordanzen der Pumpe, und wie diese sie als Nutzerin zum Handeln bewegen.

Exzerpt 5, Interview mit Linda, w, 39

- 01 L: beides=kommt,
 02 drauf an; (.)
 03 also sie (.) PIEPT, (-)
 04 sie viBRIERT=und wenn sie PIEPT UND vibriert,=
 05 =dann weiß man=jetzt: (.)
 06 is: (-)*hh
 07 und dann GIBTS,=

- 08 =so_n GANZ LAUtes Piepen.=
 09 I: =<<bestätigend> hm>—
 10 L: dann macht;
 11 sie nochmal DEUTlich auf sich AUFmerksam; (.)
 12 I: [hm.]
 13 L: [macht] ↑↑(NIEP NIEP NIEP) (.) und [darauf–] (–)
 14 I: [(lacht kurz)]
 15 L: also da (.) WIRD,
 16 =man auf jeden Fall WACH.

Die Pumpe bekommt hier eine interaktionale Agentivität zugeschrieben, die sich über die Möglichkeit zu »piepen« und zu »vibrieren« (Zeile 4) ausdrückt und von Linda zunächst deskriptiv versprachlicht wird. Das Gerät »macht; [...] DEUTlich auf sich AUFmerksam« (Zeile 10–11) und es warnt ihr Gegenüber im Falle einer Über- oder Unterzuckerung. Linda beschreibt hier auch Abstufungen in den Kommunikationsmöglichkeiten der Pumpe: vom Piepen und Vibrieren zu einem »GANZ LAUte[m] Piepen« (Zeile 8), das von Linda prosodisch im Vergleich zum Ko-Text als in der Lautstärke und Akzentuierung hervorgehobenes »NIEP NIEP NIEP« (Zeile 13) versprachlicht wird. Der Turn, der hier von der Pumpe ausgeht, hat eine Auswirkung auf das Handeln des menschlichen Gegenübers. Es wird »WACH« (Zeile 16) und reagiert im besten Fall auf die kommunizierte Information, dass ein Über- oder Unterzucker vorliegt. Das Gerät wird in diesem Fall diskursiv als Initiator der Kommunikationssituation im Rahmen seiner technischen Möglichkeiten versprachlicht, es wird von der handlungsunfähigen Projektionsfläche oder vom passiven Empfänger kommunikativer Nachrichten sukzessive zum handlungsfähigen Interaktanten. Diese Handlungsfähigkeit des Geräts, die diskursiv von Linda hergestellt wird, spiegelt sich auch in den sprachlichen Mitteln, die sie zur Darstellung wählt. Das Gegenüber der Pumpe wird in der von Linda beschriebenen Situation mit »man« generalisiert (Zeile 5 »dann weiß man«, Zeile 15–16 »also da (.) WIRD,=man auf jeden Fall WACH«). Die Pumpe selbst erfährt durch Lindas Stimme eine prosodisch akzentuierte und individuelle Verkörperung, die in der Ausdrucksform und der dreifachen Frequenz (Zeile 13, »NIEP NIEP NIEP«) der vorher eher neutralen Beschreibung der Pumpenaktivität sowie dem generalisierten »man« gegenübersteht.

Eine sprachliche Konstruktion der Geräte als kommunikationseröffnende Gegenüber bis hin zu handlungsfähigen Interaktanten sehen wir auch in den folgenden Beispielen, die alle aus konkreten Situationen während der Interviews stammen. In allen drei Fällen initiiert das Gerät durch einen Piepton oder nicht-hörbare, aber spürbare, Vibration Kommunikation mit den Nutzenden und ruft bei ihnen eine Reaktion hervor. Die Exzerpte sind besonders komplex, da die sich herausbildende Interaktion mit dem Gerät eingebunden in die Interaktion mit der Interviewerin geschieht und entsprechend konversationell von den Teilnehmenden als bedeutungsvoll bearbeitet wird.

Exzerpt 6, Interview mit Celina, w, 30

- 01 C: ... [sind.] (--)
 02 [[((Pumpe piept))]
 03 I: <<bestätigend> hm_hm,> (--)
 04 C: (ja) da MEckert sie jetzt,=
 05 =gerade

In diesem Beispiel piept die Pumpe, während Celina einen anderen Gedanken im Interview ausführt (Zeile 2). Das Piepen wird von ihr wahrgenommen und sofort in das Interviewgespräch integriert. Mir gegenüber wird es als »meckern« (Zeile 4) versprachlicht und dem Gerät damit eine kommunikative Agentivität zugeschrieben, in dem es seiner Unzufriedenheit Ausdruck geben kann. Die tonale Qualität des eindringlichen und frequenten Piepens bekommt durch die metaphorische Versprachlichung im »Meckern« als stoßweise unterbrochene Laute Valenz. Die Einordnung für die Interviewerin zeigt auch, dass die Beziehung zwischen Celina und ihrem Gerät schon vorab bedeutungsvoll ist. Wenn das Gerät piept, »bedeutet« es immer, dass etwas nicht in Ordnung ist, dass die Nutzerin eine bestimmte Handlung vollziehen muss (Insulindosis anpassen, etwas essen, Batterie wechseln etc.). Sie (die Pumpe) meldet sich nicht, wenn das Blutzuckermanagement im Rahmen der festgelegten Parameter läuft. Diese Bedeutung im Sinne einer medizinisch essenziellen Handlungsaufforderung ist durch Celina gegenüber der Interviewerin mit dem der Pumpe zugeschriebenen, umgangssprachlichen und negativ konnotierten Verb »meckern« abgedeckt.

Eine sprachliche Fassung von maschinengeleiteten Handlungen in eher positiv konnotierten Worten zeigt sich im folgenden Beispiel.

Exzerpt 7, Interview mit Linda, w, 39

- 01 L: ↑und jetzt viBRIERT sie–
 02 ↓lustigerweise geraDE? (-)
 03 I: <<leicht erstaunt> hm_hm?>
 04 L: (1.0) °h und SAGT kalibrieren–
 05 als ob sie es geHÖRT hätte–
 06 =ja;
 07 =kaliBRIEREN bis elf uhr einunVIERZIG. (-)°h
 08 (1.3) ge[nau.
 09 I: [Ah ja.]=also sie GIBT einem auch noch ne stunde zeit; (;)
 10 äh;; (;) um [das;]
 11 L: [jaja]=sie warnt schon mal–=
 12 vor; (;) genau. (-)
 13 I: [hehehehe]
 14 L: mach dich schonmal <<lachend> bereit,>

Auch in diesem Ausschnitt wird das Interview von einer Meldung des Geräts unterbrochen. Das Anzeichen der Pumpe wird zum Anlass einer Integration in den Gesprächskontext des Interviews genommen, der hier einen intertextuellen Bezug bekommt. In der Sequenz vor der Intervention der Pumpe spricht Linda über das Feature des Kalibrierens bei ihrem System. Die maschineninitiierte Kommunikation über die Vibration wird von Linda gegenüber der Interviewerin nicht negativ konnotiert, sondern im Gegenteil als lustiges Phänomen (Zeile 2) benannt. Die Pumpe wird dabei diskursiv als Akteur hergestellt: sie »viBRIERT« (Zeile 1), sie »SAGT« (Zeile 4) etwas, »sie warnt schon mal–=vor; (;)« (Zeile 11–12). Ein konstruierter Dialog findet hier schon zu Beginn der Einbettungssequenz statt. Linda spricht mit der »Stimme« der Pumpe (Zeile 1, 4 und 7) »sie [...] SAGT kalibrieren–als ob sie es geHÖRT hätte–=ja,=kaliBRIEREN bis elf uhr einunVIERZIG. °h«). Interessant ist hier auch, dass im Gegensatz zu den als wahrnehmungsunfähig

konstruierten Geräten (siehe Kapitel 5.1) von Linda eine auditive Wahrnehmung vermutet wird, was auf die thematische Ausformung des Interviews kurz vor der Vibration verweist. Die diskursiv hergestellte Agentivität des Geräts kulminiert im Sprechen von Linda *mit der Stimme* der Pumpe (Tannen 2007: 21) die die Warnung an die Nutzerin als Aufforderung formuliert (Zeile 14, »mach dich schonmal «lachend» bereit,>«).

Im letzten Beispiel unterbrechen zwei aufeinanderfolgende eindringliche Pieptöne Ausführungen des Sprechers während des Interviews.

Exzerpt 8, Interview mit Hendrik, m, 56

- 01 H: [das ist dann halt;]
 02 [((CGM-Empfänger piept))] UPS, (-)
 03 jetzt KLINGELT schon mein [system,] (1.0)
 04 [((CGM-Empfänger piept))]
 05 mein DE el be GE,⁵ (1.0)
 06 aber hab schon was geGESSen gerade, (1.3)
 07 kurz WEGdrücken SO, (.)
 08 o!KAY!, (.)
 09 un:d (-)so WAR der–

Hendrik macht die Meldung des »Systems« zum Gegenstand der Interviewinteraktion und erklärt der Interviewerin, wie genau das System sich nennt, welches »KLINGELT« (Zeile 3). Er blickt auf das Display des CGM-Empfängers, um den Anlass des Klingelns nachvollziehen zu können. Interessant ist hier, dass nicht ganz deutlich ist, an wen sich die folgenden Aussagen von Hendrik richten – das System oder die Interviewerin. Aus einer multimodalen Perspektive, die nicht nur die verbalsprachlichen Phänomene einer Interaktion einbezieht, ist Hendriks Blick für diese Sequenz aufschlussreich (Deppermann 2018: 70). Dieser richtet sich ab »aber hab schon was geGESSen gerade,« (Zeile 6) auf das Gerät, welches im Zuge der sprachlichen Realisierung

5 »DLBC« verweist hier auf einen Algorithmus eines spezifischen Diabetestechnologieanbieters, der die Schnittstelle zwischen Insulinpumpe und Glukosesensor bildet und automatisch die Insulinzufuhr berechnet.

von »mein DE el be GE« (Zeile 5) aus der Hosentasche geholt wird. Es ist, als würde er gleichzeitig zur Interviewerin und zum »System« sprechen, indem er lautlich verdeutlicht, dass die Meldung gerade keinen Handlungsdruck bei ihm auslöst, da er bereits gehandelt hat und entsprechend die Warnung »kurz WEGdrücken« (Zeile 7) kann. Erst mit dem »o!KAY!« wird der Blick wieder auf die Computerkamera und damit zur Interviewerin gerichtet, um das Interviewgespräch nach der von der Maschine initiierten Einschubsequenz fortzuführen.

In allen Beispielen ist die Kommunikation zwischen Mensch und Maschine maschineninitiiert innerhalb derer technischen Affordanzen. Eine Meldung der Geräte in Form von Piepen und Vibration veranlasst das menschliche Gegenüber zu bestimmten Handlungen im Diabetesmanagement oder – wie für den Fall der Interviewkommunikation – auch zu verbalsprachlichen Reaktionen gegenüber Gerät und Interviewerin.

5.3 Geräte als interaktionale Partner in konstruierten Dialogen

Wir haben gesehen, wie Geräte des täglichen Diabetesmanagements von den Nutzer*innen diskursiv als »sprachlose« Projektionsflächen oder Kommunikationsinitiatoren im Rahmen ihrer eigenen technischen Möglichkeiten hergestellt wurden. In den folgenden Beispielen werden die Geräte nun als voll handlungsfähige interaktionale Gegenüber und Beteiligte evident. Dies geschieht über *constructed dialogues* (Tannen 2007), welche die Interviewpersonen in der Interviewsituation enagieren und in welchen sie den Geräten Handlungs- und Wahrnehmungsfähigkeiten sowie eine Stimme zuschreiben. Aus einer diskursanalytischen Perspektive wird hier deutlich, welche Art Stimmen die Geräte von den Nutzer*innen bekommen und wie sie sprechen. Leon berichtet im Interview von den Funktionen seines AID-Systems und inwiefern das Gerät ihn bei einer vergessenen Insulinabgabe warnt.

Exzerpt 9, Interview mit Leon, m, 24

- 01 L: OH, (--)
 02 der BLUTzucker., (.)
 03 der SENSor sagt;=oh, (.)

- 04 dann AH: ja du hast vergessen was ABzugeben==
 05 =dann SCHNELL hinterHER_steuern,

Der Sensor spricht Leon hier direkt mit »du« (Zeile 4) an und gibt ihm eine direkte Handlungsaufforderung. Leon versprachlicht die Rede des Sensors in überraschend kolloquialer Form mit der Interjektion »AH: ja« (Zeile 4) und dem elliptischen Satz »=dann SCHNELL hinterHER_steuern,« (Zeile 5). Diese eher gesprochenen Konventionen folgende sprachliche Realisierung könnte mit ihrer Informalität und Unmittelbarkeit als Form von Nähesprache (Koch/Oesterreicher 1985) interpretiert werden. Ähnliche sprachliche Mittel in der Stimmgebung der Geräte durch die Interviewteilnehmenden finden wir auch in anderen Interviews, wie zum Beispiel bei Hendrik:

Exzerpt 10, Interview mit Hendrik, m, 56

- 01 H: gut das MELdet sich mal dann IRgendwann wenn's halt keinen==
 02 konTAKT mehr hat;=zur_zum; (1.0)
 03 zum SENsor- (-)ce ge EM- (-)
 04 I: <<bestätigend> hm_hm,> (.)
 05 H: sagt EY(-)
 06 ich hab nach; (-)°h
 07 ich hab; (-)ZWANzig MiNuten keinen WERT,(.)
 08 bekommen. (1.3)
 09 jetzt, (.) wo BIST du,

Hendrik spricht hier nicht vom Sensor oder von der Pumpe, sondern vom System (Zeile 1, »das«; »wenn's«). Auch das System bekommt hier eine Stimme, indem es etwas zum Nutzer »sagt« (Zeile 5). Ähnlich wie bei Leon wird das System auch hier mit umgangssprachlichen Interjektionen (Zeile 5, »EY«) und elliptischen Einschüben (Zeile 9, »jetzt«) versprachlicht. In Hendriks Imagination referenziert das System sich selbst mit »ich« (Zeile 7) und beweist seinen Agens-Status über das Wissen um fehlende Werte sowie die Mitteilung dieses Mangelzustands. Das System »fragt« dann auch direkt nach dem Verbleib des

Menschen, der sich um die Aufhebung dieses Zustands zu kümmern hat: »wo BIST du,« (Zeile 9) Auch hier wird das Gerät als vertraut mit dem imaginierten Gesprächspartner versprachlicht, was sich im informellen »du« und der Interjektion nachvollziehen lässt.

An beiden hier gezeigten Ausschnitten, aber auch an anderen Stellen, in denen die Geräte »talking voices« (Tannen 2007) bekommen, ist auffällig, dass sich die Stimmqualität der Sprechenden gar nicht bis kaum verändert, wenn sie *als* Gerät sprechen. Mit der Stimme eines anderen zu sprechen, bedeutet oft eine Abweichung zur eigenen Stimme in Hinblick auf »[...] realizations of pitch, amplitude, intonational contours, voice quality [...]« (Tannen 2007: 22), um den anderen stimmlich auf eine bestimmte Art und Weise zu positionieren. Eine für das menschliche Ohr nicht wahrnehmbare stimmliche Variation im *voicing* des Geräts kann möglicherweise an der engen körperlichen Verbundenheit zwischen Gerät und Mensch liegen. Die Geräte (Sensor und Insulinpumpe) sind am Körper befestigt, werden zuweilen von den Nutzenden selbst als Körperteile konzeptualisiert (Vallentin 2023: 24). Trotzdem sind sie im Hinblick auf interaktionale Wechselseitigkeit gegenüber, die auch als solche diskursiv hergestellt werden müssen und so überhaupt eine Stimme bekommen können.

Das Sprechen mit der gleichen Stimme, obwohl unterschiedliche Sprechende *performed* werden, ist aber nicht durchgehend im Korpus zu finden. Im Exzerpt 11 sind deutliche Unterschiede zwischen Brittas Stimme in Interaktion mit der Interviewerin und der Stimme »als Pumpe« zu erkennen.

Exzerpt 11, Interview mit Britta, w, 64

- 01 B: und- °h (.)
 02 von DAher; (;)
 03 äh;; (-)
 04 JA- (1.0) °h
 05 ich.; (-)
 06 SEhe die kurve,
 07 und SAG dann, (-)
 08 (ich mein)=ich hab_ ja jetzt- (.)

- 09 noch,=mal; (1.5)
 10 die:: (.) siebenhundert_(pf)=äh die; (.)
 11 =wat ist dat=siebenhundertachtzig g?
 12 I: <<bestätigend> hm_hm,> (.)
 13 B: ÄHM:::- (-)
 14 IS natürlich;
 15 NOCH ne ANDere;_äh:::; (1.5)
 16 kategoRIE::=wo die PUMpe,
 17 dann auch sagt; (.) ^h
 18 äh=↑da bist du jetzt zu HOCH;
 19 also das MACHEN wa=jetzt mal wieder RUNter.

Mit deutlich erhöhter Tonhöhe und veränderter Stimmqualität beginnt hier die *Voicing*-Sequenz »äh=↑da bist du jetzt zu HOCH;« (Zeile 18). Die Tonhöhe »der Pumpe« sinkt deutlich bei »also das MACHEN wa=jetzt mal wieder RUNter.« (Zeile 19), bis Brittas stimmliche Ausgangslage wieder erreicht ist und das *Voicing* der Pumpe in der Interviewsituation beendet. Das »wir«, das für die Pumpe gesprochen wird (Zeile 19, »wa«), suggeriert zum einen eine gewisse Kooperation zwischen Mensch und Maschine im gemeinsamen Vorhaben des Blutzuckersenkens. Es erinnert aber auch im Rahmen der steigend und sinkend stark markierten Prosodie und eher belehrender Stimmqualität an das Sprechen einer (medizinischen) Autorität, die überhaupt erst in der Lage ist, ein »wir« für die Aufgabe anzusetzen (Mostovaia/Fedorovskaya/Imo 2023: 119). Neben der eher auffälligen Prosodie finden sich auch in diesem Exzerpt Interjektionen (Zeile 18, »äh«, Zeile 19, »also«), informelle Anrede (Zeile 18, »du«) und Modalpartikeln wie »mal« (Zeile 19), die eher auf nächsprachliche Konzeptionen in der Imagination der*des Anderen als Interaktionspartner hinweisen.

Die Geräte erhalten in diesen Sequenzen geschlechtsspezifische Pronomen, je nachdem ob »die« Pumpe, »der« Sensor oder »das« System etwas »sagt«. Die Geräte werden als eher in Umgangssprache sprechend *performed*, was sich an den Modalpartikeln, Interjektionen und konversationellen Elementen abzeichnet. Sie verwenden eine Sprache der Nähe, indem sie ausschließlich das informelle »du« für ihr menschliches Gegenüber verwenden. Die Geräte agieren hier in einer imaginierten Wahrnehmungswahrnehmung und sind

in den konstruierten Dialogen vollwertige soziale Interaktionspartner, die ähnlich sprechen wie ihre Nutzer*innen.

6 Diabetesgeräte als (un)mögliche Interaktionspartner – Ein Kontinuum

Die Analyse der sprachlich-diskursiven Praktiken hat gezeigt, dass die Konzeptualisierung von Interaktion und die Wahrnehmung von Diabetesgeräten als (un)mögliche Interaktionspartner aus Nutzer*innenperspektive ein komplexes Beziehungsgefüge ist, das sich durch eine deutliche Heterogenität und Varianz auszeichnet. Diese Heterogenität besteht nicht nur über unterschiedliche Individuen, sondern auch innerhalb von Interviews mit der gleichen Person. Die Geräte werden im Interviewkontext entsprechend situativ hervorgebracht, mal als »Ding« im Sinne einer Aufgabenlogik, mal als »Freund ohne Stimme«, mal als interaktiver »Gesprächspartner«, der soziale Nähe herstellt. Eine uneindeutige Festschreibung und Konzeptualisierung der Maschinen als Interaktionspartner lässt sich in den Interviews, vergleichbar mit der in Kapitel 2 und 3 eingeführten Theoriedebatte, nicht ausmachen. Vielmehr zeigt sich hier ein fortwährender und teilweise widersprüchlicher Prozess der Hervorbringung und Aushandlung unterschiedlicher Interaktions- und Partnerkonzepte in Bezug auf die Geräte. Die klassischen Grenzen zwischen Kommunikation und Interaktion – insbesondere die Unterscheidung zwischen unidirektionaler Botschaftenübermittlung und wechselseitiger Beziehungs- und Bedeutungskonstitution – sowie die damit verbundenen Zuschreibungen zu den technischen und sozialen Eigenschaften der Geräte werden durch die diskursiven Aushandlungen der Nutzer*innen herausgefordert. Die Erkenntnis der Teilnehmenden, dass die Maschinen keine fühlende Wesen mit bestimmten Sinneswahrnehmungen und Wissen sind, dass sie nicht »bewusst« und »intentional« handeln können, steht im Widerspruch zu sich (trotzdem) aufbauenden und teilweise tiefen sozialen Beziehungen (»wir duzen uns«; »Freundschaft«).

Sichtbar wird ein Kontinuum, welches sich von entweder menschen- oder maschineninduzierter Kommunikation bis hin zu interaktiv(er) wahrgenommener und performter Wechselseitigkeit erstreckt. Noch komplexer wird diese Form von Kommunikations-Interaktionskontinuum, da die Geräte zum Diabetesmanagement in einer medizinisch essenziellen Verbindung mit den Nutzenden stehen. Die Aufgabe, die gemeinsam bewältigt werden muss – den

Blutzucker in einem bestimmten definierten Bereich zu halten – ist gesundheitsbestimmend und im Ernstfall lebensnotwendig. Dies wirft die Frage auf, ob mögliche situationale Agentivität für solche Maschinen anders und eventuell stärker hervorgebracht wird als für zum Beispiel Smart Watches oder Computerprogramme. Auch der Bezug der Maschinen zu den Körpern der Betroffenen ist hier mitzudenken. Anders als andere Nahkörpertechnologien (Kaerlein 2018) wie das Smartphone sind die Geräte direkt mit dem Körper verbunden, sie durchdringen ihn gar oder werden als eigenständige Körperteile wahrgenommen (Valentin 2023: 24). Gerade dann müssen die Geräte in Momenten der Interaktion oder in konstruierten Dialogen mit den Geräten als (mehr oder weniger) partnerschaftliches Gegenüber versprachlicht werden.

Ein Blick auf die diskursiven Suchbewegungen der Teilnehmenden zur sprachlichen Konstruktion ihrer Geräte in Interaktionen zeigt letztlich auch den fühlenden, sozialen und sich in Beziehung setzenden Menschen. Die Möglichkeit, sich mit seinem medizinisch notwendigen Gerät in (eine interaktiv imaginierte) Beziehung setzen zu können hilft, den Alltag mit der Krankheit besser zu navigieren. Es lässt die Geräte weniger als *Fremdkörper* erscheinen, sondern als Teile vertrauter menschlicher Beziehungen. Der Begriff der Interaktion erweitert sich damit in Räume imaginiertes Anthropomorphisierung, aber auch in die ganz konkrete Konstruktion sozialer Beziehungen zu technischen medizinischen Hilfsmitteln. Die damit einhergehenden Entgrenzungen sprachwissenschaftlicher Grundannahmen zu Interaktion und Kommunikation sowie der an ihnen Beteiligten eröffnen somit Horizonte für das Verständnis dynamischer und hybrider Austauschprozesse, die traditionelle anthropozentrische Paradigmen überwinden und technologische sowie posthumanistische Perspektiven integrieren.

Literaturangaben

- Boy, Guy A. (2011): Introduction. A Human-Centered Design Approach. In: *The Handbook of Human-Machine Interaction: A Human-centered Design Approach*, Farnham, Surrey, England; Burlington, VT: Ashgate, 1–20.
- Cameron, Deborah (2001): *Working With Spoken Discourse*, London: Sage.
- Coeckelbergh, Mark (2011): Humans, Animals, and Robots: A Phenomenological Approach to Human-Robot Relations. In: *International Journal of Social Robotics* 3:2, 197–204. doi.org/10.1007/s12369-010-0075-6.

- Coeckelbergh, Mark (2020): When Machines Talk: A brief Analysis of some Relations between Technology and Language. In: *Technology and Language* 1:1, 22–27. doi.org/10.48417/TECHNOLANG.2020.01.05.
- Couper-Kuhlen, Elizabeth/Margret Selting (2018): *Interactional Linguistics. Studying Language in Social Interaction*, Cambridge: Cambridge University Press.
- Deppermann, Arnulf (2018): Sprache in der multimodalen Interaktion. In: Deppermann, Arnulf/Silke Reineke (Hg.), *Sprache im kommunikativen, interaktiven und kulturellen Kontext*, Berlin/Boston: De Gruyter, 51–85.
- Deppermann, Arnulf/Axel Schmidt (2016): Partnerorientierung zwischen Realität und Imagination: Anmerkungen zu einem zentralen Konzept der Dialogtheorie. In: *Zeitschrift für germanistische Linguistik* 44:3, 369–405. doi.org/10.1515/zgl-2016-0021.
- Ehrmantraut, Luise (2020): »Spinnt sie« oder »Spinnt die«? Konkurrenten pronominaler Wiederaufnahme und ihre Verwendungsbedingungen im gesprochenen Deutsch. Examensarbeit an der Universität des Saarlandes.
- Esposito, Elena (2017): Artificial Communication? The Production of Contingency by Algorithms. In: *Zeitschrift für Soziologie* 46:4, 249–265. doi.org/10.1515/zfs0z-2017-1014.
- Fischer, Kerstin (2016): *Designing Speech for a Recipient: The Roles of Partner Modeling, Alignment and Feedback in so-Called »Simplified Registers«*, Amsterdam: John Benjamins Publishing Company.
- Fortunati, Leopoldina/Autumn Edwards (2020): Opening Space for Theoretical, Methodological, and Empirical Issues in Human-Machine Communication. In: *Human-Machine Communication* 1, 7–18. doi.org/10.30658/hmc.1.1.
- Geurts, Bart (2018): Making Sense of Self Talk. In: *Review of Philosophy and Psychology* 9:2, 271–85. doi.org/10.1007/s13164-017-0375-y.
- Goffman, Erving (1967): *Interaction Ritual. Essays on Face-to-Face Behaviour*, New York: Pantheon.
- Gordon, Cynthia (2011): Conversation and Interaction. In: Mesthrie, Rajend (Hg.), *The Cambridge Handbook of Sociolinguistics*, Cambridge University Press, 105–121.
- Gouvrit, Florence (2013): Empathy and Human-Machine Interaction. In: *International Journal of Synthetic Emotions* 4:2, 8–21. doi.org/10.4018/ijse.2013070102.
- Habscheid, Stephan/Tim Hector/Christine Hrnca (2023). Human and Non-Human Agency as Practical Accomplishment: Interactional Occasions for

- Ascription and Withdrawal of (Graduated) Agency in the Use of Smart Speaker Technology. IN: *Social Interaction. Video-Based Studies of Human Sociality*, 6(1). <https://doi.org/10.7146/si.v6i1.137378>
- Hausendorf, Heiko (2015): Interaktionslinguistik. In: Eichinger, Ludwig M. (Hg.), *Sprachwissenschaft im Fokus*, Berlin/München/Boston: De Gruyter, 43–70.
- Hausendorf, Heiko/Reinhold Schmitt (2018): Sprachliche Interaktion im Raum. In: Deppermann, Arnulf/Silke Reineke (Hg.), *Sprache im kommunikativen, interaktiven und kulturellen Kontext*, Berlin/Boston: De Gruyter, 87–118.
- Hoc, Jean-Michel (2010): From Human-Machine Interaction to Human-Machine Cooperation. In: *Ergonomics* 43:7, 833–843. doi.org/10.1080/001401300409044.
- Höflich, Joachim R (2013): Relationships to Social Robots: Towards a Triadic Analysis of Media-Oriented Behavior. In: *Intervalla* 1, 35–48.
- Kaerlein, Timo (2018): *Smartphones als digitale Nahkörpertechnologien. Zur Kybernetisierung des Alltags*, Bielefeld: transcript.
- Koch, Peter/Wulf Oesterreicher (1985): Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. In: Deutschmann, Olaf/Hans Flasche/Bernhard König/Margot Kruse/Walter Pabst/Wolf-Dieter Stempel (Hg.), *Romanistisches Jahrbuch* 36, Berlin/New York: De Gruyter, 15–43.
- Leblebici, Didem (2024): ›You are Apple, why are you speaking to me in Turkish?‹: The role of English in voice assistant interactions. In: *Multilingua* 43:4, 455–485. doi.org/10.1515/multi-2023-0072.
- Lind, Miriam (Hg.) (2022): *Mensch – Tier – Maschine: Sprachliche Praktiken an und jenseits der Außengrenze des Humanen*, Bielefeld: transcript.
- Luginbühl, Martin (2019): Mediale Durchformung. Fernsehinteraktion und Fernseh­mündlichkeit in Gesprächen im Fernsehen. In: Marx, Konstanze/Axel Schmidt (Hg.), *Interaktion und Medien. Interaktionsanalytische Zugänge zu medienvermittelter Kommunikation*, *OraLingua*, Heidelberg: Winter, 125–146.
- Luhmann, Niklas (1981): Kommunikation mit Zettelkästen: Ein Erfahrungsbericht. In: Baier, Horst/Hans Mathias Kepplinger/ Kurt Reumann (Hg.): *Öffentliche Meinung und sozialer Wandel: Für Elisabeth Noelle-Neumann*, Opladen: Westdeutscher Verlag, 222–228.
- Luhmann, Niklas (1984): *Soziale Systeme: Grundriß einer allgemeinen Theorie*, Frankfurt a.M.: Suhrkamp.

- Luhmann, Niklas (2008): *Einführung in die Systemtheorie*. 4. Aufl, Baecker, Dirk (Hg.), Heidelberg: Carl-Auer-Verlag.
- Mostovaia, Irina/Victoria Fedorovskaya/Wolfgang Imo (2023): *Wir Beide und Мы с Вами* (›wir mit Ihnen‹): Strategien zur Vagheitsreduktion im Gebrauch des Personalpronomens der 1. Person Plural in Deutschen und Russischen Diagnosemitteilungs- und Therapieplanungsgesprächen. In: *Zeitschrift für Germanistische Linguistik* 51:1, 88–123. doi.org/10.1515/zgl-2023-2003.
- Pennycook, Alastair (2018): *Posthumanist Applied Linguistics*. First edition, London/New York: Routledge/Taylor & Francis Group.
- Pitsch, Karola (2019): Referential Practices for a Museum Guide Robot. Human-Robot-Interaction as a Methodological Tool to Investigate Multimodal Interaction. *Workshopband MuC*, 366–368. doi.org/10.18420/MUC2019-WS-661.
- Pustejovsky, James/Nikhil Krishnaswamy (2021): Embodied Human Computer Interaction. In: *KI – Künstliche Intelligenz* 35:3-4, 307–327. doi.org/10.1007/s13218-021-00727-5.
- Saussure, Ferdinand de (2016): *Grundfragen der allgemeinen Sprachwissenschaft: eine Auswahl*, Jahraus, Oliver (Hrsg), Stuttgart: Reclam.
- Schleidgen, Sebastian/Orsolya Friedrich/Selin Gerlek/Galia Assadi/Johanna Seifert (2023): The Concept of ›Interaction‹ in Debates on Human–Machine Interaction. In: *Humanities and Social Sciences Communications* 10:1, 551. doi.org/10.1057/s41599-023-02060-8.
- Schmidt, Philipp/Sophie Loidolt (2023): Interacting with Machines: Can an Artificially Intelligent Agent be a Partner? In: *Philosophy & Technology* 36:3, 55. doi.org/10.1007/s13347-023-00656-1.
- Schneider, Britta (2022): Multilingualism and AI: The Regimentation of Language in the Age of Digital Capitalism. In: *Signs and Society* 10:3, 362–387. doi.org/10.1086/721757.
- Searle, John R. (1969): *Speech Acts, an Essay in the Philosophy of Language*, Cambridge: Cambridge University Press.
- 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 Mertzluft/Christian Meyer/Miriam Morek/Frank Oberzaucher/Jörg Peters/Uta Quasthoff/Wilfried Schütte/Anja Stukenbrock/Susanne Uhmann (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10,

- 353–402. Abrufbar unter: <https://www.gespraechsforschung-online.de/fileadmin/dateien/heft2009/px-gatz.pdf>
- Shannon, Claude E. (1968): *The Mathematical Theory of Communication*. In: *The Mathematical Theory of Communication*, Urbana: Univ. of Illinois Press, 31–125.
- Shannon, Claude E./Warren Weaver (1964): *The Mathematical Theory of Communication*, Urbana: Univ. of Illinois Press.
- Su, Hang/Wen Qi/Jiahao Chen/Chenguang Yang/Juan Sandoval/Med Amine Laribi (2023): Recent Advancements in Multimodal Human–Robot Interaction. In: *Frontiers in Neurorobotics* 17:1084000. doi.org/10.3389/fnbot.2023.1084000.
- Tannen, Deborah (2007): *Talking Voices*. 2nd ed, Cambridge: Cambridge University Press.
- Trujillo, Anna C./Irene M. Gregory/Kasey A. Ackerman (2019): Evolving Relationship between Humans and Machines. In: *IFAC-PapersOnLine* 51:34, 366–71. doi.org/10.1016/j.ifacol.2019.01.015.
- Vallentin, Rita T. (2023): Erste Ergebnisse einer Interviewstudie zur Selbstwahrnehmung von Menschen mit Typ-1-Diabetes. In: *Diabetes Congress Reports* 2, 22–26.
- Weaver, Warren (1964): Recent Contributions to the Mathematical Theory of Communication. In: *The Mathematical Theory of Communication*, Urbana: Univ. of Illinois Press, 1–28.

Dr. Rita Tamara Vallentin
Europa-Universität Viadrina
Große Scharrnstrasse 59
15230 Frankfurt (Oder)
vallentin@europa-uni.de
<https://www.kuwi.europa-uni.de/de/professuren-mitarbeitende/wimi-vallentin/index.html>

Sprache, Sprechen und Interagieren in mediatisierten Kontexten

Kommunikationspraktiken mit Vertreter*innen der Spezies Oktopus

Filmische, literarische und naturwissenschaftliche Evidenz

Konstanze Jungbluth

Abstract Wer den Film *My Octopus Teacher* (Ehrlich/Reed, ZA 2020) sieht, kann selbst den Aufbau der Paarbeziehung zwischen dem Taucher und Autor Foster und »einem Oktopus-Weibchen« (*Octopus* 2020 min. 01:06:10), einem flüssigen »gallertartigen Weichtier mit einer hohen Intelligenz« (*Octopus* 2020 min. 00:22:40-00:22:44), in Südafrika miterleben. »Man muss gewillt sein zuzuhören« (Montgomery 2017: 157). Diese Kontaktaufnahme wird als multimodal beschrieben, als ein visuelles Ereignis (Augenkontakt), als Berührung (Körperkontakt), als räumliche Kontingenz (»Du darfst jetzt in meine Oktopuswelt kommen«, *Octopus* 2020 min. 00:28:05-00:29:08). *Liquide Körper, liquide Praktiken und sogar liquide Sprachen* (Schneider im Druck; Canagarajah 2021) bestätigen das Potential posthumaner Denkansätze »whose turn to be liquefied has now come« (Bauman 2000: 6). Sie zeigen den Sprachgebrauch im kommunikativen Raum der Mensch-Tier- und auch der Mensch-Mensch-Interaktionen als entgrenzt. »Man überschreitet eine Grenze, wenn man sich in das Leben von Tieren einmisch« (*Octopus* 2020 min. 00:45:20-00:45:25).

Keywords Interaktion; Kommunikation; Human Animal Studies; Liquidität; Agency

dí núnýá {search knowledge} ›be curious‹ (Ewe)
(Yakpo im Druck)

1 Einleitung

Mein Beitrag hat zum Ziel, die Praxis ›Sprache zu entgrenzen‹ als Teil einer liquiden Praxis am Beispiel der menschlich-tierlichen Interaktion zu zeigen, die auch im Kontext der kulturwissenschaftlich geprägten *Human Animal Studies* erforscht wird. Allerdings wurde den von mir ausgewählten Meerestieren bislang kaum Beachtung geschenkt, vielleicht weil sie auf einem Human-Animal-Kontinuum – jedenfalls im Fall von Oktopoden – nicht in direkter Nachbarschaft zu *Humans* angesiedelt werden würden. Tierpfleger*innen, Besucher*innen in Aquarien und Taucher*innen selbst in freier Natur jedoch vergleichen ihre interaktiven Erfahrungen mit einzelnen Individuen der Spezies Oktopus mit solchen, die sie mit einem Haustier (»a cat or a dog« Octopus 2020¹ min. 00:34:58-00:35:01) oder sogar mit einem Kind (Montgomery 2022: 66) gesammelt haben. In unerwartet pointierter Form erfinden Mensch-Oktopus-Paare multimodale Praktiken, für die im und am Wasser visuelle, taktile und olfaktorische Mittel neben anderen bedeutsam sind. Voraussetzung für solche interaktiven Beziehungen ist nicht nur das Interesse füreinander, sondern auch die beidseitige Anerkennung von *Agency*, die eigenständiges Handeln zwischen den Individuen erst möglich macht.

Welche Evidenz für die unter zwei Individuen artenübergreifend entwickelten Kommunikationspraktiken und ihre Wirksamkeit kann die Zusammenschau eines Kinofilms mit einem literarisch geprägten Sachbuchtext und naturwissenschaftlichen Quellen hervorbringen? Aus einem Interesse an der Interaktion sind die folgenden Fragen bedeutsam: Ist die seit Aristoteles beobachtete Neugier des Oktopus in der Mensch-Tier-Begegnung ein Zeichen für Dummheit (Aristoteles zitiert nach Grasso 2014) oder für seine Intelligenz (Grasso 2014)? Was für eine Rolle spielt die unterschiedliche Medialität der Quellen, die einerseits schriftlich im Fall des Sachbuchs und der naturwissenschaftlichen Artikel die Rezeption als Leser*in (*Readership*), andererseits audiovisuell im Fall des Films die Rolle die*der Zuschauer*in (*Spectatorship*)

1 ›Ich begleitete ihn etwa 80 % seines Lebens‹ (01:08:40-1:08:45); ›Seine Intelligenz ist vergleichbar mit Katzen, Hunden oder sogar niederen Primaten. Dabei sind sie Weichtiere.‹ (00:34:58-00:35:01).

auslösen? Inwiefern stellen die Ergebnisse die Annahme in Frage, Sprache sei abschließend als Sprachsystem zu begreifen (Saussure 1916; Chomsky 1957)? Ich möchte zeigen, welche Dimensionen dieser traditionelle Blick ausgeblendet hat und das Potential posthumaner Denkansätze »whose turn to be *liquefied* has now come« (Bauman 2000: 6; Hervorhebung KJ; vgl. Schneider im Druck) im Folgenden ausloten.

Der Beitrag umfasst fünf zentrale Kapitel, die folgende Themen entfalten: Individuelle *Agency* des Mensch-Oktopus-Paares; liquides Zeichenrepertoire mittels Berührung, farbiger Muster und Geruchssinn; mediale Perspektiven liquiden Sprechens; Fiktion im Film: lebenslang ein Paar?; tierlich-menschliche Neugier im liquiden Raum. Schließlich führe ich die Erkenntnisse zu liquiden Körpern, liquider Praxis und liquiden Sprachen zusammen und weise auf interdisziplinär einzulösende Forschungsdesiderata hin.

2 Agency

Anschließend an die auch in den Kulturwissenschaften geführte Diskussion zum historischen Wandel von *Agency* wird auch in den Studien der *Human Animal Studies* (HAS) dieser Begriff zentral gesetzt (»key concept« bei Böhm/Steen 2023: 15). Dem Tier bzw. dem Lebewesen schlechthin wird eine ihm eigene Handlungs- und »Wirkungsmacht« (Roscher 2015) zugeordnet, die auch in die durch menschliche Fürsprecher*innen geführte Verteidigung² der für den Erhalt oder die Veränderung seines eigenen Lebensraums bedeutenden Rechte (Singer 1975)³ münden kann. Basierend auf dieser dem tierlichen Individuum und seiner Gruppe zustehenden Selbständigkeit ist eine Begegnung mit dem

2 Vgl. »Guardianship« (Stone 1972). Ich danke Moritz Wiegand für diesen Hinweis.

3 Nicht zu verwechseln ist die hier entfaltete, in den Sozial- und Geisteswissenschaften eingeführte Bedeutung von *Agency* mit dem in der Diskursdomäne der Rechtswissenschaften verwendeten Begriff. Dort werden mit *Agency* die Regelung von Beziehungen zwischen Akteur*innen, die im Auftrag anderer tätig werden, gefasst, beispielsweise Spediteure, »forwarding agents« im britischen Kontext, »commissionaire de transport« in Frankreich. Ihre rechtliche Position als Vermittelnde zwischen Auftraggeber*innen und Kund*innen hinsichtlich unterschiedlicher Selbständigkeiten und darauf aufbauender Verantwortung ist in verschiedenen europäischen Ländern unterschiedlich gelöst (Zweigert/Kötz 1996). Sogenannte »Brokers« oder Makler*innen und ihre Rechte sind ein anderes Beispiel für solche Mittler*innen. Zur historischen Handlungsmacht von »Brokers«, vgl. Jungbluth im Druck.

Menschen auf Augenhöhe erst möglich. So reflektiert Foster im Film: »Dieser Oktopus lehrte mich, dass ich kein Besucher bin, sondern ein Teil dieser Welt, ein enormer Unterschied« (Oktopus 2020 min. 01:22:43–01:22:45).

Aufbauend auf dem von Bourdieu (1979) eingeführten Begriff des ›Habitus‹, Habermas' (1995) Konzept des ›kommunikativen Handelns‹ und Reckwitz (2016) Verständnis der ›sozialen Praxis‹ wurde der Begriff der *Agency* in den Sozial- und Kulturwissenschaften etabliert. Insbesondere in den Politikwissenschaften wird mit diesem Begriff die zu überwindende Praxis der Einflussnahme einer politischen Elite auf eine aktive Veränderung gesellschaftlich relevanter Verhältnisse bezeichnet, die verkörpert durch staatliche Verantwortung tragende Männer, aber auch durch die im Marxschen Sinne verstandene Arbeiterklasse in historisch überwundenen Zeiten beobachtet werden kann (vgl. ›structure/agency‹ Raithelhuber 2008). Mindestens seit Martin Luther Kings Rede 1963 und den Aufbruchsbewegungen in der Folge der 1968er Jahre in Europa wurde diese Handlungsmacht nach und nach auch auf Mitglieder der zuvor ausgeschlossenen Gruppen erweitert, die sich durch ihre ethnische Zugehörigkeit oder ihr Geschlecht, ihre Religion oder andere soziale Merkmale von den traditionellen, bis dahin die gesellschaftliche Entwicklung, ihre Strukturen und ihre Handlungspläne maßgeblich bestimmenden Personen und Gruppen unterscheiden.

Im Kontext des Klimawandels, der nicht nur für Menschen, sondern auch für Tiere und Pflanzen gravierende Folgen hat, wird im Rahmen posthumaner Forschung das Konzept der *Agency* auch für andere Lebewesen geltend gemacht (Oliver 2017). In diesem Beitrag zu den Oktopus-Mensch-Begegnungen schließe ich insbesondere an den Aspekt der ›autonom als Kooperationspartner*innen handelnden [tierlichen] Individuen‹ (Geese 2015: 227; Ergänzung KJ) als Ausweis von *Agency* an. Geese (2015; 2017) entfaltet am Beispiel der Führungshunde tierliche ›Agency in Mensch-Tier-Triaden⁴. Diese Triaden bestehen aus zwei Menschen und dem für die Begleitung von Menschen mit eingeschränktem Sehvermögen ausgebildeten Hund. Die wechselseitig eingeräumte Handlungsmacht, ihre reziproke *Agency*, ist sowohl für die Beziehungen zwischen Menschen als auch für solche zwischen Mensch und Tier bedeutsam, wenn sie gemeinsam kommunikativ handeln wollen. Die Konstruktion des Selbst (›Self‹) und der* des Anderen (›Other‹) als Kommunikationspartner*in basiert auf der

4 Geese (2015; 2017) richtet ihr Interesse auf die Aushandlung der bei solchen dreieitigen Konstellationen ausgedrückten Stigmata, die manchmal gegenüber der visuell eingeschränkten Person geäußert werden.

reziproken Anerkennung der Individualität der in dieser Beziehung verbundenen Akteur*innen (›agents‹ Glasersfeld 1991: 113–128). In den im Film *My Octopus Teacher* (Ehrlich/Reed, ZA 2020) und im Sachbuch *Rendezvous mit einem Oktopus* (Montgomery 2020) dargestellten tierlich–menschlichen Treffen (›encounters‹) hängen die tatsächlich vollzogenen Handlungen auch von den Rollen der Interagierenden (›relationship‹) und den damit verknüpften Hierarchien ab, die auf Machtverhältnisse verweisen und nicht ausgeblendet werden dürfen⁵. Vielmehr bestimmt der immer wieder neu zu verhandelnde Spielraum die zwischen den Partner*innen aufgebaute Beziehung (›partnership‹).

Ich möchte mich im Folgenden der Frage nähern, welche Art von *Agency* tatsächlich in der Oktopus-Mensch-Begegnung beobachtet werden kann, indem ich bereits jetzt weitere in meinem Beitrag zur Sprache kommenden Akteur*innen mit ihren jeweiligen Rollen vorstelle. Im Unterschied zu der Triade, von der Geese (2015) als einem gleichzeitigen Zusammentreffen dreier Akteur*innen am gleichen Ort spricht, bilden die Rezipient*innen des Films oder des Sachbuchs ihrerseits eine Triade mit den handelnden menschlich–tierlichen Paaren, die ihnen auf der Leinwand oder auf den Buchseiten begegnen. Die Leserin begegnet der Besucherin im Aquarium und dem Oktopus Weibchen Athena nicht in Wirklichkeit, sondern im Text an einem von der Leserin gewählten Ort zu einer von ihr bestimmten Zeit (›zerdehnte‹ Kommunikation am Beispiel der Briefkorrespondenz und der Boten: Ehlich 1994; Jungbluth 2016b). Ihre *Agency* bezieht sich auf die Rezeption im Feld des Mediums Film und in dem der Schriftlichkeit. Letztere beruht auf Alphabetisierung und Lesekompetenz, die mit gesammelter Erfahrung verknüpft ist: ›readership‹. Auch Filmrezeption beruht auf Medienerfahrung, die unter anderem auf dem Verständnis des cineastischen Aufbaus und des Bild und Ton gestützten ›story telling‹ beruht. Die früher irrtümlich als passiv verstandene Rolle der auch emotional angesprochenen Kinobesucher*innen wird daher treffend von Greifenstein und Schmitt (2014) als ›spectatorship‹ gefasst. Auch hier findet die Triade nur im Kopf der Filmzuschauer*innen statt, in ihrer Vorstellung. Diese beiden Formen von *Agency* wurzeln in kulturellen sozialen Praktiken, die jahrelange Ausbildung voraussetzen und den Menschen vorbehalten bleiben. Das Gleiche

5 Auch im Tierreich spielt Macht artenübergreifend eine Rolle, z.B. zwischen Oktopus und Hai: »Man sieht, dass das Oktopus Weibchen jetzt die Oberhand hat [als es ihr gelingt, auf dem Rücken des Hais sitzend], ihn komplett [zu] überliste[n]« (min. 01:05:45–01:05:49).

gilt nicht für die Rolle aktiver Beobachter*innen, die ich als ›observership‹⁶ bezeichne, und die nicht nur für den Taucher und die Besucherin im Aquarium, sondern auch für ihre jeweiligen tierlichen Gegenüber geltend gemacht werden kann. Mit Anderson et al. (2010: 270) verstehe ich ihre Rolle als eine mögliche Form der mit dem Begriff ›relationship‹ gefassten Beziehung zwischen den Akteur*innen.

Was im vorliegenden Kontext der Mensch-Oktopus-Kommunikation strittig ist, bezieht sich auf die spezifisch menschlichen Äußerungsformen der Sprache und des Denkens (›propriatorship of language and thought‹ Massumi 2014). Meiner Meinung nach spiegelt sich darin eine seit alteuropäischen Zeiten tradierte Sichtweise, die beispielsweise in der von Aristoteles überlieferten Bewertung des Oktopus als einem ›dummen‹ Tier repräsentiert wird: ›Aristotle denigrated octopuses' intelligence for behaviors [as for example stretching out his arms and investigating the human with his suckers...] that might be called ›curiosity‹« (Grasso 2014: 94; Ergänzung KJ; vgl. Yakpo im Druck). Grasso selbst betont »that cephalopods possess cognition« (2014: 94) und führt weiter aus:

His [Aristotle's] definition would be too anthropomorphic to admit an intelligence that did not agree with the typical social [!] conceptions of intelligence in humans. *The definition [for intelligence] that I will use [...] is broader: any adaptive process which connects sensing in the world to action in the world.* The word adaptive in this definition requires some elaboration. I use adaptive in the sense that it is used in the field of cybernetics: that of a process that changes with experience. (Grasso 2014: 95; Ergänzungen und Hervorhebung KJ; vgl. ›embodied mind‹ Gallese/Lakoff 2005 beispielhaft verkörpert durch individuelles »sensing in the world«)

[In octopus evolution,] neurons concentrated at the front of the animal, forming something more and more like a definite brain. Further, much of a cephalopod's nervous system is not found within the brain at all, but spread

6 Dieser Begriff wird bislang vor allem in den Politikwissenschaften für den Beobachterstatus im politischen Kontext verwendet, etwa wenn noch nicht der EU zugehörige Staaten berechtigt sind, an bestimmten Beratungen als Zuhörende teilzunehmen. Ich erweitere den Begriff hier auf Tierbeobachtungen durch Menschen und Menschbeobachtungen durch Tiere. Als Beispiel sei die Vogelbeobachtung als soziale Praxis erwähnt, die besonders von zahlreichen (Hobby-)Ornitholog*innen aus Großbritannien ausgeübt wird. Diese Expert*innen zeichnen sich ebenfalls durch ›observership‹ aus.

through the body [where his/her eight arms are particularly prominent].
(Godfrey-Smith 2018: 67; Ergänzung K)

Ergänzend sei darauf hingewiesen, dass die molekularbiologische Forschung ebenfalls den Fokus auf *Agency* legt, worunter sie Evidenz für den Energiefluss zwischen Zellen fasst (Giesen et al. 2020). ›Doing‹ und ›action in the world‹, also mit Energie verknüpftes Handeln über die Grenzen der Zellwände hinweg oder zwischen den Subjekten kann als gemeinsamer Kern der Bedeutungen festgehalten werden und weist zugleich auf seinen liquiden Charakter hin.

3 Liquides Zeichenrepertoire mittels Berührung, farbigen Mustern und Geruchssinn

In diesem Kapitel möchte ich die materiellen und verkörperten Aspekte der Interaktion zwischen Menschen und Oktopoden in den Mittelpunkt stellen⁷. Mir scheint es bedeutsam, dass es sich im Film wie auch im »erzählenden Sachbuch« (Montgomery 2022: 4) nicht um frisch geschlüpfte, sondern um von mir vorläufig als adoleszent bezeichnete Oktopoden handelt, die im Lauf der Erzählung geschlechtsreif werden und sterben. »[Octopuses] have very short lives: just one or two years« (Godfrey-Smith 2017: 159). Es ist schon erstaunlich, dass die Individuen dieser Spezies in einer im Vergleich zu der humanen Entwicklung kurzen Zeit und quasi im Selbststudium einen derart kompetenten Umgang in ihrer Habitat ausbilden können⁸, wie er im Film und im Buch dargestellt wird. »Dass ein Lebewesen derart schnell denken und überlebensnotwendige Entscheidungen treffen kann, ist einfach unglaublich« (Oktopus 2020 min. 01:07:46-01:07:55). Wird im Mensch-zu-Mensch-Kontext sprachlicher Austausch, der Dialog mit lautlichen Mitteln, regelmäßig zentral gesetzt, so treten in der Mensch-Oktopus-Begegnung im liquiden Raum nicht nur Berührungen an ihre Stelle, sondern es kommen auch andere Mittel hinzu, beispielsweise Farben und Muster des ganzen Körpers, die neben olfak-

7 Ich danke Dorothea Horst für ihre Hinweise zu diesen Aspekten.

8 »[Die geschlüpfte Oktopode als Jungtiere] verkörpern Autopoiesis, da ihnen die Eltern als Lernmodelle fehlen. Sie müssen mit Hilfe ihrer Tentakel alles selbst lernen. Dabei verlassen sie sich auf ihr dezentrales Gehirn in den Tentakeln, mittels derer sie in Kommunikation mit ihrer Mitwelt treten« (Graeff 2022; Ergänzung K).

torischen Aspekten diese Kommunikation in ganz besonderer Weise prägen (von Böhm/Steen 2023; Uexküll 1958).

Vielfarbig und mehr oder weniger glatt oder stark strukturiert spiegelt die Haut der Oktopoden ihre Stimmung wider. In der Interaktion werden insbesondere die Veränderungen von Farbe und Struktur der Oktopushaut vom menschlichen Gegenüber zugleich auch als kommunikative Handlungszüge verstanden (Leon 2022: 504–505). Dabei steht den Tieren ein großes Farbspektrum zur Verfügung: »[Forming part of the cephalopod skin] the pigment-containing chromatophore sacs [...] can contain yellow, red, or brown/black pigimentary color« (Mather 2004: 193, Ergänzung KJ). Die Farbe Weiß, auch in der Variation von ›pale‹, zeigt, dass die beobachteten Oktopoden entspannt sind. »Was für ein inniges Einvernehmen herrscht da zwischen uns«, so interpretiert Sy Montgomery (2022: 192) ihr Wohlbefinden und vielleicht auch das ihrer Oktopuspartnerin. Tasten und Riechen sowie Bewegung⁹ sind hier weit verbreitete Mittel der Interaktion (siehe Kapitel 2): »[R]eaching one or two arms toward [the human]« (Anderson et al. 2010: 263; Ergänzung KJ). »[The arms] have not only the sense of touch, but also the capacity to sense chemicals – to smell, or taste« (Godfrey-Smith 2018: 67, Ergänzung KJ).

Dass Oktopoden ›good learner[s]‹ sind, belegt einerseits ihre *Agency* in der Form des ›learnership‹, andererseits weist diese Fähigkeit sie als »justifiably known as the invertebrates with intelligence«¹⁰ (Mather 2004: 209) aus. Kommunikation mittels Farben ist zweifelsohne ein Gebiet, auf dem der Oktopus dem Menschen überlegen ist¹¹. Er ist fähig, »[a] ritualized communication« (Mather 2004: 207) mittels einer auffälligen Hautfärbung beispielsweise als Warnung zu zeigen. »[Selbst wenn es] sich bei den verschiedenen Displays der

9 »An aggressive move« (Godfrey-Smith 2017: 185).

10 Dazu gibt es unterschiedliche Interpretationen: »Das Hirn des Oktopus ist vergleichsweise klein. Die meisten Sinneszellen liegen in seinen unabhängig voneinander operierenden Armen« (Meijer 2018: 113). »Oktopoden sind gallertartige, wabernde und wirbellose Kopffüßer. Sie besitzen acht Tentakel mit Saugnäpfen. Jeder Tentakel ist in der Lage, eigenständig zu denken und zu fühlen. Das Gehirn eines Oktopus befindet sich nicht zentral im Kopf, sondern im ganzen Körper verteilt. Man könnte sagen, der Körper, in dem übrigens drei Herzen schlagen, ist das Gehirn« (Graeff 2022).

11 Obwohl sich auch unsere eigene ›animality‹ (Massumi 2014: 3; Derrida 2006) in Farbwechseln unserer Haut äußert: »blass vor Angst« oder »rot vor Scham« sind dafür Beispiele. Traditionelle Bemalungen indigener Völker und moderne Tätowierungen sind ebenfalls zu nennen, auch wenn sie nicht in gleicher Weise wandelbar sind. Ich danke Moritz Wiegand für diesen Hinweis.

Akteur*innen in semiotischer Hinsicht um Symptome (z. B. für Angst, Stress) [...] handel[t]« (Böhm/Steen 2023: 1; Ergänzungen KJ), kann der Farbwechsel trotzdem auch als Medium der Interaktion zwischen Arten (›interspecies communication«; vgl. Montgomerys Erfahrung mit Athena bei Leon 2022: 504f.; siehe auch Abschnitt 5) verstanden werden. Diese Interpretation liegt nahe, wenn die*der Interaktionspartner*in ihre*seine Zuwendung beispielsweise unterbricht. Evolutionär sind die lebhaften Farb- und Strukturmuster dieser Tiere als Verteidigung gegenüber ihren Fressfeinden entwickelt worden. So können sich Oktopoden in Form, Farbe und Struktur (›saddle spot«, ›reduced-area saddle« Mather 2004: 205) an ihre Umgebung in einem Maße anpassen, »that so nicely help the camouflaged cephalopod to disappear before vertebrate eyes« (Mather 2004: 208), nämlich den Augen ihrer Feinde, darunter auch den Menschen, die Oktopoden fangen, um sie selbst zu verspeisen oder zu verkaufen.

Die Kreativität der Oktopoden ist auch sichtbar in den Techniken des Versteckens in selbst zusammengetragenen Muschel- oder Steinhäufen (min. 00:11:54–00:11:58) oder mittels drapierter Blätter, die im *Great African Seaforest* noch allgegenwärtig sind. Das in seiner Evolution schon lange nicht mehr durch Muschelschalen geschützte Tier kombiniert vielfältige Praktiken, um sein Leben zu schützen, und versteht es auch, seine Signale kommunikativ bedeutsam zu graduieren (›variations in signal form are a common modulation of signal intensity« Mather 2004: 205). Es sind diese beiden Funktionen, Camouflage und Signalisieren, die zur Entstehung und dem Gebrauch der Farbwechsel beigetragen haben (Godfrey-Smith 2017: 126). Beide sind zugleich auch auf die Beziehungsebene mit nicht der gleichen Art zugehörnden Interaktionspartner*innen gerichtet unabhängig davon, ob sie tierlich oder menschlich sind.

Wie aber steht es mit der Kommunikation innerhalb dieser Tierart? Tatsächlich leben die Oktopoden einsam und treffen sich nur, wenn sie geschlechtsreif sind. Im Wettstreit um ein zur Paarung bereitendes Oktopus-Weibchen kann ein besonders ausdrucksstarkes Muster beobachtet werden »[agonistic zebra] produced for the cephalopod visual system as receiver« (Mather 2004: 206). »Zebra is produced only by mature squid, but both by male and female [...]. [Z]ebra appears in situations [...] of challenge, such as males competing for the consortship of a mature female« (Mather 2004: 206). Diese Praxis des ›fellowship« oder ›partnership« unterstreicht die Agency des Oktopus-Weibchens einerseits und den Prozess des Suchens und Findens (›seeking out«) der Oktopus-Männchen andererseits.

Die Gegenüberstellung der Perzeption der gewählten Formen bei Kommunikation mit Artgenossen im Unterschied zur Kommunikation mit Fremden in ihrem Habitat (Table 11.1 Mather 2004: 204) zeigt, dass das Profil (outline) einfach (clarity) ist unter Gleichen, aber komplex bei der artenübergreifenden Verwendung zur Täuschung, vorwiegend zum Verstecken vor Anderen. Auch die Farbmuster sind gegenüber den Verfolgern vielfältig, während sie innerhalb der Art gerade auch im Fall des Zebra-Musters sich auf einen einfachen Kontrast beschränken. Farbe spielt gegenüber Fremden eine Rolle, kaum aber in der Kommunikation mit ihresgleichen (Mather 2004: 204f). Die multimodale Kommunikation von Oktopoden stellt zweifelsohne für die menschlichen Partner*innen eine große Herausforderung dar. Es kann ihnen nur durch wiederholende Begegnungen gelingen, bestimmte auffällige Muster, bestimmte Farben und/oder Duftstoffe oder auch Bewegungsschemata als mit den jeweils spezifischen Handlungszügen verknüpft zu erkennen und ihre Erfahrung für weitere Treffen gedanklich vorzuhalten.

Vielleicht ist ja für den in der Regel solitär lebenden Oktopus diese zuletzt genannte Aufgabe gar nicht so unbekannt. Sein Aufwachen ist bereits durch die Interaktion mit anderen Arten bestimmt. Die Kommunikation mit seinesgleichen bleibt letztlich auf das Kräftemessen in der Brunft mit männlichen Rivalen und den Akt der Übergabe der Spermazellen an das Weibchen beschränkt, wobei über den Zeitpunkt ihrer Nutzung das Oktopus-Weibchen allein bestimmt, ihr also eine besondere und für die Arterhaltung bedeutsame *Agency* vorbehalten bleibt. Die mit unterschiedlichen Sinnen jenseits des Hörens verknüpften Kommunikationsrepertoires weisen deutlich über den sprachlichen Kontext hinaus. Sie repräsentieren Handlungszüge jenseits der Sprache, wie es unser Buchtitel *Sprache entgrenzen* ankündigt.

4 Mediale Perspektiven liquiden Sprechens

Ich habe mich entschieden, das Bild der »[f]ließende[n] Übergänge, verschwimmende[n] Körper, [...] [und der] verflüssigten Sprache« (Dreckmann/Meis 2022; Ergänzungen KJ), letztere mit Bezug zu Bauman/Palese (2013) in der Kollokation ›liquide Sprache‹ zu nutzen, um zunächst die Ungleichheit des menschlichen Körpers, –selbst beim Tauchen im Meer–, gegenüber dem Körper des Oktopus zu adressieren. Diese Vorstellung des Liquiden möchte ich aber dann auch auf die hinsichtlich ihrer zugrundeliegenden Mediali-

tät unterschiedlichen Daten anwenden, um sie einerseits als vergleichbar, andererseits als sich komplementär ergänzend beschreiben zu können.

Ich bleibe dabei, den Film voranzustellen, obwohl er in seiner Historizität und seiner spezifischen audiovisuellen Performativität und technischen Bedingtheit eine medial jüngere Form repräsentiert als der literarisch gestaltete und als Sachbuch publizierte Text, der seinerseits sich von den naturwissenschaftlichen Aufsätzen unterscheidet, welche den jeweiligen mit ihrer akademischen Disziplin verknüpften (naturwissenschaftlichen) Normen folgen.

Was den Film angeht, so wird die Illusion der Filmzuschauer*innen in der erzählten Story durch eine Ausblendung der tatsächlichen Produktionsumstände erzeugt:

Was im Film nach dem zarten Beginn von ungestörter Zweisamkeit aussieht, wird [...] hochprofessionell mitgefilmt. Nach rund sieben Jahren, in denen Foster fast täglich im Kelpwald getaucht ist, fängt die Regisseurin und Kamerafrau Pippa Ehrlich an, ihn auf seinen Ausflügen in der *False Bay* zu begleiten. Auch sie taucht die meiste Zeit frei, ohne Flasche, ohne Neopren. Nur ist sie nie im Bild. In dem Dokumentarfilm wird das Entstehen im Team und als professionelle Produktion nicht thematisiert. Foster erwähnt weder Ehrlich noch das restliche Filmteam, obwohl er seine täglichen Tauchgänge scheinbar minutiös dokumentiert und die einzelnen Tage mit Inserts durchzählt. (Lüdemann 2020)

Der Dokumentarfilm schöpft seine technischen Möglichkeiten aus, indem er in Art einer Collage durch Ton- und Bildschnitte die Fiktion einer zwischen zwei Individuen sich entfaltenden Mensch-Tier-Interaktion kreiert. Der Film ordnet Aufnahmen verschiedener Oktopoden gekonnt so an, dass es den Zuschauer*innen so vorkommt, als würde immer ein- und dasselbe Oktopus-Weibchen dem Taucher begegnen. Für die Entwicklung der Filmgeschichte im liquiden Raum sind neben den Aufnahmen unter Wasser auch die wiederholend eingeschobenen »kitchen scenes« bedeutsam, die der Regisseur James Reed angeregt hat. Foster als Erzähler kommentiert darin seine Rolle. Er spricht frontal zur Kamera. Die*der Zuschauer*in wird direkt angesprochen als Dialogpartner*in. Diese Inszenierung seiner individuellen Reflexion über seine Beziehung mit dem Oktopus-Weibchen (»Ich bin lieber so eine Art amphibisches Lebewesen« min. 00:09:17-00:09:20) steht im Kontrast zur interaktiven Unmittelbarkeit der unter Wasser gefilmten Begegnungen mit dem Oktopus-Weibchen, bei der der*dem Zuschauer*in lediglich die Rolle

der*des Betrachterin*s zukommt. »Seine zweitausend Saugnäpfe setzt er [der Krake] unabhängig voneinander ein, als hätten wir Menschen zweitausend Finger« (min. 00:34:50-00:34:55; Ergänzung KJ). »Solche Begegnungen regen die Intelligenz dieser Tiere an und vielleicht bereitete es ihm auch eine gewisse Freude« (min. 00:31:08-00:31:10). »Irgendwie begriff dieser Oktopus, dass dieses Ding nicht gefährlich ist, und trat in Interaktion mit diesem Menschen« (min. 00:31:11-00:31:13).

Wie der deutsche Titel ›Mein Lehrer. Der Krake‹ ausdrückt, begreift er sein Verhältnis hierarchisch als ihr untergeordnet und reklamiert für sich die Rolle des Schülers, der ihrer im deutschen Filmtitel vorangestellten Rolle komplementär ist (»Dieser Oktopus lehrte mich, dass ich kein Besucher bin, sondern ein Teil dieser Welt« min. 01:22:40-01:22:43). Während der englische Originaltitel keine Genuszuweisung vornimmt, ›My Octopus Teacher‹, weist der deutsche Titel eher auf einen Mann hin: ›Mein Lehrer. Der Krake‹, obwohl die Paarbeziehung im Film auch entlang einer traditionellen Rollenteilung interpretiert werden kann (»Ich habe mich in dieses Lebewesen verliebt, aber auch in die wilde Natur, die es verkörpert und die mich gewandelt hat« min. 01:22:16-01:22:18; »Ich dachte nur noch an dieses Tier im Wasser oder an Land« min. 00:33:05-00:33:12; »Es wurde fast zu einer Art Besessenheit. Ich wollte es jeden Tag besuchen« min. 00:33:14-00:33:26). Das Oktopus-Weibchen ist die Hauptperson (»she« min. 00:29:35; 00:34:03; »see her« 00:46:01). Der Taucher ist als therapiebedürftiger Patient von ihr abhängig, was im Kontrast zu seiner Handlungsmacht, seiner *Agency* steht, denn es ist an ihm, die Begegnungen mit ihr zu suchen, indem er sie regelmäßig an ihrem Ort trifft. Die den Zuschauer*innen vermittelte hierarchisch geordnete Tier-Mensch-Beziehung, die im Titel des Films explizit und hochrangig gesetzt ist, wird mehrfach gebrochen. An einer Stelle überlegt der Taucher, die Haie, die das Oktopus-Weibchen bedrohen, zu verscheuchen (min. 00:30:44-00:30:50). Er fühlt sich als bestimmt, sie, die hier als schwach gezeigt wird, zu verteidigen. Nicht als ein hierarchisch unterlegener Schüler, sondern als ihr handlungsmächtiger Beschützer imaginiert er sich. Auch die eingestreuten Szenen in der Küche, die die Geschichte vorantreiben, stellen diese Hierarchie in Frage. »In gewisser Weise spiegelten sich unsere Leben« (Octopus 2020 min. 00:47:00-00:47:06). Aus der Beobachtung, dass dem Oktopus-Weibchen ihr Arm nachwächst, leitet der Taucher ab, dass er auch selbst geheilt werden und seine Krise überwinden wird. Die Erfahrung des Tauchers, dass seine »Beziehung zu den Wäldern des Meeres und ihren Geschöpfen [...] mit jeder Woche, jedem Monat, jedem Jahr [tiefer wird]« (min. 01:21:37-01:21:48; Ergän-

zung KJ), führt dazu, dass Foster sich nicht nur »in Verbindung mit diesem Ort« (min. 01:21:57-01:21:58) fühlt, sondern es ihm so vorkommt, als ob »er [der Ort] mit einem auf sichtbare Weise [spricht]« (min. 01:21:59-01:21:60). Der Gebrauch von Sprache wird demnach nicht nur für Tiere, sondern sogar für natürliche Räume reklamiert. Bedeutsamer aber erscheint mir die Verortung der Interakteur*innen im Sinne Bühlers in der Origo (Bühler 1934), die dreifach durch den Ort und die Zeit sowie die im kommunikativen Austausch stehenden Akteure bestimmt ist. Ihre zweiseitige Bedingtheit (›zwiefältig‹ Buber 1923) und wechselseitige Orientierung im Raum determiniert ihre zur Kommunikationsdyade erweiterte Beziehung (›Gesprächdyade‹ Jungbluth 2005: 59–75). Mit dem dyadischen Modell lässt sich die Auswahl materiell distinktiver kommunikativer Mittel entlang der drei Anordnungen der Akteure einander gegenüber (›face-to-face‹) im Unterschied zu einer solchen nebeneinander (›side-by-side‹) und einer sehr viel seltener zu beobachtenden Anordnung hintereinander (›face-to-back‹) erfassen.

5 Fiktion im Film: Lebenslang ein Paar

Tatsächlich folgen die Interaktionen einerseits zwischen Taucher und ›Tintenfischfrau‹ (Lüdemann 2020) in freier Natur am südafrikanischen Kap (Ehrlich/Reed, ZA 2020) und andererseits zwischen Besucherin und ebenfalls einem ›Oktopus-Weibchen‹ (Montgomery 2022: 71) namens Athena im künstlich gestalteten New England Aquarium in New Hampshire bekannten Phasen zwischenmenschlicher Begegnungen.

[Der psychisch angeschlagene Taucher] mit Maske, Schnorchel und Flossen [...] findet heraus, wo sie wohnt, wartet auf sie vor ihrer Felsenwohnung. Und so treffen sie sich öfter. Anfangs nur zaghaft beginnt auch die Tintenfischfrau, den *Freediver* zu inspizieren, [...] [was sich darin äußert, dass] die Oktopusfrau eine[...] ihrer Tentakel[n] bis zu Fosters Fingerspitze ausstreckt, seine Hand berührt und Saugnapf für Saugnapf abtastet. (Lüdemann 2020; Ergänzungen KJ)

Die Kontaktaufnahme zwischen Mensch und Tier beginnt mit einem reziproken Anschauen: »[The octopus] starred me on in a little gap« (min. 00:12:05-00:12:11; Ergänzung KJ) und in umgekehrter Richtung: »Dann starrte ich nur noch in die Augen dieser unglaublichen Kreatur« (min. 00:30:16-00:30:34). »Er

beobachtete mich. Ich beobachtete ihn« (Montgomery 2022: 12), wodurch sich eine beidseitig ausgeübte *Agency* des ›observership‹ bestätigt. Die menschlich-tierliche visuelle Kontaktaufnahme erfolgt aus sicherer Distanz, wie es auch in Mensch-zu-Mensch Begegnungen bei einem ersten Kontakt und beispielsweise zur Vorbereitung einer verbalen Instruktion (vgl. ›fokussierte Interaktion‹ als Teil einer multimodale[n] Kommunikation; Müller/Bohle 2007) geschieht. Oktopoden können Menschen als Individuen wiedererkennen. Der Blickkontakt steht an erster Stelle.

That octopuses can recognize individual humans¹² is one more reminder that scientists and keepers form what Davis and Balfour (1992) described as ›the inevitable bond‹ (p. 3) with their animals, that there is more going on in these interactions than just provision of food presentation or of stimuli for testing. They described *the bonding process as a relationship* between the observer (a person) and the observed (animal), although they note *the relationship is two-way*. Our results are a reminder that such a relationship may form between pairs as phylogenetically distant as humans and octopuses. (Anderson et al. 2010: 270)

Wenn eine wechselseitige Aufmerksamkeit beispielsweise durch einen Blickkontakt geschieht, kann diese tierlich-menschliche Beziehung: möglicherweise weiter entfaltet werden. Prozesshaft kann die momentane Beziehung in Form eines Beobachten|›observership‹ zu einer Paarbeziehung|›relationship‹ weitergeführt werden, die auch den Zwischenschritt einer fürsorglichen Beziehung|›careship‹¹³ einschließen kann. Schließlich können die beteiligten Individuen sich auch als ein besonderes Paar begreifen, das zunehmend

12 Die Individualität ist auf beiden Seiten bedeutsam: »[...] one message of octopus experiments is that there is a great deal of individual variability. [...] Charles was probably an octopus with a particularly feisty temperament« (Godfrey-Smith 2018: 54). »[J]ets of water through the funnel« [may be observed], when octopuses are ›irritated‹ (Anderson et al. 2010: 263, Ergänzung KJ). Keinesfalls verhalten sich alle Vertreter*innen einer Art gleich. »Octopuses [...] have learned to turn off the lights by squirting jets of water at the bulbs when no one is watching« (Godfrey-Smith 2018: 55). »Mit seinen Wasserspielchen [im Coburger Sea-Star-Aquarium] sucht [Oktopus] Otto Aufmerksamkeit. Beschäftigt sich niemand mit ihm, räumt er schon mal sein ganzes Becken um oder stößt kleine Einsiedlerkrebse von der Spitze seiner Pyramide« (Kummer 2010, Ergänzungen KJ). Dieses Verhalten wurde unter in Aquarien gefangenen gehaltenen Oktopusen in Deutschland beobachtet, aber auch in Neuseeland (Godfrey-Smith 2018: 215f.).

13 Zu ›careship‹ in Westafrika, besonders Ghana, vgl. Chachu/Amfo (im Druck).

wechselseitige Verantwortung für ihr*sein Gegenüber übernimmt und diese Begegnung auch über eine mehr oder weniger ausgedehnte Zeitspanne verstetigt. Wechselseitige Erwartungen und die frequenten Begegnungen selbst sind ein Hinweis auf soziale Kontexte, die in tierlichen und menschlichen Gemeinschaften innerhalb einer Art entwickelt und poetisch auch als Liebesbeziehung |>pairship</>|>bonding</> bezeichnet werden können.

Die artenübergreifende Beziehung jedoch ist ungewöhnlich. Im Folgenden zeichne ich ihre schrittweise Intensivierung nach. Wie in jeder anderen Beziehung zwischen Lebewesen der gleichen Art können die in einer Mensch-Tier-Beziehung verbundenen Individuen gleichfalls ihr reziprokes »observership« (>the two-way relationship<) in Form eines zweiten Modus durch Berührungen stärken:

When you encounter¹⁴ and approach an octopus in the wild and pause in front of it, in at least some species the octopus sends out one arm to inspect you. Often a second arm follows. (Godfrey-Smith 2018: 67)

Eine solche grüßende Interaktion kann nicht nur wie zwischen Menschen mittels Worten, sondern wie ein Handschlag oder ein flüchtiges Küssen durch Berührung eingeleitet werden. Kurz, es handelt sich um eine die Interaktion eröffnende Grußhandlung, um »intendierte kommunikative Signale (z. B. als Drohung, Begrüßung)« (Böhm/Steen 2023: 153), die paarförmig durch mindestens zwei sich wechselseitig bedingende gestische oder sprachliche Handlungen repräsentiert wird¹⁵. Der Gruß kann auch durch eine Kombination sprachlicher und non-verbaler Mittel kommuniziert werden (Erikson 2007). Der in allen Quellen betonte dialogische Charakter spiegelt sich auch im »receptive[n]

14 >Encounter< wird hier im Sinne von tatsächlich den Anderen anschauen (»seeing«) verwendet mit dem möglichen Ergebnis »of being surprised and shocked into a new way of thinking« (Böhm 2022: 147). Vgl. zu dieser empathischen Interpretation von »encounter« Derrida (2006) und Haraway (2008) sowie Midgley (1984). Buber (1923) kreiert das Wort »Vergegnung« als Gegensatz zu Begegnung, womit er den markierten Fall einer ausbleibenden Verbindung zwischen Menschen trotz des physischen Zusammentreffens bezeichnet.

15 Vgl. »speech acts« Austin 1962; Searle 1969: ein Sprechakt ist eine sozial relevante Handlung. Über die wechselseitigen Erwartungen zwischen Mensch-Tier-Paaren berichten nicht nur naturwissenschaftlich fundierte Quellen wie die zitierte, sondern auch das erzählende Sachbuch und der Film.

[.] [V]erhalten« (Coulmas 1977: 70). Coulmas betont außerdem die Notwendigkeit, das zwischen den (menschlichen) Interaktionspartner*innen praktizierte Aushandeln von »Bedeutung als Prozess zu konzeptualisieren«¹⁶:

Als Sy Montgomery einige Wochen später Athena, [ihre Oktopuspartnerin] füttern darf, erkennt diese sie wieder und lässt sich nach der Mahlzeit an Kopf und Körper von ihr streicheln. Montgomery staunt, wie weich sich die Haut des Tiers anfühlt. Sie spürt, dass Athena sich über das Essen und die freundschaftliche Berührung freut, und sie ist stolz auf die Verbindung [bond], die sich zwischen ihnen entwickelt hat. (Leon 2022: 504f., Ergänzung K)

In der Wortwahl ›Essen‹ spiegelt sich die vermenschlichte Perspektive der Berichterstatte¹⁷, die auf eine gesteigerte Nähe gegenüber dem Oktopus-Weibchen und eine damit verknüpfte größere emotionale Bindung zu ihr hinweist¹⁸:

Im Nu sind meine beiden Hände und Unterarme umschlungen von Dutzenden weicher, mich abtastender Saugnäpfe. (Montgomery 2022: 17)

Schließlich entsteht wechselseitiges Vertrauen nicht nur durch taktile Kommunikation, sondern auch aus räumlicher Kontingenz: »Du darfst in meine Oktopushöhle kommen« (min. 00:21:22-00:21:30). So stellt sich Foster die Einladung seiner Oktopus-Interaktionspartnerin vor. Mit dem Ziel, das in den Bildern Sichtbare für das Publikum auch emotional erfahrbar zu machen, spricht der Taucher *als* Tier:

Im Vergleich mit der sonst üblichen Konstruktion tierlicher Emotionen in einem Sprechen *über* Tiere, können durch die mediale Praktik des Sprechens *als* und *für* Tiere [.] die Zuschauer*innen stärker emotional in die dargestellte Handlung involviert werden. (Steen 2023: 342)

16 Zu artinternen Begrüßungsritualen zwischen Tieren u.a. Vögeln und Säugetieren: Meijer 2018: 50ff; zu Graugänsen im Besonderen: Kleindorfer/Mc-Allister-Käfer 2024; zu Language Use as Doing: »Linguistics as a performative science«: Rajagopalan 2013.

17 Der Film über die menschlich-tierliche Interaktion im natürlichen Umfeld zeigt keine Fütterung im Unterschied zur Begegnung im Ozeaneum, in dem diese Fürsorge (careship« siehe oben) durch den Menschen unabdingbar ist.

18 Mit Steen (2023: 342) verstehe ich diesen Gebrauch als Mittel einer emotionalen Steigerung, die sich an ihre Leser*innen richtet.

Indem Foster dem Oktopus-Weibchen seine Stimme leiht (Armbruster 2015), wodurch auch er sein Gegenüber vermenschlicht, ermöglicht er den Zuschauer*innen die emotionale Erfahrung ihrer Rolle (›spectatorship‹ Greifenstein/Schmitt 2014; ›poiesis of film-viewing‹ Kappelhoff/Wedel 2015; siehe Abschnitt zwei). Die scheinbar passive Rolle des ›spectatorship‹ wird zu einer körperlichen Erfahrung, indem sie um »embodied, attentional, and affective dimensions« erweitert wird. Die medial erzeugte emotionale Nähe verdichtet sich, wozu die Kamerawinkel, Einstellungsgröße und Mise-en-scène, insbesondere auch ›staging‹ beitragen¹⁹. So erleben die Zuschauer*innen buchstäblich am eigenen Körper, wie der Taucher mit dem Oktopus-Weibchen einen zunehmend intimeren Umgang entwickelt: »Eine derartige Verbindung zu so einem Lebewesen zu erleben, ist absolut atemberaubend« (Octopus 2020 min. 00:32:00-00:32:11).

Ohne dass es im Film thematisiert wird, sind die Begegnungen zwar stark limitiert auf die Zeitspanne, die ein Apnoetauchen erlaubt, was sich auch bei erfahrenen Taucher*innen auf etwa zehn bis fünfzehn Minuten begrenzt, aber die regelmäßige raumzeitliche Kopräsenz erlaubt dem tierlich-menschlichen Paar intensive Momente des Kennenlernens. Tatsächlich findet die im Film fiktiv als eine mehr als ein Jahr andauernde Tier-Mensch-Begegnung im Fall von Foster ihren Höhepunkt in der Umarmung seines nackten Oberkörpers durch das Oktopus-Weibchen mit allen ihren acht Armen (min. 00:32:26-00:32:46). Ich stimme Graeff (2022) zu, wenn sie diese Berührungen als ein »kommunikatives Spiel« bezeichnet, das zum Prozess der Heilung Fosters beiträgt:

Seine Haut hat direkten Kontakt mit der Haut des Oktopus. Auch hier sind die Tentakel[n] bedeutsam, die im Verlauf des Films mit den menschlichen Fingern, Händen, Armen und Beinen interagieren. Das kommunikative Spiel unter Wasser verändert Foster auch mental. (Graeff 2022; Ergänzung K)

Der Taucher ahnt nicht, dass diese Begegnung die letzte zwischen ihnen sein wird. Auch Montgomery erlebt ihre letzte Begegnung als sehr emotional, denn »nur eine Woche später wird ihr per E-Mail mitgeteilt, dass Athena gestorben ist – eine Nachricht, bei der sie in Tränen ausbricht« (Nachwort von Donna Leon in Montgomery 2022: 505). Sowohl der Film als auch der literarische Text erfüllen die Erwartungen ihrer Rezipient*innen an ein kulturhistorisch geläufiges Narrativ, das den Bogen von der ersten Begegnung bis zum Höhepunkt

19 Ich danke Dorothea Horst für diese Hinweise.

der Interaktionsgeschichte dramatisch steigert: Der Oktopus wird gejagt (»das macht sie [die Pyjamahaie] zu todbringenden Oktopusjägern« min. 00:17:15-00:17:20), er wird verletzt und verschwindet dann eine Zeit lang, was auf den von den Begegnungen mit ihr abhängigen Taucher dramatisch wirkt, und zuletzt stirbt das Oktopus-Weibchen²⁰. Indem die dramatische Spannung die mit den Rollen des ›readership‹ und des ›spectatorship‹ verknüpften ästhetischen Erwartungen einlöst, weist das Narrativ als fiktionales Erzählen sowohl im literarischen Text als auch im Film zugleich auf die für beide Prozesse geltenden kulturellen Praktiken hin. In dieser Hinsicht konvergieren Buch- und Filmszenierung und ihr Narrativ der Mensch-Oktopus Begegnung im Sinne einer anthropomorphisierten Interaktionsperspektive.

6 Tierlich-menschliche Neugier im liquiden Raum

Was haben die Beziehungsgeschichten zwischen Menschen mit solchen zwischen Menschen und Tieren gemeinsam? Nicht nur Bühler (1934) verweist auf die Bedeutung der Beziehungsebene zwischen den Akteur*innen, die einer kommunikativen Interaktion vorläufig ist und sie bestenfalls zu fundieren vermag. Wenn eine*r ursprünglich Fremde*r ein*e Kommunikationspartner*in werden soll, muss ihr eine eigene *Agency*, eine eigene Handlungsmacht²¹ zuerkannt werden (siehe Abschnitt 2). Die Interaktion selbst besteht nach meinem Verständnis, wie es in vielen Modellen zur Kommunikation auch ausgedrückt wird, nicht allein in der Produktion von in der Regel multimodalen Zeichen auf der Seite desjenigen, der sich ausdrückt (Sender*in; Sprecher*in) und auf ihrer Interpretation auf Seiten des oder der Interaktionspartner*innen (Empfänger*in; Hörer*in). Alterität drückt eine reziproke Beziehung zwischen den Subjekten aus (vgl. das Reflexivpronomen in: *sich* begrüßen, *sich* unterhalten)²², die für Sprache zentral gesetzt wird (Schlieben-Lange 1998: 44).

20 Ich danke Rita Vallentin und Dorothea Horst für ihre Kommentare und unsere Diskussion zur dargestellten Dramatik.

21 Für das Octopus-Weibchen kann hier exemplarisch auf die Interpretation des Tauchers hingewiesen werden: »Der Hai wurde komplett überlistet« (Octopus 2020 min. 01:06:54-01:07:00).

22 Vgl. ›two way relationship‹ im Oktopus-Mensch Paar (Anderson et al. 2010). Wenn die Beziehung emotional mit engeren Erwartungen für zukünftige Begegnungen verknüpft wird: ›bonding‹ (Davis/Balfour 1992; Rault et al. 2020).

Darunter fällt auch der Rollenwechsel im Gespräch. Die Vielfalt der Handlungen, die eine Begrüßung auszudrücken vermögen, machen es möglich, dass sie nicht nur sprachlich, nicht nur lautlich, sondern auch gestisch, beispielsweise durch Blicke verkörpert werden kann. Die traditionell angenommene Grenze zwischen Menschen und Tieren stellt also keine unüberwindbare kommunikative Grenze dar und bestätigt die Bemühungen um eine Betrachtung ›Beyond the Human-Animal Divide‹²³ (Oliver 2017). Daher reklamieren meines Erachtens zurecht Böhm und Steen (2023) Alterität als bedeutsam auch für tierlich-menschliche Paarbeziehungen. In solchen Kontexten ist die Anerkennung der *Agency* des Gegenübers in besonderer Weise an die kreative Lösung des Miteinander geknüpft. »Wir dürfen nicht vergessen, dass *das Aufeinander-Eingestelltsein* ein wechselseitiges Verhältnis ist: Tiere beeinflussen uns in unserem Zusammenleben mit ihnen genauso wie wir sie« (Meijer 2018: 39, Hervorhebung KJ).

Wenn die Handlungsmotivation zur Interaktion zwischen Tier und Mensch regelmäßig mit Neugier verknüpft wird, wie es auch in dem diesem Beitrag vorangestellten Datenausschnitt der westafrikanischen Sprache Ewe (Yakpo im Druck) der Fall ist und im Film das Oktopus-Weibchen in ihrer Rolle als Lehrerin auszeichnet (»[My octopus teacher was] very curious, very interested, very curious«, min. 00:18:12-00:18:20; Ergänzung KJ), so wird damit die mit Kreativität verknüpfte Offenheit und zugleich die Beziehungsebene unmittelbar angesprochen. »[A]uch scheint sie [die Tintenfischfrau] genauso neugierig auf mich zu sein« (Montgomery 2022: 20; Ergänzungen KJ). Oktopoden sind Einzelgänger und wachsen auch nicht mit ihresgleichen auf. »Oktopoden haben keine Vorbilder, sie müssen sich alles selbst beibringen [...]« (Graeff 2022). Sie lernen beim Beobachten²⁴ anderer Tiere in ihrer Habitat (Fiorito/Scotto 1992), was sich auch auf die Interaktion mit Taucher*innen und Besucher*innen ausweiten lässt. Spielen, auch das artenübergreifende

23 Vgl. auch »Zwischen-Leiblichkeit« als Teil tierlicher *Agency* (Balgar 2015) und als Brücke zwischen Subjekten, die sich durch *Agency* auszeichnen, ohne dass eine Abgrenzung zwischen Tier und Mensch vorgenommen wird. Die fehlende Abgrenzung unterstreicht den liquiden Charakter der Interaktion.

24 »Octopuses are highly visual, especially when it comes to navigation and learning. They have lateralized vision, and are able to use a single eye for perceptual and learning tasks. Signals received via one eye are transmitted and processed in its ipsilateral optic lobe, which sends this information further upstream for ›cross-brain transfer‹ (Mather 2021: 408)« (Carls-Diamante 2022: 3).

gemeinsame Spielen²⁵ (»play behavior« Mather 2004: 202) löst einen über die Zeit verlaufenden Lernprozess nicht nur bei Haustieren aus »comprising last interactions, present, and [even] predicting future ones« (Rault et al. 2020: 2). Der naturwissenschaftliche Blick stellt die Sicht, dass Tiere »ausschließlich im Hier und Jetzt« (Meijer 2018: 43) leben, in Frage. Als autopoietisches System ausgelöst durch sich wiederholende Begegnungen mit bestimmten Menschen kann auch das Tier eine Erwartungshaltung ausbilden. »The time at which a relationship is formed remains difficult to determine, but it can be defined as the time at which the animal forms expectations of its interaction with humans« (Rault et al. 2020: 2).

Der mit solchen Erwartungen verknüpfte Prozess des Oktopus wird biologisch²⁶ (Grasso 2014; Godfrey-Smith 2016: 65–69; auch auf molekularer Ebene: Giesen et al. 2020)²⁷ mit Energieflüssen in einem neuronalen, auf chemischen Rezeptoren aufbauenden System (vgl. »embodied mind« Gallese/Lakoff 2005) in Verbindung gebracht. Seine Systemhaftigkeit und die messbaren, »lebendige[n] Knoten mobiler animierter Verbindungen« (Haraway 2018: 257) sind neuronalen Ursprungs und liegen auch unseren Gedächtnisleistungen zugrunde. Beide Ansätze unterscheiden sich hinsichtlich ihrer Dynamik und der wechselseitigen Bedingtheit. Die solipsistisch, lediglich das Individuum betrachtende und damit auch statische Auffassung²⁸ der im Gehirn gespeicherten Informationen, wie sie von Gallese/Lakoff (2005)

25 »Spiel«, konstatiert der kanadische Philosoph Massumi (2014) [..], »ist Ausdruck unmittelbarer Kreativität« (Meijer 2018: 133; zu »animal play«: Burghardt 2005).

26 »Our results demonstrate general principles by which single proteins and cells provide immensely flexible signal processing and highlight the importance in considering each functional component of a neural system toward understanding information coding and behavior« (Giesen et al. 2020: 603). »[W]e establish C(hemotactile) R(eceptor)s [CRs] as one distinguishing molecular feature [e.g. a protein] of this advanced and unique nervous system. CRs are capable of extensive signal filtering and coding, well suited to contribute to peripheral processing in the distributed, semi-autonomous nervous system of the octopus arm« (Giesen et al. (2020: 601; Ergänzung K)). »CRs provide diverse stimulus detection and transduction mechanisms, which could modify tonic spiking activity to transmit distinct neural signals in a chemical-dependent manner.« (Giesen et al. (2020: 602).

27 »Whole cell patch-clamp recordings revealed that cells with short, rounded dendritic endings (mechanoreceptors [like the suckers of cephalopod's arms]) could be defined by intrinsic mechanosensitive responses« (Giesen et al. 2020: 595; Ergänzung K).

28 Ich danke Dorothea Horst für diesen Hinweis.

vertreten wird, fassen Haraway und auch Foster zu Recht dynamisch, indem sie auf ihre Reziprozität im interaktiven Kontext verweisen (vgl. auch Zitate aus dem Film im folgenden Abschnitt 7). Kreativität im Sinne einer unangepassten und offenen Wahl zwischen möglichen Mitteln der Kommunikation und ihrem möglichst zwanglosen Gebrauch, der häufig auch mit Neugier und Spiel verknüpft ist, wird nicht erst bei der Schaffung der materiellen Mittel der Kommunikation und ihrem sequenziellen oder kombinierten Einsatz im Prozess der Verständigung relevant. Sie wirkt bereits auf die gemeinsame Gestaltung des Handlungsraums ein. Die Positionierung der Kommunikationspartner*innen zueinander und ihre Distanz ist auch für die Wahl der möglichen Kommunikationsmittel bedeutsam (Jungbluth 2018). Diese Leiblichkeit der Kommunikation (Weinrich 1988), die für alle Partner*innen gültig ist, auch für solche mit liquiden Körpern, stellt auch den sachlich gegebenen und unhintergehbaren Ausgangspunkt für den Aufbau und die möglichen Handlungszüge in der Beziehung zwischen den tierlich-menschlichen Kommunikationspartner*innen dar. Die Behauptung von Massumi (2014), Kreativität sei dem Menschen allein vorbehalten, ist aus dieser Perspektive nicht haltbar.

Wie aber wird Bedeutung geschaffen? Harms differenziert aus der Perspektive der ›naturalist theory of meaning‹ (Harms 2004: 40ff.) zwischen fünf Stadien der Emergenz von Bedeutung. Die erste Stufe ist die *Perzeption* einer Äußerung, die dem Empfänger allein vorbehalten ist. ›Natural meaning‹ entsteht beispielsweise auch durch olfaktorische, taktile oder visuelle Wahrnehmung (zu letzterer vgl. Albertazzi/Tonder/Vishwanath 2024). Auf der zweiten Stufe entsteht durch *Kooperation* zwischen dem Sender und dem Empfänger ein ›primitive content‹. Hier werden Konventionen zwischen den Interagierenden ausgehandelt, die Mehrdeutigkeiten ausschließen können, indem die Produktion und Interpretation der Zeichen, z. B. ›warning cries‹²⁹, Signale, die

29 Using the example of warning cries, Harms shows »that you can translate only signals between systems with very similar functions. This is why you can translate (well enough) between German and English, but not between vervet [monkey] and English« (Harms 2004: 40; Ergänzung KJ). Cheney and Seyfarth (1981) studied these monkeys who »employ a system of warning calls in which each of three types codes for the presence of a particular predator (snake, eagle, leopard). Animals hearing the call respond with behavior that is appropriate to the danger posed by the predator: hearing an eagle call, they descend from a tree; hearing a snake call, they stand and scan the ground; hearing a leopard call, they climb up the tree« (Sinha 2004: 217).

die Zu- oder Abwendung der Interagierenden und die Wege für die Weiterführung der Interaktion steuern, immer besser aufeinander abgestimmt werden³⁰. Im Prozess, Bedeutung zu schaffen, wird auf Stufe drei ›*Pure Indication*‹ situiert. Hier geht es um Futtersuche, ›rational choice‹ Entscheidungen (vgl. Esser 1991) und *Handlungsziele* wie ›belief‹ oder ›desire‹.

Auch hier finden wechselseitige Anpassungen und kooperative Abstimmungen zwischen Sender*in und Empfänger*in statt. Die Stufen vier und fünf der Bedeutungsentfaltung aber sind nach dieser Auffassung Menschen vorbehalten, da sie auf (abstrakten) sprachlichen Zeichen basieren: ›Symbolic reference/propositional contents‹ und ›hierarchical representation‹. Auf diese Inhalte und ihre Strukturierung wird auch mit dem Begriff der Semantizität referiert, der zweifelsohne dem menschlichen Sprachgebrauch vorbehalten ist.

Aber wie genau passiert es, dass aus bloßen Signalen eine Sprache wird? Harms stützt sich bei der Suche nach einer Antwort auf diese Frage auf »the teleofunctional theory of meaning as pioneered by Millikan (1984)« (Harms 2004: 40).

First, we must no longer assume that the world is the basic unit of meaning and that truth and falsity emerge only when simple referring symbols are combined to form complete representations. [...]. Second, meaning derives from historical patterns of successful coordination and comes in as many varieties [...] rather than being an abstract universal (i.e. the proposition) which somehow attaches to the signaling systems with sufficient complexity. [...] *Meaning, properly speaking, is fundamentally a cooperative phenomenon.* (Harms 2004: 41; Hervorhebung KJ)

Die Verknüpfung einer referentiellen Bedeutung (»What is the case« Millikan 2004: 18) mit einer Handlungsaufforderung (»What to do?« Millikan 2004: 18) sieht Millikan am Beispiel der Bienen in den sogenannten ›Pushmi-pullyu representations (PPRs)‹. Sie werden als einstellige, quasi ›in einem Atemzug‹ hervorgebrachte Äußerung gewertet. PPRs können insofern als ein weiterer Schritt auf dem Weg vom primitiven Zeichen zur Sprache betrachtet werden.

30 Daneben wird auch auf eine kompetitive Abstimmung zwischen beiden Interaktanten hingewiesen, die im Kontext sexueller Abläufe oder solcher des Futterneids sowie gegenüber Fressfeinden bedeutsam sind (Stage 2a ›Exploited perception‹). Zu Signalen gegenüber Verfolgern: Cheney/Seyfarth 1981.

Durch die Aufführung von als Tanz beschriebenen Bewegungsmustern, die bereits Frisch (1923) als ›Sprache‹ bezeichnete, geben die als Kundschafterinnen handelnden Bienen bei ihrer Rückkehr in den Bienenstock artintern Auskunft über Pollen- und Nektarsammelgebiete an die Sammlerinnen weiter. Millikan stützt damit die Sichtweise auf Sprache, auch die der Tiere, als einer für die Kooperation entwickelten sozialen Praxis (Jungbluth 2016; Reckwitz 2016). Mit Bezug auf Peirce adressiert Francescoli (2021) in ähnlicher Weise die interaktive Praxis des ›(s)emiotic scaffolding‹ als »[communicative] processes of sign action developing at many organizational levels [...] ›focusing energy flow and agency of the system [...] upon a constrained repertoire of possibilities‹ (Emmeche 2015)« (Francescoli 2021: 201; Ergänzung KJ).

Sinha (2004) focuses on the growth of higher-order communicative structures from primitive »signal« background seen in much of animal communication, but his view offers suggestions [...] about how the process of »elaboration« to higher-order »symbolic« structures occur. (Oller/Griebel 2004: 9)

Zwei Prozesse spielen dabei eine Rolle: Referenzierung und Interpretation (›construal‹). Erstere verlangt eine ungeteilte Aufmerksamkeit von Empfänger*in und Sender*in. Für die Interpretation oder Deutung verweist Sinha auf Langacker (2009 [1987]), dessen kognitive Grammatik ein Kontinuum zwischen Grammatik, Semantik und Lexikon skizziert und so diese zuvor getrennt betrachteten Teilbereiche der Linguistik integriert. Allerdings ist meines Erachtens der Schritt zur Analyse kommunikativer Handlungen zwischen Mensch und Tier entlang dieser an der menschlichen Sprachfähigkeit entwickelten Konzeption kein einfacher. Kognition wird hier lediglich als individuelle menschliche Gedächtnisleistung betrachtet. Im Unterschied dazu, hat beispielsweise der auch von Cowley und Fester-Seeger (2023) vertretene Ansatz, Kognition als Prozess einer Interaktion zwischen Akteur*innen und ihrer Umgebung zu begreifen³¹, das Potential (vgl. [cephalopod's] ›social cognition‹ Rault et al. 2020: 2; Ergänzung KJ), die kooperative Handlung nicht nur im Kontext der Mensch-Maschinen-Interaktion, sondern auch in der Mensch-Tier-Kommunikation abzubilden.

Diese Oppositionen, die bereits in der Systemtheorie von Maturana/Varela (1973/²1994) und im Anschluss daran auch bei Luhmann (1982) eine Rolle spielen, sind ebenfalls solche zwischen *máquinas* ›Maschinen‹ und *sistemas vivien-*

31 Ich danke Rita Vallentin für diesen Hinweis.

tes »lebendigen Systemen«. Jedes lebendige System repräsentiert zugleich *una máquina autopoietica* und als solche besitzt es Individualität³² im Unterschied zu anderen Systemen, die nicht autonom sind wie etwa Autos, die Beispiele für alopoietische Maschinen sind. Ihre Identität hängt von dem*der Beobachter*in ab, und sie reproduzieren sich nicht selbst. Individualität bleibt lebendigen Systemen vorbehalten, da sie *Agency* voraussetzt unabhängig davon, wie umfangreich ihr Radius ist. Die Zugehörigkeit der in der Tier-Mensch-Dyade versammelten Akteur*innen zur Kategorie der lebendigen Systeme ist unstrittig. Die Perspektive der Dyade in ihrer raumzeitlichen Dimension betont das Miteinander, das ich als eine erweiterte Entwicklung der Autopoiesis begreife, die Haraway (2018) »Sympoiesis« nennt. Im Prozess des aufeinander bezogenen Handelns (vgl. auch »Resonanz« Rosa 2016) lassen sich individuelle Aktionen und Reaktionen als Teile eines liquiden Ganzen identifizieren. So kann beispielsweise der bloß dialogische sequenzielle Austausch durch unterschiedliche Formen der Überlappung komplementiert werden, was nicht zum Abbruch der Kommunikation führen muss, sondern vielmehr häufig dichte, emotional herausragende Stellen im Austausch indexiert.

Welche Rolle spielen Erfahrung und das Gedächtnis? Evolutionär haben Oktopoden ihre Schalen verloren und gelernt, sich an ihr Habitat entsprechend anzupassen (vgl. »adaptiver Prozess« Grasso 2014: 95). Ihre auch im Film vielfach sichtbare Fähigkeit, neue Höhlen zu finden und für sich selbst durch Materialanhäufung ein Versteck kreativ zu bauen, wird ebenfalls als Ausweis ihres Gedächtnisses bewertet. In engem Zusammenhang damit wird auch die räumliche Orientierung (Mather 1991; Boal 2000) als neuronal gesteuert und gedächtnisrelevant bewertet:

Octopuses' capacities for memory are also highlighted in their use and occupancy of dens. Denning behaviour³³ is exhibited by many octopus species, wherein a hole is dug in the seabed or any other soft substrate, and used as a residence for several days to a few weeks. In some cases, octopuses collect stones and arrange them around the opening of the den. Octopuses usually capture prey by going on hunting trips that can last up to several hours

32 »Las máquinas autopoieticas poseen individualidad« Maturana/Varela 1973/²1994: 60; »autopoiesis« (Rodríguez/Torres 2003).

33 Man beachte die ausgedrückte Prozesshaftigkeit: »denning«, »hunting«, »identifying«, die in »determining its significance in a given context« gipfelt, was nicht nur auf die Bewegung, sondern auch auf den Schaffensprozess der Bedeutung hinweist (Grasso 2014; Giesen et al. 2020; Prozess der »Kontextualisierung« Gumperz 1992).

and cover large distances, after which they return to the den with the prey to eat [often using different ways for coming back than the ones used for going out]. [...] In the context of navigating using environmental landmarks, conditional discrimination is expressed as identifying a certain feature as distinct from similar ones and determining its »significance« in the given context. (Carls-Diamante 2022: 6, Ergänzung KJ)³⁴

Andere Forscher*innen ergänzen ein »what-where-when [it happened on the basis of a specific past experience]« Gedächtnis, das sie als »episodic-like memory« charakterisieren (Jozet-Alves/Bertin/Clayton 2013: 1033–1035).

[T]his is evidence of parallel evolution of intelligence *within* the cephalopods. [...] There was an expansion of the nervous system within the octopus line [of evolution], and another one, in parallel, in the other cephalopods. (Godfrey-Smith 2017: 198; Ergänzung KJ)

Eine Auffassung, die aus der Perspektive der *Critical Animal Studies*³⁵ ein Kontinuum zwischen Menschen und Tieren annimmt, lässt es zu, dass auch die Verantwortung des Menschen für die Welt als eine mit anderen Lebewesen geteilte akzeptiert werden kann.

The hope is that [...] we might move beyond our anthropomorphism as regards ourselves: our image of ourselves as humanly standing apart from other animals; our inveterate vanity regarding our assumed species identity, based on the specious grounds of our sole proprietorship of language, thought, and creativity. (Massumi 2014: 3)

Wenn Kreativität und raumzeitliche Orientierung für Handlungen, wie ich oben ausgeführt habe, durchaus auch bei Oktopoden beobachtet werden

34 »Among these [advanced cognitive capacities] are the ability to form mental maps of areas surrounding their dens (Hanlon/Messenger 1996), the capacity for concept formation manifested as being able to recognize a given feature of the environment from different angles, and conditional discrimination or the ability to »discriminate between potential cues [present in the environment] and show context (condition) sensitivity« (Hvorecny et al., 2007, p. 449)« (Carls-Diamante 2022: 6; Ergänzungen KJ).

35 »We cover the potential mechanisms involved in the development and maintenance of positive human–animal relationships from the perspective of the animal. This encompasses habituation, associative learning, and possibly attachment or bonding based on communication and social cognition« (Rault et al. 2020: 2).

können, muss die *Agency*, die der Rolle der*des Besitzerin*Besitzers zukommt (»propriatorship« Massumi 2014), mit diesen Mollusken mindestens in mancherlei Hinsicht geteilt werden, was sicherlich, *mutatis mutandi*, auch für andere Tiere zu einem jeweils artspezifischen Maße göltig sein wird.

7 Liquide Körper – Liquide Kommunikation – Liquide Sprachen

Welche Schlüsse lassen sich aus der Betrachtung der Kommunikation des Tauchers mit dem Oktopus und allgemeiner aus den tierlich-menschlichen Beziehungen ableiten? Bauman/Palese (2013) haben den Begriff des Liquiden als einen Leitbegriff des epistemischen Denkens im 21. Jahrhundert eingeführt, das durch Mobilität und Austausch geprägt ist. Der Oktopus verkörpert die Dimension des beständigen Formwechsels und der beständigen Anpassung an wechselnde Umstände in einzigartiger Weise. Er zeigt uns mit seinem liquiden Körper, was es heißt, »constantly ready and prone to change (its shape)« (Bauman/Palese 2013: 2) zu sein. Sicht- und spürbar ist aber auch artenübergreifend die menschlich-tierliche Interaktion selbst als eine liquide Praxis zu begreifen, und in dieser Perspektive als mit der Interaktion zwischen Erwachsenen und Kindern oder Haustieren vergleichbar (Abschnitt 1 und 6). Liquide heißt hier, dass die beispielsweise in Lexika gelistete Entsprechung von Lemma und Bedeutung beim jeweiligen Miteinander-Kommunizieren nicht vorausgesetzt werden kann, sondern gemeinsam ausgehandelt werden muss. In einem literarischen Format wird diese Willkürlichkeit der Benennungen, Arbitrarität der Zeichen, sehr anschaulich in Peter Bichsels Geschichte *Ein Tisch ist ein Tisch* (1995) vorgeführt. Auch Interaktionen zwischen ungleichen Gesprächspartner*innen in der Mensch-Mensch-Interaktion, die sich beispielsweise im Alter (Kinder, Greise), in ihrer (sprachlichen) Herkunft (»semiotic repertoires deployed by agentive individuals [and their] material ecologies«: Canagarajah 2021) und ihrer materiellen Einbettung, in ihrer Zugehörigkeit oder in anderen Aspekten unterscheiden, können sich nicht allein auf eindimensional, beispielsweise nur sprachlich gebundene Repertoires stützen, sondern müssen mehrdimensionale Repertoires einsetzen und die Bedeutung der darin geordneten Zeichen aktiv aushandeln, um zukünftig einen zunehmend konventionalisierten Gebrauch vorzubereiten. Liquide sind die Zeichen selbst, aber insbesondere auch ihre Bedeutungszuordnungen und ihre Aktualisierung in der Interaktion. Sogar Sprachen kommt mit Schneider (im Druck) ein liquider Charakter zu, insbesondere in mehrsprachigen

Kontexten, wobei je nach historischem und lokalem, regionalem sowie auch nationalem Kontext zu dieser Vielsprachigkeit auch sogenannte Kreolsprachen, Dialekte oder Mundarten bis hin zu besonderen, selbst ideolektale Sprechweisen beitragen können.

Für Tier-Mensch- wie für Mensch-Mensch-Interaktionen gilt gleichermaßen, dass jede Begegnung einzigartig ist. »Man muss gewillt sein zuzuhören« (Montgomery 2017: 157). Diese Maxime ist nicht nur metaphorisch, sondern als Auftrag an alle in die Interaktion eingebundenen »agents« zu verstehen, wie es sich auch im Film in dem einleitend gesprochenen Satz »Das Tier nahm Kontakt mit mir auf« (min. 00:19:14-00:19:17) zeigt. Diese Kontaktaufnahme unter Wasser ist multimodal. Sie wird als visuelles Ereignis (Augenkontakt: Day 1 »[das Tier] starrte mich an« min. 00:14:26-00:14:37), als Berührung (Körperkontakt »Er kam heraus« min. 00:18:07-00:18:09; »Ich streckte meine Hand aus [der Oktopus tastet sie mit seinen Saugnäpfen ab]« min. 00:18:38-00:18:42), als räumliche Kontingenz (»Du darfst in meine Oktopushöhle kommen« min. 00:21:22-00:21:25; cf. »spatial learning of octopuses« Boal 2000) und sogar als synchrone Fortbewegung beider Akteure im Wasser für die Zuschauer*in erfahrbar gemacht. Die Entgrenzung der den Sprachgebrauch auslassenden Kommunikation spiegelt sich im Satz: »Grenzen schienen sich aufzulösen zwischen mir und dem Oktopus« (min. 00:21:25-00:21:29), und sogar als Empathie, wenn der Taucher den Verlust eines ihrer Arme durch einen Hai, der den Oktopus angreift, »als ob es mir zugestoßen wäre« (Day 26 00:18:07-00:18:09) wahrnimmt. Empathisch fragt Foster sich, »ob es ih[r] gut geht« (min. 01:09:40-01:09:45) bis er sich schließlich eingesteht: »Ich habe mich in dieses Lebewesen verliebt, in diesen Ort« (min. 01:10:15-01:10:21).

Auch bei der Kommunikation unter Menschen sind die von ihnen neben anderen Mitteln der Kommunikation benutzten Sprachen in ihrer Vielfalt, besonders auch angesichts ihrer sich ständig wandelnden Varietäten als liquide zu begreifen (zu Belize in der Gegenwart: Schneider im Druck; zum Sprachgebrauch des »Kust-Portugees« im 15.-19. Jahrhundert in Westafrika: Dakubu 2012). Für solche Kontaktvarietäten, die sich in mehrsprachigen Kontexten in der von Migration geprägten Geschichte der Menschheit immer wieder herausgebildet haben, verwendete Schuchardt im 19. Jahrhundert für eine lusophone Sprechweise entlang der Küsten Westafrikas die Bezeichnung »Negerportugiesisch«. Damit bezeichnete er eine im Kontext europäisch-afrikanischer Handelsbeziehungen gebrauchte Lingua Franca (Schuchardt 1882; Jungbluth im Druck). Keinesfalls können sie abschließend anhand eines Lexikons und einer Grammatik allein beschrieben werden. Im postkolonialen

Kontext zeigen insbesondere plurilinguale Gesellschaften in Lateinamerika, beispielsweise in Belize (ebenfalls: Schneider im Druck) als auch in Afrika (Yakpo 2019), dass der in Europa historisch entwickelte Bedarf für einen sprachlichen Standard ideologisch fundiert und national besetzt, keineswegs jedoch allgemein verbreitet ist. Vielmehr sind angesichts ihres historischen Gewordenseins alle unsere Sprachen als »creolized tongues« (Aboh 2019) zu beschreiben. Ein *wechselseitiges Aufeinander eingestellt Sein* (Meijer 2018: 39) ist nicht nur in der tierlich-menschlichen Begegnung, sondern auch in der Beziehung zwischen Menschen bedeutsam.

Vergleichbar den in diesem Beitrag fokussierten menschlich-tierlichen Begegnungen sind Kommunikations- und Sprechweisen immer an unterschiedliche Gesprächspartner*innen und ihre Repertoires anzupassen, die längst nicht mehr einförmig sind, es eigentlich auch nie waren, wie die Dialektologie und die Variationslinguistik an Forschungen in allen Erdteilen in Geschichte und Gegenwart zigfach nachgewiesen haben (stellvertretend: Wenker 1881; Gilliéron/Edmont 1897–1900). Die Vorstellung, Sprache sei abschließend als Sprachsystem zu beschreiben, wie es beispielsweise aus einer strukturalistischen Sicht Saussure (1916) oder aus einer generativen Perspektive Chomsky (1957) unternommen haben, ist meines Erachtens irreführend. Solche Konstrukte sind eng mit nationalen Vorstellungen und Positionierungen der Überlegenheit verknüpft, die heutzutage überwunden sein sollten. Bedeutung entsteht in der Interaktion, sie entsteht als Prozess und ist immer als unabgeschlossen zu begreifen. Im Fortführen der Kommunikation und im Handeln selbst verändert sie sich laufend. *A fortiori* zeigt sich ihr liquider Charakter in der Interaktion mit Fremden, nicht nur zwischen Menschen, sondern auch artenübergreifend in der Begegnung mit Tieren, besonders pointiert mit Meerestieren, weil ihr Raum den Gebrauch aller akustischer Zeichen stark einschränkt. Bedeutung bleibt immer vorläufig. Sie ist bestenfalls ein Zwischenergebnis der Kooperation zwischen den an der Kommunikation Beteiligten, die nicht nur hinsichtlich ihrer Zugehörigkeiten entgrenzt sind und sich mit jedem Gegenüber von Neuem positionieren. Jede dieser Beziehungen ist einzigartig.

Immer wieder aufs Neue muss es darum gehen, die wechselseitige Verbindung des Ego und Alter als »partnership« aufzubauen, zu pflegen und weiterzuentwickeln, wobei der Rollentausch zwischen den Interaktionspartner*innen grundlegend ist. Ziel ist ein *two way relationship* (Anderson et al. 2010), eine Beziehung, welche die »leiblichen Bedingungen [des menschlichen und gegebenenfalls auch des tierlichen Körpers], wie sie sich aus dem Sitz der Kommu-

nikationsorgane für die Blickstellung in der kommunikativen Dyade ergeben« (Weinrich 1988, ²1995: 81; Jungbluth 2005: 19–22; vgl. Merleau-Ponty 1945) angemessen berücksichtigt, was auch für die hier fokussierte Mensch-Oktopus-Dyade zutreffend ist.

Forschung zur artenübergreifenden Kommunikation anzustreben, ist zweifelsohne ein interdisziplinäres Handlungsziel, das nur gemeinsam erfolgreich sein kann. Möglicherweise hat die Inklusion naturwissenschaftlicher Forschung das Potential, Grenzen zu überwinden und eine über die Geisteswissenschaften hinausreichende Regelmäßigkeit der Züge (›turns‹) dieser Interaktionen zu verallgemeinern. Es sind die Individuen, die sich darüber klar und einig sein müssen:

Die wechselseitige Orientierung des Handelns mehrerer Individuen und der Aufbau von sozialen Beziehungen (angefangen bei der Paar-Beziehung, der sog. Dyade [.]), ist also nur möglich, wenn sich die Individuen an gemeinsam gekannte und akzeptierte Standards (Regeln) des Zusammenlebens halten. (Schäfers 1992: 26)

Dass gegenseitiges Vertrauen sogar in den beiden Mensch-Oktopus-Dyaden bedeutsam ist, kann uns zuversichtlich machen. Der Taucher leiht seiner Partnerin seine Stimme (Armbruster 2015), wenn er sagt: »It's like ›I totally trust this human and I'm coming out of the den‹« (min. 00:20:16–00:20:23). Zugleich eignet er sich die Rolle des Beobachters (›observership‹) an (Glaserfeld 1991; 1996), die auch seiner tierlichen Partnerin eigen ist. Sie verlässt »sich auf ihr dezentrales Gehirn in den Tentakeln, mittels derer sie in Kommunikation mit ihrer Mitwelt [tritt]« (Graeff 2022; Ergänzung KJ).

Die Oktopus-Mensch-Beziehung ist in ihrer kommunikativen Bewältigung keineswegs so besonders, wie sie vielleicht zunächst erscheinen mag. Auch alltägliches kommunikatives Handeln unter Menschen und der darin sichtbar werdende Sprach(en)gebrauch zeigen, dass allem Sprechen ein liquider Charakter eigen ist. Dieser erlaubt es, situativ angemessen Sprachen und Sprechweisen miteinander und mit anderen Kommunikationsmitteln kreativ zu verbinden und *in situ* Bedeutung auszuhandeln. Das Ziel, Bedeutung gemeinsam zu entfalten, schafft die Grundlage für den Aufbau und Erhalt jeder vertrauensvollen Beziehung. Mit den Forschungen der *Human Animal Studies* hat dieses Handeln insofern Überschneidungen, als es auch mit Peter Singer darum geht, Machtpositionen aufzugeben:

Je soutiens qu'il ne peut y avoir aucune raison — hormis le désir égoïste de préserver les privilèges du groupe exploiteur — de refuser d'étendre le principe fondamental d'égalité de considération des intérêts aux membres des autres espèces. (Singer 1975: 1)

Ich bin davon überzeugt, dass es keinen Grund dafür gibt –außer dem egoistischen Wunsch, die Privilegien einer ausbeuterischen Gruppe zu bewahren–, das grundlegende Prinzip der Gleichheit der Interessen jedes anderen Lebewesens zurückzuweisen. (Übersetzung KJ)

Literaturangaben

- Aboh, Enoch O. (2019): Our creolized tongues. In: Doron, Edit/Malka Rappaport Hovav/Yael Reshef/Moshe Taube (Hg.), *Language contact, continuity and change in the genesis of Modern Hebrew*, Amsterdam: Benjamins, 287–320.
- Albertazzi, Liliana/Gert J. van Tonder/Dhanraj Vishwanath (2024): *Perception beyond Inference. The Information Content of Visual Processes*, Cambridge: MIT Press.
- Anderson, Roland D./Jennifer A. Mather/Mathieu Q. Monette/Stephanie R. M. Zimsen (2010): Octopuses recognize individual humans. In: *Journal of Applied Welfare Science* 13, 261–272.
- Armbruster, Karla (2015): What do we want from talking animals? Reflections on Literary Representations of animal voices and minds. In: DeMello, Margo (Hg.), *Speaking for animals. Animal autobiographical writing*, New York: Routledge, 17–33.
- Austin, John L. (1962): *How to do things with words*, Oxford: OUP.
- Balgar, Karsten (2015): Leiblichkeit und tierliche Agency. Die Handlungsfähigkeit von Tieren im Kontext von Leiblichkeitskonzepten. In: Wirth, Sven/Anett Laue/Markus Kurth/Katharina Dornenzweig/Leonie Bossert/Karsten Balgar (Hg.), *Das Handeln der Tiere*, Bielefeld: transcript, 137–148. doi.org/10.14361/9783839432266-005.
- Bauman, Zygmunt (2000): *Liquid Modernity and Beyond*, Cambridge: Polity.
- Bauman, Zygmunt/Emma Palese (2013): Individual and society in the liquid modernity. In: *SpringerPlus* 2:191. doi.org/10.1186/2193-1801-2-191.
- Bichsel, Peter (1995): *Ein Tisch ist ein Tisch*, Frankfurt a.M.: Suhrkamp.
- Boal, Jean Geary/Andrew W. Dunham/Kevin T. Williams/Roger T. Hanlon (2000): Experimental evidence for spatial learning in Octopuses. In: *Journal of Comparative Psychology* 114, 246–252.

- Böhm, Alexandra (2022): Teaching Empathy and Emotions: J.M. Coetzee's The Lives of Animals and Human-Animal Studies. In: Hübner, Andreas/Micha Edlich/Maria Moss (Hg.), *Teaching Human-Animal Studies*, Berlin: Neofelis, 131–148.
- Böhm, Alexandra/Pamela Steen (2023): *Mehrdeutigkeit und Unsicherheit in Tier-Mensch-Begegnungen: Linguistische und literaturwissenschaftliche Zugänge, Editorial*, *LiLi* 53:4, 153–159.
- Bourdieu, Pierre (1979): *Entwurf einer Theorie der Praxis auf der ethnologischen Grundlage der kabyliischen Gesellschaft*, Frankfurt a. M.: Suhrkamp.
- Buber, Martin (1923): *Ich und Du*, Leipzig: Insel-Verlag.
- Bühler, Karl (1934): *Sprachtheorie. Die Darstellungsfunktion der Sprache*, Stuttgart: G. Fischer.
- Burghardt, Gordon M. (2005): *The genesis of animal play: Testing the limits*, Cambridge: MIT Press.
- Canagarajah, Suresh (2021): Materialising semiotic repertoires: challenges in the interactional analysis of multilingual communication, In: *International Journal of Multilingualism* 18:2, 206–225.
- Carls-Diamante, Sidney (2022): Where Is It Like to Be an Octopus? In: *Front Syst Neurosci.* 16:840022. doi.org/10.3389/fnsys.2022.840022.
- Chachu, Sewoenam/Nana Aba Appiah Amfo (im Druck): Accessing Healthcare across Borders: Linguistic Strategies Employed in Multilingual Cross-border Healthcare Interactions in West Africa. In: Nossem, Eva (Hg.), *Border Languageing: Multilingual Practices on the Border*, Serie: Border Studies. Cultures, Spaces, Orders, Baden-Baden: Nomos.
- Cheney, Dorothy L./Robert M. Seyfarth (1981): Selective forces affecting the predator alarm calls of vervet monkeys. In: *Behaviour*, 76:1-2, 25–61.
- Chomsky, Noam (1957): *Syntactic Structures*, Berlin: Mouton.
- Coulmas, Florian (1977): *Rezeptives Sprachverhalten: eine theoretische Studie über Faktoren des sprachlichen Verstehensprozesses*, Hamburg: Buske.
- Cowley, Stephen J./Marie-Theres Fester-Seegeer (2024). Coming to conceive: Radical Embodied Thinking in Action. In: King, Mark E./Paul J. Thibault (Hg.), *Learning as Interactivity, Movement, Growth and Becoming*, Volume 1: Ecologies of Learning in Higher Education, London and New York: Routledge, 33–45.
- Dakubu, Mary Esther Kropp (2012): The Portuguese language on the Gold Coast, 1471–1807. In: *Ghana Journal of Linguistics* 1:1, 15–33.
- Davis, Hank/Dianne Balfour (1992): *The inevitable bond*, Cambridge: CUP.

- Derrida, Jacques (2006): *Le animal que donc je suis*, Paris: Galilée; 2016. Deutsche Übersetzung: *Das Tier, das ich also bin*, Wien: Passagen; 2002. Portugiesische Übersetzung des ersten Kapitels nach den Kongressakten aus Cerisy, France: *O animal que logo sou*, São Paulo: UNESP.
- Dreckmann, Kathrin/Verena Meis (2022): *Fluide Mediale. Medialität, Materialität und Medienästhetik des Fluiden*, Berlin: De Gruyter.
- Ehlich, Konrad (1994): Funktion und Struktur schriftlicher Kommunikation. In: Günther, Hartmut/Otto Ludwig (Hg.), *Schrift und Schriftlichkeit*. Handbuch Sprache und Kommunikation 10:1, Berlin/Boston: De Gruyter, 18–41.
- Emmeche, Claus (2015): Semiotic scaffolding of the social self in reflexivity and friendship. In: *Biosemiotics* 8, 275–289.
- Erikson, Lars. H. (2007): Sich-Begrüßen, Sich-Verabschieden. Formeln der aktuellen standarddeutschen Umgangssprache. In: *Studie zur heutigen Landeskunde in den deutschsprachigen Ländern*. Abrufbar unter: <https://web.archive.org/web/20070116152629/>
- Esser, Hartmut (1991): *Modellierung sozialer Prozesse, Alltagshandeln und Verstehen. Zum Verhältnis von erklärender und verstehender Soziologie am Beispiel von Alfred Schütz und »Rational Choice«*, Tübingen: Mohr.
- Fiorito, Graziano/Pietro Scotto (1992): Observational learning in Octopus vulgaris. In: *Science* 256, 545–547.
- Francescoli, Gabriel (2021): »Semiotic Canalization«: a Process Directing the Use and Interpretation of Signals in Animal Interactions? In: *Biosemiotics* 14, 199–207.
- Frisch, Karl von (1923): Über die »Sprache« der Bienen. In: *Zoologische Jahrbücher: Zeitschrift für Systematik, Geographie und Biologie der Tiere. Abteilung für allgemeine Zoologie und Physiologie der Tiere*, Bd. 40, 1–186.
- Gallese, Vittorio/George Lakoff (2005): »The Brain's Concept: The Role of the Sensory-Motor System. In: *Cognitive Neuropsychology* 22:3-4, 455–479.
- Geese, Natalie (2015): Autonom handelnde Individuen, Kooperationspartner_innen, Natur- oder Kulturwesen? Der Beitrag von Führungshunden zur Herstellung von Agency in Mensch-Tier-Triaden. In: Wirth, Sven V./Anett Laue/Markus Kurth/Katharina Dornenzweig/Leonie Bossert/Karsten Balgar (Hg.), *Das Handeln der Tiere*, Berlin/Boston: De Gruyter, 227–244.
- Geese, Natalie (2017): Stigmatisierungen in Mensch-Führungshund-Triaden. In: Burzan, Nicole/Ronald Hitzler (Hrsg.), *Auf den Hund gekommen. Erlebniswelten*, Wiesbaden: Springer VS, 139–155.
- Giesen, Lena van/Peter B. Kilian/Corey A. H. Allard/Nicholas W. Bellono (2020): Chemotactile Sensation in Octopus. In: *Cell* 183:3, 594–604.

- Gilliéron, Jules/Edmond Edmont (1897–1900): *Atlas linguistique de la France*.
 Abrufbar unter: https://www.lexilogos.com/atlas_linguistique_france.htm
- Glasersfeld, Ernst von (1991; 1996): Farewell to Objectivity. In: *Systems Research* 13:3, 279–286. Revised and expanded version of Abschied von der Objektivität. In: Watzlawick, Paul/Peter Krieg (Hg.), *Das Auge des Betrachters*, München: Piper.
- Godfrey-Smith, Peter (2017): *Other Minds. The Octopus and the evolution of intelligent life*, London: Collins.
- Graeff, Alexander (2022): Der queere Blick: Tentakel ausbilden, *Frankfurter Rundschau* 7.2.2022. Abrufbar unter: <https://www.fr.de/kultur/gesellschaft/queer-queerness-geschlechteridentitaeten-oktopus-der-queere-blick-tentakel-ausbilden-91282095.html>
- Grasso, Frank W. (2014): The octopus with two brains: How are distributed and central representations integrated in the octopus central nervous system? In: Darmaillacq, Anne-Sophie/Ludovic Dickel/Jennifer Mather (Hg.), *Cephalopod Cognition*, Cambridge: CUP, 94–122.
- Greifenstein, Sarah/Christina Schmitt (2014): Cinematic communication and embodiment. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva Ladewig/David McNeill/Jana Bressemer (Hg.), *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*. HSK 38:2, Berlin/Boston: De Gruyter Mouton, 2061–2070.
- Gumperz, John (1992): Contextualization revisited. In: Auer, Peter/Aldo di Luzio (Hg.), *The Contextualization of Language*, Amsterdam: Benjamins, 39–53.
- Habermas, Jürgen (1995): *Theorie des kommunikativen Handelns*, Frankfurt: Suhrkamp.
- Haitzinger, Nicole (2019): Der Oktopus als gegenläufig fluide und vielsinnige Figur in den szenischen Künsten. Vortrag, Dresden: Universität Dresden.
- Hanlon, Roger T./John B. Messenger (1996): *Cephalopod Behaviour*, Cambridge: CUP.
- Haraway, Donna (2018): *Unruhig bleiben: Die Verwandtschaft der Arten im Chthuluzän*, Frankfurt a.M.: Campus.
- Haraway, Donna (2008): *When Species meet*, Minnesota: University of Minnesota Press.
- Harms, William F. (2004): Primitive Content, Translation, and the Emergence of Meaning in Animal Communication. In: Oller, Kimbrough/Ulrike Griebel (Hg.), *Evolution of Communication Systems: A Comparative Approach*, Cambridge: MIT Press, 31–48.

- Hvorecny, Lauren M./Jessica L. Grudowski/Carrie J. Blakeslee/Tiffany L. Simmons/Paula R. Roy/Jennifer A. Brooks/Rachel M. Hanner/Marie E. Beigel/Miranda A. Karson/Rachel H. Nichols/Johanna B. Holm/Jean G. Boal (2007): Octopuses (*Octopus Bimaculoides*) and cuttlefishes (*Sepia Pharaonis*, *S. Officinalis*) can conditionally discriminate. In: *Animal Cognition* 10, 449–459.
- Jozet-Alves, Christelle/Marion Bertin/Nicola S. Clayton (2013): Evidence of episodic-like memory in cuttlefish. In: *Current Biology* 23:23, 1033–1035.
- Jungbluth, Konstanze (2005): *Pragmatik der Demonstrativpronomina*, Berlin/Boston: De Gruyter. Imprint Niemeyer.
- Jungbluth, Konstanze (2016a): Co-Constructions in Multilingual Settings, In: Fernández-Villanueva, Marta/Konstanze Jungbluth (Hg.), *Beyond Language Boundaries: Multimodal Use in Multilingual Contexts*, Boston/Berlin: De Gruyter Mouton, 151–166.
- Jungbluth, Konstanze (2016b): Mensageiros e porta-vozes. Transmissão de assuntos jurídicos aos não-letrados. Uma prática discursiva histórica entre oralidade e escritura, In: Martins, Marco Antonio/Lucrécio Sá (Hg.), *Rumos da linguística Brasileira no século XXI: historiografia, gramática e ensino*, São Paulo: Blucher, 203–215.
- Jungbluth, Konstanze (2018): Positions – Constellations – Practices: Referring to the space in front of the hearer, In: Federica da Milano/Paolo Zublena (Hg.), *Archivio glottologico italiano*, 1031, Le Monnier Firenze. Abrufbar unter: <https://riviste.mondadorieducation.it/archivio-glottologico-italiano/rivista/>
- Jungbluth, Konstanze (im Druck): Brokers on the move. Encounters between Europeans and Africans in the Portuguese Seaborne Empire. In: Enoch Aboh/Salikoko Mufwene (Hg.), *Uniformitarianism in Language Speciation*. Series Cambridge Approaches to Language Contact, CUP.
- Kappelhoff, Herrmann/Michael Wedel (seit 2015; org.): *Cinipoetics. Poetologien audiovisueller Bilder*, Berlin: FU Berlin: DFG Kollegforschungsgruppe. Abrufbar unter: <https://www.cinipoetics.fu-berlin.de/>.
- Kleindorfer, Sonia/Patricia McAllister-Käfer (2024): *Die erstaunliche Welt der Graugänse. Wie sie leben, kommunizieren und füreinander sorgen*, Wien: Brandstätter.
- Kummer, Elfi (2010): *Der schlaue Oktopus*, München: TZ. Abrufbar unter: <http://www.tz.de/bayern/otto--der-schlaue-oktopus-70263.html>.
- Langacker, Ronald W. (2009): *Investigations in Cognitive Grammar*, Berlin/Boston: De Gruyter Mouton.

- Leon, Donna (2022): *Die Seele eines Tintenfischs. Nachwort zu Sy Montgomery, Rendezvous mit einem Oktopus*, Zürich: Diogenes, 500–517.
- Lüdemann, Dagny (2020): Verliebt in einen Oktopus, *ZEIT ONLINE*. Abrufbar unter: <https://www.zeit.de/kultur/film/2020-09/my-octopus-teacher-netflix-dokumentation-tierfilm-craig-foster>.
- Luhmann, Niklas (1982): Autopoiesis, Handlung und kommunikative Verständigung. In: *Zeitschrift für Soziologie* 11, 366–379.
- Massumi, Brian (2014): *What Animals teach us about Politics*, Durham: Duke University Press.
- Mather, Jennifer A. (2019): What is in an octopus's mind? In: *Animal Sentience* 26:1. doi.org/10.51291/2377-7478.1370.
- Mather, Jennifer A. (2004): Cephalopod Displays: From Concealment to Communication. In: Oller, D. Kimbrough/Ulrike Griebel (Hg.), *Evolution of Communication Systems*, Cambridge, MA: MIT Press, 193–213.
- Maturana Romesin, Humberto/Francisco J. Varela Gracia (1973; ²1994; ⁶2005): *De máquinas y seres vivos. Autopoiesis: la Organización de lo vivo*, Barcelona: Lumen.
- Meijer, Eva (2018): *Die Sprachen der Tiere*, Berlin: Matthes & Seitz.
- Merleau-Ponty, Maurice (1945): *Phénoménologie de la perception*, Paris: Gallimard. Deutsche Übersetzung (1966): *Phänomenologie der Wahrnehmung*, Berlin: De Gruyter.
- Midgley, Mary (1984): *Animals and why they matter*, Athens: University of Georgia Press.
- Millikan, Ruth Garrett (2004): On reading signs: some differences between us and the others. In: Oller, Kimbrough/Ulrike Griebel (Hg.), *Evolution of Communication Systems: A Comparative Approach*, Cambridge: MIT Press, 15–30.
- Millikan, Ruth Garrett (1984): *Language, Thought, and Other Biological Categories: New Foundations for Realism*, Cambridge: MIT Press.
- Montgomery, Sy (2017): *Rendezvous mit einem Oktopus*, Zürich: Diogenes. Englische Originalausgabe (2015): *The Soul of an Octopus: A Surprising Exploration into the Wonder of Consciousness*, New York: Atria.
- Müller, Cornelia/Ulrike Bohle (2007): Das Fundament fokussierter Interaktion. In: Schmitt, Reinhold (Hg.), *Koordination. Analysen zur multimodalen Interaktion*, Tübingen: Narr, 129–166.
- Oliver, Kelly (2017): Earth Ethics and Creaturely Cohabitation. In: Ohrem, Dominik/Roman Bartosch (Hg.): *Beyond the Human-Animal Divide: Creaturely Lives in Literature and Culture*, New York: Palgrave Macmillan, 21–41.

- Oller, Kimbrough/Ulrike Griebel (2004): Theoretical and Methodological Tools for Comparison and Evolutionary modeling of Communicative Systems. In: Oller, Kimbrough/Ulrike Griebel (Hg.), *Evolution of Communication Systems: A Comparative Approach*, Cambridge: MIT Press, 3–11.
- Raithelhuber, Eberhard (2008): Von Akteuren und agency – eine sozialtheoretische Einordnung der structure/agency-Debatte. In: Homfeldt, Hans G./Wolfgang Schröer/Cornelia Schweppe (Hg.), *Vom Adressaten zum Akteur*, Opladen: Barbara Budrich, 17–46.
- Rajagopalan, Kanavillil (2013): Linguistics as a performative science. In: Vlian Jr, Orlando/Cida Caltabiano (Hg.), *Lingua(gens) e suas múltiplas faces*, Campinas, SP: Mercado de Letras, 27–36.
- Rault, Jean-Loup/Susanne Waiblinger/Xavier Bolvin/Paul Hemsworth (2020): The Power of a Positive Human–Animal Relationship for Animal Welfare. In: *Frontiers in Veterinary Science* 7:590867. doi.org/10.3389/fvets.2020.590867.
- Reckwitz, Andreas (2016): *Kreativität und soziale Praxis*, Bielefeld: transcript.
- Rodríguez M., Dário/Javier Torres N. (2003): Autopoiesis, la unidad de una diferencia. Luhmann y Maturana. In: *Sociologias* 5:9, 106–140. Deutsche Übersetzung von Peter Birle und Jessica Zeller: *Autopoiesis, die Einheit einer Differenz: Luhmann und Maturana*, Berlin: IAI.
- Rosa, Hartmut (2016): *Resonanz – eine Soziologie der Weltbeziehung*, Berlin: Suhrkamp.
- Roscher, Mieke (2015): Zwischen Wirkungsmacht und Handlungsmacht. Sozialgeschichtliche Perspektiven auf tierliche Agency. In: Wirth, Sven/Anett Laue/Markus Kurth/Katharina Dornenzweig/Leonie Bosser/Karsten Balgar (Hg.), *Das Handeln der Tiere*, Bielefeld: transcript, 43–66.
- Saussure, Ferdinand de (1916/1989): *Cours de Linguistique Générale*, Edition Critique Vol. 1, Wiesbaden: Harrassowitz.
- Schäfers, Bernhard (2000): *Grundbegriffe der Soziologie*, Opladen: Leske.
- Schlieben-Lange, Brigitte (1998): Alterität als sprachtheoretisches Konzept. In: *LiLi* 110, 41–57.
- Schneider, Britta (im Druck): *Liquid Languages. Constructing Language in Late Modern Cultures of Diffusion* (Series: ›Cambridge Approaches to Language Contact‹), Cambridge: CUP.
- Schuchardt, Hugo (1882): *Kreolische Studien I: Über das Negerportugiesische von S. Thomé*, Wien: Buchhändler der Kaiserlichen Akademie der Wissenschaften.

- Searle, John R. (1969): *Speech acts. An essay in the philosophy of language*, Cambridge: CUP.
- Sinah, Chris (2004): The Evolution of Language: From Signals to Symbols to Systems. In: Oller, Kimbrough/Ulrike Griebel (Hg.), *Evolution of Communication Systems: A Comparative Approach*, Cambridge: MIT Press, 217–236.
- Singer, Peter (1975): *La libération animale*, Paris: Payet.
- Steen, Pamela (2023): ›Sprechende‹ Tiere im Zoo: Animation, Empathie und Fiktion in Zoo-Doku-Soaps. In: Böhm, Alexandra/Pamela Steen (Hg.), *Mehrdeutigkeit und Unsicherheit in Tier-Mensch-Begegnungen: Linguistische und literaturwissenschaftliche Zugänge*, *LiLi* 53:2, 317–349.
- Stone, Christopher D. (1972): Should trees have standing? Toward Legal Rights for Natural Objects. In: *Southern California Law Review* 45, 450–501.
- Uexküll, Jakob von (1958): *Streifzüge durch die Umwelten von Tieren und Menschen. Bedeutungslehre*, Hamburg: Rowohlt.
- Weinrich, Harald (1988; ²1995): Über Sprache, Leib und Gedächtnis. In: Gumbrecht, Hans-Ulrich (Hg.), *Materialität der Kommunikation*, Frankfurt a.M.: Suhrkamp, 80–93.
- Wenker, Georg (1881): Sprach-Atlas von Nord- und Mitteleuropa auf Grund von systematisch mit Hilfe der Volksschullehrer gesammeltem Material aus circa 30.000 Orten. Abth. I, Lief. 1, Straßburg/London: Trübner. De Gruyter. doi.org/10.1515/9783111579061.
- Yakpo, Kofi (2019): *A Grammar of Pichi*, Berlin: Language Science Press. doi.org/10.5281/zenodo.2546450.
- Yakpo, Kofi (im Druck): The Evolution of Copula Systems in West African Pidgin: A Uniformitarian Perspective. In: Mufwene, Salikoko/Enoch Aboh (Hg.), *Uniformitarianism in Language Speciation*, Series CALC, Cambridge: CUP.
- Zweigert, Konrad/Hein Kötz (1996): *Einführung in die Rechtsvergleichung*, Tübingen: Mohr.

Filmverzeichnis

- My Octopus Teacher (2020). Directed by Pippa Ehrlich and James Reed. Produced by Craig Foster. Deutsche Synchronisation von Gesine Hirsch: Mein Lehrer, der Krake. Los Gatos, USA.

Prof. Dr. Konstanze B. Jungbluth

Europa-Universität Viadrina

Große Scharnstr. 58

15230 Frankfurt (Oder)

Deutschland

jungbluth@europa-uni.de

https://www.borders-in-motion.de/de/team/mitglieder/jungbluth_konstanze/index.html

Die Stimme als Interface

Entgrenzungen von Mensch und Maschine anhand von Spike Jonzes Film HER

Dorothea Horst

Abstract *Der Beitrag befragt anhand von Spike Jonzes Film HER (USA, 2013) etablierte anthropozentrische und philosophische Vorstellungen der Stimme. In dem Film wird einem körperlosen Operating System mit dem Namen Samantha eine – scheinbar menschliche – Stimme gegeben. Er eröffnet so Fragen nach deren Status vor dem Hintergrund aktueller technologiebedingter kommunikativer Verschränkungen und Transgressionen, und Raum zur Reflexion über etablierte (sprach-)wissenschaftliche und kulturelle Epistemologien zu Stimme und Sprache: ihre menschliche Exklusivität, ihre Vergeschlechtlichung und ihre unbedingte Verkörperung. Auf der Grundlage filmwissenschaftlicher Perspektiven auf die Stimme nähert sich der Beitrag anhand ausgewählter Szenen der Überschreitung solcher etablierter Paradigmen. Er entwirft Implikationen für eine entgrenzte kulturwissenschaftliche Linguistik jenseits einer anthropozentrischen Welt-, Subjekt- und systemischen Sprachvorstellung.*

Keywords *Stimme; Körper; Leiblichkeit; Medium; Mensch-Maschine-Interaktion im Film*

1 Eine Stimme ohne Körper?

Im zweiten Buch von »De Anima« definiert Aristoteles (2006) die Stimme als Laut eines beseelten Wesens und legt damit implizit den Grundstein zu ihrer späteren Lesart als distinktives Merkmal des Menschen. Im Zeitalter von Mensch-Maschine-Interaktionen ist das Privileg einer Stimme allerdings nicht mehr allein »dem Menschen« vorbehalten. Sprachassistenten wie Alexa, Siri oder Cortana »lernen« Sprache und fordern dazu auf, mit ihnen zu spre-

chen (z. B. ›ruft‹ Alexa in Amazons Werbeclip potenzielle Käufer*innen explizit auf: »Sprich mit mir!« bzw. lautet der entsprechende Hashtag »#FragAlexa«; siehe die Beiträge von Fester-Seeger, Leblebici oder Schneider i. d. Bd.). Wie ist der Status von Stimme – und Menschsein – vor dem Hintergrund dieser kommunikativen Verschränkungen und Transgressionen zwischen Mensch/en und Maschine/n heute zu denken?

Der Beitrag adressiert diese und weitere Fragen nach der kategorischen Differenz von Mensch und Maschine, nach dem Verhältnis von Stimme und Subjekt oder von Stimme und Geschlecht ausgehend von Spike Jonzes Film *HER* (USA 2013). *HER* erzählt die Geschichte einer bemerkenswerten Liebesbeziehung zwischen den beiden Hauptfiguren Theodore und Samantha. Bemerkenswert ist diese Liebe deshalb, weil Samantha ein ›körperloses‹ Operating System auf dem Computer ist, das mit dem leibhaftigen Menschen Theodore lediglich in stimmlicher Gestalt in (eine intime emotionale) Beziehung tritt. Aus einer anthropozentrischen Perspektive ist eine solche, rein akusmatische Präsenz (Chion 1994), die einen Körper lediglich indiziert, aber nicht besitzt, prekär, denn: »who can conceive of a voice without a body?« (Doane 1980: 33) Mit der ausgestellten Körperlosigkeit Samanthas rückt die Stimme in ihrer Performativität und Medialität, ihrer Ambiguität und ihrem Begehren ins Zentrum. Anhand des Films reflektiert und hinterfragt der Beitrag über die Stimme deshalb klassische Vorstellungen eines vergeschlechtlichten menschlichen Körpers (Quinlivan 2017), eines selbstidentischen Subjekts (Silverman 2016; Žižek 1996) und einer kategorischen Differenz »zwischen human und non-human« (Angerer 2015: 58).

Über die unsichtbare, flottierende Stimme als Schnittstelle für Begegnung und Interaktion rücken längst bestehende Verflechtungen und Beziehungen zwischen Mensch und Maschine bzw. Technik ins Zentrum. Vor diesem Hintergrund lassen sich etablierte (sprach-)wissenschaftliche und kulturelle Epistemologien zu Stimme und Sprache kritisch befragen: Gehört die Stimme eindeutig zu einem menschlich und geschlechtlich gelesenen Körper? Gelangen Stimme und denkendes Subjekt stets zur Deckung? Sind erkenntnistheoretische Binaritäten wie Mensch vs. Nicht-Mensch, Natur vs. Kultur, Geist vs. Körper unstrittig aufrechtzuerhalten? Der Beitrag nähert sich diesen Fragen zunächst über eine Bestandsaufnahme zum Status der Stimme in der Sprachwissenschaft (Abschnitt 2). Dabei zeichnet er nach, wie die Stimme vor dem Hintergrund eines logosorientierten und intellektualistischen Sprachkonzepts sowie einer traditionell systemisch ausgerichteten Linguistik lange ein Randphänomen ist. Demgegenüber zeigt Abschnitt 3, wie sie im Tonfilm

in Fusion mit der visuellen Ebene zu einem genuin *audiovisuellen* Phänomen wird und wie feministische filmwissenschaftliche Ansätze diesen scheinbar selbstevidenten Zusammenschluss von Bild und Ton kritisch reflektieren. Anhand von Jonzes Film *HER* werden in Abschnitt 4 schließlich Überschreitungen damit verbundener Paradigmen (zum Beispiel die menschliche Exklusivität der Stimme und ihre unbedingte Verkörperung) hinsichtlich dreier Aspekte anhand ausgewählter Szenen nachgezeichnet: die Partialität und Vielheit der Stimme, ihre Queerness und ihr nonkonformistischer Entzugscharakter. Im Ergebnis der Analyse reflektiert der Beitrag, inwiefern *HER* als posthumane Perspektive – im Sinne einer Verflechtung von Mensch und Technologie, Kultur und Natur, Geist und Körper, Ratio und Emotio anstelle ihrer Binarität (siehe Pennycook 2018) – betrachtet werden kann und welche Implikationen diese für eine entgrenzte kulturwissenschaftliche Linguistik hat (Abschnitt 5).

2 Der Status der Stimme in der Linguistik

Multimodalität sprachlicher Kommunikation ist mittlerweile sowohl in der Linguistik als auch in ihren Nachbardisziplinen, z. B. den Medienwissenschaften oder Sozialwissenschaften, ein anerkanntes Wesensmerkmal (z. B. Klug/Stöckl 2016; Elleström 2010; Jewitt 2014), das in die theoretische wie empirische Arbeit einbezogen wird. Oralität und Literalität (Ong 2016), redebegleitende Gestik (Müller et al. 2013), audiovisuelle bewegte Bilder (Schmitt 2015) oder Abbildungen (Stöckl 2004) werden als spezifische Modalitäten einzeln, und im Zusammenspiel mit anderen Modalitäten, auf ihre bedeutungstragenden und gesamtbedeutungsherstellenden Aspekte hin in unterschiedlichsten kommunikativen Kontexten untersucht. Dessen ungeachtet stellt die Stimme einen lange Zeit in der sprachwissenschaftlichen Auseinandersetzung mit Sprache eher unberücksichtigten Forschungsgegenstand dar – trotz ihrer zentralen Rolle beim Sprechen, indem sie durch Phonation und weitere physikalische Vorgänge die Artikulation von Rede ermöglicht. Ein naheliegender Grund dafür dürfte die systemische und strukturalistische Tradition sowohl der europäischen als auch der amerikanischen Linguistik sein. Ferdinand de Saussure hat die Disziplin mit seinem auf ein überindividuelles System von Zeichen (die *langue*) ausgerichteten Sprachkonzept seit Beginn des 20. Jahrhunderts grundlegend geprägt (siehe auch Metten 2014: 10–27). Die interne Strukturierung und Gesetzmäßigkeiten dieses Systems bildeten, anstatt seiner vielfältigen konkreten medialen Realisierung, lange den Kern des

Erkenntnisinteresses. Die Lösung von diesem Paradigma und die verstärkte Hinwendung zu materiellen und medialen Eigenschaften von Sprache ist sukzessive erst in den vergangenen zwei bis drei Jahrzehnten erfolgt. Wie die Philosophin Sibylle Krämer ausführt, hat auch die pragmatische Wende u.a. durch Austins und Searles Sprechakttheorie, an diesem systemischen und amedialen Fokus nichts geändert, weil sie gleichermaßen auf überindividuelle und transsituative Gelingensbedingungen des Sprechens abhebt (Krämer 2001). Diesbezüglich macht die Hinwendung zum Embodiment, die sich im Rahmen der Kognitiven Linguistik und dort spezifisch innerhalb der konzeptuellen Metapherntheorie Lakoffs und Johnsons vollzieht, gleichermaßen keine Ausnahme. Sie fragt nach geteilten bedeutungstragenden kognitiven Schemata die ›unser‹ Wahrnehmen, Denken und Handeln grundsätzlich prägen, nicht nach spezifischem Wahrnehmungserleben einer kommunikativen Situation zu einem bestimmten Zeitpunkt zwischen konkreten Interaktionspartner*innen (vgl. Horst i. E., 2020).

Den Grund für diese systemische und amediale Ausrichtung sieht Krämer im Primat eines logosorientierten bzw. intellektualistischen Sprachkonzepts (das teilweise bis heute wirksam ist) (Krämer 2001; Krämer/König 2002). Die Folge dieser Perspektive: »Es ist die *stumme*, es ist eine *entkörperte* Sprache, die zum Gegenstand wird« (Krämer 2005: 223, Herv. DH). Zu einer ähnlichen Einschätzung kommt Per Linell. Er attestiert der Sprachwissenschaft einen »writen language bias« (Linell 2005), durch den stimmliche Aspekte wie Rhythmus, Sprechtempo oder Pausen unberücksichtigt bleiben, weil sie als nicht zugehörig zum wissenschaftlichen Gegenstand ›Sprache‹ gelten, sondern allenfalls als parasprachlich (Linell 2005: 60). Mit der Priorisierung eines autonomen bedeutungstragenden, abstrakten und überindividuell geteilten Zeichensystems kommt der medialen Realisierung der Sprache – zum Beispiel als stimmliche Artikulation – allenfalls der formale Status einer nachträglichen Verlautbarung im Sinne einer medialen Verpackung zu, die per se keine eigenen, für den Akt des Sprechens bedeutungskonstitutiven Wesensmerkmale besitzt.¹ Dagegen könnte nun eingewendet werden, dass stimmliche Aspekte sehr wohl Bestandteil der sprachwissenschaftlichen Forschung sind, und zwar im Rahmen der beiden Teilbereiche Phonetik und Phonologie. Hier ist allerdings darauf hinzuweisen, dass die Phonetik nicht auf die medienpezifischen Aspekte der

1 Diese lange vorherrschende sprachwissenschaftliche Medienvergessenheit wird auch an dem teilweise bis heute rezipierten Sender-Empfänger-Modell der Kommunikation (Shannon/Weaver 1976) evident.

Stimme abhebt, sondern diese hinsichtlich der Produktion, Artikulation und Perzeption von Lauten als physikalische (und damit messbare) Prozesse be- greift und beschreibt. In der Phonologie werden die Sprachlaute – und damit auch die Stimme – grammatikalisiert, d.h. hinsichtlich ihrer Funktion inner- halb des priorisierten abstrakten Zeichensystems kategorisiert, und gehen da- mit wiederum ihrer Medienspezifik, ihrer Stimmlichkeit verlustig. Den Philo- sophen Mladen Dolar hat dieser Umstand zu der lakonischen wie drastischen Feststellung veranlasst, die Sprachwissenschaft habe die Stimme getötet:

Der erste Schritt der Phonologie war also die totale Reduktion der Stimme als Substanz der Sprache. Getreu ihrer apokryphen Etymologie, war die Pho- nologie auf das Töten der Stimme aus – an ihrem Ursprung findet sich das griechische *phoné*, Stimme, aber genauso zutreffend kann man *phonos* hören, Mord. (Dolar 2002: 235)

Unabhängig davon, wie man zu diesem Verdikt stehen mag – verstummt ist die Stimme in der Linguistik nicht. In der Konversationsanalyse bei- spielsweise gehört sie unabdingbar als multimodale Ausdrucksressource zu einem vollständigen Gesprächsdatum. Hier wird sie primär als eingesetztes prosodisches Mittel für die Handlungsorganisation der Interaktionsteilneh- mer*innen angesehen und analysiert (Deppermann 2014: 20). In der Sozio- linguistik gilt sie als akustischer Index für den Körper, dem sie entspringt. Als solcher verortet sie den Körper im sozialen Raum, indem sie sowohl aus der Nutzungs- als auch aus der Wahrnehmungsperspektive mit Kategorien wie Gender, *race* oder Sexualität in Verbindung gebracht wird (Buchholtz/Hall 2016: 179).

Ob als geistiges Verlautbarungsmittel oder als körperliches Index – die Frage, welche Rolle/n und Funktion/en die Stimme innehat, ist komplex. Schützeichel (2011: 89) weist in diesem Zusammenhang auf einen zweiten Diskussionsstrang hin, der die ›Wahrhaftigkeit‹ der Stimme befragt und sich zwischen den beiden Polen Authentizität und Arbitrarität bewegt. Während im ersten Fall von einer bruchlosen Passung zwischen stimmlichem Ausdruck und Sprecher*in ausgegangen wird (ein identitäres Stimmig-Sein, z.B. in Herders Sammlung ›Stimmen der Völker in Liedern‹), ist es im zweiten Fall der flexible Umgang mit verschiedenen akustischen Masken (Canetti 1975) oder Redeweisen (Goffmann 2005) zum situationsadäquaten Eindrucksma- nagement. Geteilte Grundannahme beider Positionen ist, dass die sprechende Person Macht und Kontrolle über ihre Stimme besitzt, jedoch: »Unsere Stimme

hat ein Eigenleben, sie ist uns oft eine Quelle der Befremdung« (Schützeichel 2011: 89). Im plötzlichen Brüchig werden oder Verstummen beispielsweise offenbart sich ein Entzugs- und Überschusspotenzial, das sich der Kontrolle von Sprecher*innen entzieht. Dieser Aspekt berührt den leiblich expressiven Charakter der Stimme, der unter anderem in der Ausdruckstheorie (beispielsweise bei Bühler 1933 oder Cassirer 2004) thematisiert wird. Ausdruck und Ausdruckswahrnehmung sind situationsbezogen affektiv unmittelbar miteinander verschränkt und bedingen einander wechselseitig (Plessner 2016). Stimmlicher Ausdruck gehört damit nicht einer Person, sondern ist eingebettet im, und durchdrungen vom kommunikativen Kontext, in dem er stattfindet.

Mit dem Aufkommen und der zunehmenden Präsenz synthetischer Stimmen in Alltag und Beruf stellt sich die Frage des leiblichen Ausdrucks und der stimmlichen Wahrhaftigkeit noch einmal hochaktuell. Smartphones (beispielsweise Siri von Apple oder der Google Assistant) reagieren auf gesprochene Kommandos und antworten darauf. Künstliche Intelligenz zur Steuerung von Technik und Geräten (beispielsweise Alexa von Amazon) ist in der Lage, sich lernend den sprachlichen Eigenheiten und Wünschen ihrer Nutzer*innen anzupassen. Gleichzeitig etablieren solche soziotechnologischen Praktiken Sprachhierarchien und können sowohl dominante gesellschaftliche Diskurse über Sprache (und Stimme) als auch linguistische Sprachepistemologien prägen (vgl. Leblebici 2021, i.d.Bd.; Schneider 2022, i.d.Bd.). So führt ein den Technologien inhärenter *Algorithmic Bias* (siehe u.a. Weizenbaum 1976) unter anderem zur Homogenisierung von Sprache durch eine Spracherkennung, die primär auf männliche weiße Sprecher und Standardsprache anstelle nicht standardisierter Varietäten ausgerichtet ist (Markl 2022). Während Dolar (2007: 34) noch feststellt, dass »der unpersönlichen Stimme, der mechanisch produzierten Stimme (des Anrufbeantworters, des Computers und so weiter) [...] stets etwas Unheimliches an[hafte]t«, hat sich durch die rasante Weiterentwicklung der Technologien, gestützt von Investitionen großer Konzerne wie Amazon, Apple, Google oder Microsoft, viel verändert. Mittlerweile wird es zunehmend schwieriger, synthetische von menschlichen Stimmen zu unterscheiden, weil KI-basierte Sprachprogramme nicht mehr wie früher Lautschnipsel kombinieren (und deshalb stimmlich erkennbar roboterartig klingen), sondern anhand von Sprachinput lernen, trainieren und schließlich in stände sind, diesen eigenschaftstreu stimmlich zu imitieren (Dworschak 2017).

Solche medientechnologisch bedingten Prägungen und Hybridisierungen von Stimme eröffnen – ebenso wie ihre indexikalische und expressive Dimension – eine Schnittstelle zu medienwissenschaftlichen Reflexionen der Stimme im Film. Vor allem die filmgeschichtliche Entwicklung vom Stummfilm zum Tonfilm und die audiovisuelle Multimodalität, auf deren Grundlage Zuschauer*innen akustische und visuelle Phänomene synthetisieren, machen den Film zu einem interessanten Spielraum künstlerischer Exploration und theoretischer Reflexion. Der folgende Abschnitt zeigt theoretische wie analytische Implikationen dieser Stimm-Körper-Fusionierung ebenso auf wie Möglichkeiten ihrer subversiven Unterwanderung im feministischen Kino.

3 Die verkörperte und entkörperte Stimme im Film

Die Bezeichnung des Films als *audiovisuelles* Medium stellt seine multimodale Prägung durch Ton und Bild in den Vordergrund. Deren Verständnis als individuelle modale Ebenen, wie es insbesondere die semiotisch geprägte Multimodalitätsforschung (z.B. Bateman/Schmidt 2012; Forceville 2009; Kress/van Leeuwen 2001) vertritt, legt eine separate Analyse und ein additives Zusammenspiel im Prozess der Bedeutungsherstellung nahe. Tatsächlich aber vollzieht sich der Prozess des In-Einklang-Bringens von Bild- und Tonspur in der Zuschauer*innenwahrnehmung so selbstverständlich, dass die bewusste Zergliederung im Prozess des Filmesehens schwierig ist. Seit der Einführung und raschen Verbreitung des Tonfilms in den 1920er Jahren änderten sich Seh- und Hörgewohnheiten im Kino von zuvor sichtbar getrenntem Bild und Ton² hin zur vermeintlich ganzheitlich sinnlich erfahrbaren Filmwelt.

Filmtheoretische Reflexionen, beispielsweise von Rick Altman oder Michel Chion, hinterfragen diese Selbstevidenz kritisch, ebenso wie die damit einhergehende implizite Unterordnung der Ton- unter die Bildebene: »[T]he concerns of the sound track have remained excluded from the nodal points of film criticism« (Altman 1980a: 3f.). Chion zufolge verleiht der Ton dem Bild einen expressiven und informativen Mehrwert, »with which a sound enriches a given image so as to create the definite impression [...] that this information or expression ›naturally‹ comes from what is seen and is already contained in

2 Orchester, Klaviere oder Pianolas begleiteten im Vorführraum das auf der Leinwand ausgestrahlte Geschehen.

the image itself»³ (Chion 1994: 5). Dieser Eindruck sei jedoch falsch, so Altman. Er bezeichnet ihn gar als eine grundsätzliche Lüge und spricht vom Kino als einem Bauchredner (Altman 1980b). Dies betrifft sowohl realistisch anmutende Fusionen wie die von sich bewegenden Lippen und hörbaren Worten, als auch nicht reale Situationen, wie beispielsweise die Faustkämpfen unterlegten Geräusche. Im Zuge dieser Offenlegung des Konstruktionscharakters von zusammenwirkendem Ton und Bild wird ersterem eine mediale Spezifik attestiert, die sich nicht im Status eines bloßen Anhängsels oder einer akustischen Dopplung visueller Inhalte erschöpft. Laut Chion ist insbesondere die Stimme im Film gegenüber sämtlichen anderen Tönen und Klängen privilegiert. Der Film ist für ihn »a vococentric or, more precisely, a verbocentric phenomenon« (Chion 1994: 5). Mit dieser Dominanz der Stimme bzw. des gesprochenen Worts komme er menschlichem kommunikativen Verhalten gleich. Der Einzug des Tons in den Film geht folglich mit einer körperlichen Dimension einher. Er führt zum lebendigen Eindruck eines belebten und beseelten, eines vollständigen und vollwertigen (menschlichen) Körpers.

Die wiederhergestellte Vollständigkeit eines sichtbaren und hörbaren Stimm-Körpers durch das Aufkommen des Tonfilms ist für die Filmphilosophie und feministische Filmwissenschaft jedoch ein Phantasma; und dies nicht lediglich mit Blick auf transnationales Kino und Synchronisation. Kritisch stellt Mary Ann Doane heraus: »The addition of sound to the cinema introduces the possibility of representing a *fuller (and organically unified) body*, and of confirming the status of speech as an individual property right« (Doane 1980: 34, Herv. DH). Mit dieser anthropozentrischen Engführung, die mithin aus Sicht der feministischen Filmtheorie und Filmkritik vergeschlechtlicht, d.h. männlich geprägt ist, eröffnet sich Raum für kritische Hinterfragung, auch aus dekolonialer Perspektive. Feministische, psychoanalytisch geprägte Annäherungen an den Film waren lange in der binären Geschlechterlogik von (weißem) Mann und (weißer) Frau gefangen, mit einem klaren Fokus auf privilegierter männlicher Subjektivierung und viktimisierter weiblicher Objektivierung. Einschlägig ist hier Laura Mulveys Essay »Visual Pleasure and Narrative Cinema« (1975), in dem sie institutionalisierte Formen und Verhal-

3 Chion (1994) hat hierfür den Begriff der *Synchrese* geprägt, über den er den Aspekt der Synchronizität und der Synthese zusammenbringt: »Synchresis [...] is the spontaneous and irresistible weld produced between a particular auditory phenomenon and visual phenomenon when they occur at the same time« (Chion 1994: 63).

tensmuster einer patriarchalen Gesellschaft in einem männlichen Blickregime (*male gaze*) im Hollywoodfilm widergespiegelt sieht.

Der binären Logik von Mann versus Frau ist auch Kaja Silverman ver-schrieben. Sie richtet ihren Blick jedoch nicht auf die männliche Schaulust, sondern auf die Stimme. Ihr zufolge fungiert die Frau im Mainstream-Kino zum Zweck des männlichen Ganzheitsempfindens als Projektionsfläche für psychisch empfundenen Mangel und Kastration, was sich nicht nur visu-ell, sondern auch stimmlich manifestiere (Ingelfinger/Penk Witt 2004: 18). Das männliche Subjekt nimmt damit die Rolle des sprechenden Subjekts im faktischen wie fiktionalen Diskurs ein, während das weibliche Subjekt ein »unzulässiges, verhandeltes, fügsames Sprechen« enagiert bzw. vollzieht (Silverman 2016: 72), keine Stimmautorität besitzt und vom Diskurs ausge-schlossen bleibt. Für Silverman ist dieser hierarchisierte Dualismus eng mit dem Status des Körpers verbunden: Die privilegierte und mächtige männliche Stimme erscheint häufig als körperloses Voiceover aus dem Off und nimmt ihren Platz im Symbolisch-Abstrakten von Sprache und Diskurs ein. Demge-genüber ist die weibliche Stimme an einen konkreten und sichtbaren Körper innerhalb der Story gebunden, mit dem sie am bedeutungsvollen Diskurs keinen Anteil hat, sondern sich »mit unzuverlässigem, verhandeltem oder fügsamem Sprechen« (Silverman 2016: 72) bemerkbar macht.⁴

Was die Gegenüberstellung einer entkörpernten unsichtbaren (männli-chen) Stimme gegenüber einer qua Synchronizität verkörpernten (weiblichen) Stimme impliziert, lässt sich mit Michel Chions Ausführungen zum *Akus-maître* greifbar machen. Eine Stimme, die, bevor sie erklingt, noch nicht visualisiert wurde, wird in der Filmwissenschaft *Akusmaître* genannt: »[M]an [hat] es mit einem Wesen ganz besonderer Art zu tun, einem sprechenden und handelnden Schatten« (Chion 1996: 51). Aus seiner Unsichtbarkeit zieht dieser Schatten seine Kraft: sein Allwissen, seine Allmacht und Magie (Chion 1996: 54). Mit Verweis auf die Figur des Zauberers von Oz aus dem gleichna-migen Film argumentiert Chion, dass infolge von Deakusmatisierung jede zuvor unsichtbare Stimme qua körperlicher Festschreibung ihrer Magie und Macht beraubt und menschlich wird. Als »Inkarnation der Stimme« macht die

4 Mulvey wurde bezüglich ihrer Übernahme der binären biologischen Kategorisierung von *sex* in Männlich und Weiblich sowie der Nichtberücksichtigung von Klasse, *race*, Sexualität kritisiert. Silverman übernimmt in ihrem Konzept der weiblichen Stimme im Kino beide problematischen Aspekte und reproduziert damit das binäre Geschlechterstereotyp ebenso wie das eurozentristische »einheitliche weiße weibliche« Subjekt.

Deakusmatisierung »das akusmatische Wesen sterblich [...], indem sie ihm einen Ort zuweist und sagt: ›Hier ist dein Körper, du wirst dich an einem Ort befinden und nicht woanders«⁵ (Chion 1996: 57).

Diese zuordnende, vergeschlechtlichte Fixierung der Stimme kritisiert die feministische Filmwissenschaft der 1970er und 1980er Jahre insbesondere mit Blick auf das Phantasma eines »bruchlosen Abgleich[s] von menschlicher Gestalt und menschlicher Stimme [...] – d.h. [...] die Repräsentation eines homogenen, denkenden Subjekts, bei dem äußere Erscheinung und Inneres kongruent sind« (Silverman 2016: 73). Silverman weiß den Konstruktionscharakter der Zuordnung von Stimme und (vergeschlechtlichtem) Körper jedoch konstruktiv zu wenden und Film als metareflexiven Raum feministisch zu nutzen. Deshalb spricht sie sich, z.B. durch Mehrfach-Zuordnungen oder akusmatische Stimmen für weiblich gelesene Stimmen, für eine offensive Entkopplung von Stimme und Körper im feministischen Kino aus:

Damit wäre die Möglichkeit eröffnet, dass die Frau am phallischen Diskurs teilhat, wodurch sie der Befragung zu ihrem Ort, ihrer Zeit und ihren Begehren entkäme, die sie beständig sicher-stellt. [...] In der Tat: Das weibliche Subjekt derart zu ent-körpern, würde jede ihrer früheren Konzeptualisierungen, mittels derer wir Kenntnis von ihr hatten, in Frage stellen, da sie gerade als Körper konstruiert ist. (Silverman 2016: 77)

Dem Plädoyer für die Befreiung von Vereindeutigung, Lesbarkeit und *male gaze* durch die Entkörperung der weiblichen Stimme scheint Jonzes Film *HER* mit der Figur der Samantha nachzukommen. Obwohl als weibliche Stimme – gesprochen von Scarlett Johansson – lesbar, bleibt Samantha in ihrer Körperlosigkeit im Hinblick auf ihren Status (als Maschine), ihr Gender, ihre Sexualität etc. nie ganz fassbar und anschaulich. Ein wesentlicher Grund scheint die rein stimmliche Präsenz ihrer Figur eines Operating Systems zu sein. Laut Marie-Luise Angerer (2015: 57) hat sich der filmische Diskurs über das Verhältnis von Mensch und Maschine bislang primär auf das Interface zwischen beiden konzentriert, »also auf das, was man üblicherweise als Kommunikationskanal bezeichnet, oder aber auf das Outfit der Maschine: »je menschenähnlicher, umso besser«. Im Falle von *HER* ist diese Schnittstelle zwischen Mensch und

5 Dem widerspricht Dolan (2007: 95) entschieden: »Letztendlich gibt es so etwas wie Deakusmatisierung nicht. Der Ursprung der Stimme ist niemals zu sehen, sie entspringt einem geheimen, strukturell verborgenen Inneren und kann nie dem entsprechen, was wir zu sehen bekommen.«

Maschine unsichtbar und gleichzeitig erkennbar prekär geworden. Obendrein kommt zum Mangel einer evidenten visuellen noch der einer akustischen Unterscheidbarkeit, weil das stimmliche Interface im Gegensatz zu der »unpersönlichen [...], [...] mechanisch produzierten Stimme« (Dolar 2007: 34) allzu menschlich klingt.

Bietet das Sichtbare keinerlei körperlichen Anhaltspunkt, so gewinnt die Stimme in ihrer Materialität und Sinnlichkeit, als »Tongeste« (Balázs 1972: 210) mit Klang-, Timbre- und Rhythmusqualitäten, einen herausgehobenen, autonomen Charakter. Der Filmtheoretiker Bela Balázs (1972: 127) differenziert mit diesem Begriff klar den »akustisch-sinnlichen Eindruck« von der allzu häufigen Engführung der Stimme auf den Inhalt, als Rede. Als Ausdrucksphänomen nimmt sie eine Zwischen(leibliche)-Stellung ein: »in between body and language, in between biology and culture, in between inside and outside, in between subject and other, in between mere sound or noise and meaningful articulation« (Shaviro 2006). Mit dieser Entfaltung eines Zwischenraums, der etablierte Kategorien überschreitet und aufbricht, eignet sowohl die Stimme im Allgemeinen als auch Samantha im Besonderen Entgrenzungspotenzial an. Dies wird in der folgenden Analyse anhand der Aspekte (1) einer zusammengesetzten Körperlichkeit und subjektiven Vielheit, (2) der Überschreitung vergeschlechtlicher Körperlichkeit und vergeschlechtlichten Begehrens, und (3) eines feministischen Nonkonformismus am Beispiel ausgewählter Szenen des Films *HER* ausgeführt.

4 Grenzgänge, Überschreitungen, Ambivalenzen in *HER* (2013)

Spike Jonzes Film spielt das inzwischen längst Alltag gewordene Onlinedating futuristisch auf der nächsten Stufe durch, indem die Technologie nicht das Mittel ist, um Leute miteinander in Kontakt zu bringen, sondern selbst zum begehrten Gegenüber wird. Bei den Oscars 2014 gewann der Regisseur eine Trophäe für das beste Originaldrehbuch. Die Nominierung von Scarlett Johansson bei den Golden Globes als beste Nebendarstellerin für ihre Rolle als Samantha wurde hingegen aus einem interessanten (und äußerst »sprechenden«) Grund abgelehnt: Sie trete im Film körperlich (sichtbar) nicht in Erscheinung; ihre Präsenz sei rein stimmlich (Prüßmeier 2013). Das Credo der eindeutigen Zuordnung einer weiblich gelesenen Stimme zu einem sichtbaren, weiblich gelesenen Körper, das die feministische Filmwissenschaft kritisiert hatte (siehe Kap. 3), scheint nach wie vor wirkmächtig zu sein.

Bevor die Analyse anhand der oben genannten drei Punkte illustriert, auf welche Weise die Stimme Samanthas entgrenzend wirksam ist, soll im Folgenden zunächst der Inhalt von HER kurz nachgezeichnet werden.

4.1 Eine menschliche Stimme – ein künstliches Wesen

Die Geschichte des Films spielt in der nahen digitalisierten Zukunft, in der Technologien den Alltag, die Arbeit sowie die Beziehungen zu anderen organisieren: Privater und öffentlicher Raum sind miteinander verschaltet. Theodore Twombly (Joaquin Phoenix) arbeitet für ein Unternehmen, dessen Geschäftsmodell das Verfassen von Liebesbriefen für Kund*innen ist. Die Erbringung dieser Dienstleistung erfolgt technologiegestützt: Theodore spricht den Inhalt qua Sprachsteuerung ein; das System überführt seine gesprochenen Worte in einen handschriftlich anmutenden geschriebenen Text, der anschließend ausgedruckt und als analoger Brief versendet wird. Er lebt nach der Trennung von seiner Frau Catherine, der er nachtrauert, allein und hat allenfalls Kontakt zu seiner Nachbarin Amy.

Eines Tages wird Theodore über Werbung auf ein neuartiges Operating System (OS) aufmerksam, das er kauft und bei sich installiert. Das OS, das User*innen als Stimme durch den Alltag begleitet, nennt sich selbst Samantha (Scarlett Johansson). Es kann »nicht nur – wie die Siri-Software von Apple – verbal mit seinem Besitzer kommunizieren [...], sondern [prägt] im Lauf der Zeit eine Persönlichkeit [aus]« (Keazor/Wübbena 2015: 16). Schnell entwickelt sich das Verhältnis zwischen Theodore und Samantha über eine reine Anwendungslogik und Dienstleistungserbringung hinaus zu einer Liebesbeziehung. Diese bringt nach einer ersten Phase der Euphorie zunehmend auch Schwierigkeiten im Miteinander der beiden ans Licht. Samanthas Wunsch nach einer körperlichen Beziehung und bewusstseins-technischen Weiterentwicklung sowie differierende Vorstellungen von Verfügbarkeit und Exklusivität der Liebesbeziehung führen zu Konflikten zwischen ihr und Theodore, bis sie sich schließlich von ihm trennt.

Auf eine reine Abhängigkeits-Dystopie der Menschen von Technologien scheint HER dennoch nicht reduzierbar zu sein. Der Film frage vielmehr

nach unserem Umgang mit den durch die Technologie gegebenen Möglichkeiten, um auf diese Weise einen Blick auf unsere Befähigung dazu sowie auf das sich auf dieser Grundlage von uns ausgeprägte Verhältnis zu eben dieser Technologie zu werfen. (Keazor/Wübbena 2015: 17)

Es gehe um »das sich daraus ergebende Wechselverhältnis zwischen Mensch und Computer« (ebd.) und um beider wechselseitige Prägung, nicht nur im gemeinsamen Sprechen, sondern auch im jeweiligen Wahrnehmen und Fühlen. Ein solcher Blick eröffnet posthumanistische Entgrenzungsperspektiven auf Sprache und Mensch jenseits eines anthropozentrischen Weltbildes, eines abstrakten Sprachsystems und eines selbstidentischen Subjekts, wie sie im Folgenden aufgezeigt und entfaltet werden sollen.

4.2 Partielle Körperlichkeit und subjektive Vielheit

Indem Samantha als Stimme, aber nicht als sichtbarer Körper im Film in Erscheinung tritt, weicht ihre Figur von der verkörperten weiblichen Stimme des Mainstream-Kinos ab, die die feministische Filmwissenschaft kritisiert hat. Sie stellt auch ein zentrales Epistem westlichen Denkens infrage: das eines ganzheitlichen, mit sich selbst identischen Subjekts, bei dem Hör- und Sichtbares, Innen und Außen einander »stimmig« ergänzen. Samanthas stimmliche Präsenz überschreitet das Bild eines vollständigen und nach außen abgegrenzten menschlichen Körpers und weist der Stimme auf diese Weise einen herausgehobenen partiellen Charakter zu. Die Stimme ist aufgrund dieser Partialität allerdings nicht defizitär verfasst, sondern »funktioniert« – weil sie eben nicht in das Bild einer körperlichen Überordnung eingepasst ist – autonom. Der Philosoph Slavoj Žižek (1996: 92) stellt eine genuine Zusammengehörigkeit einer Stimme zu »ihrem« Körper grundsätzlich infrage: »The voice acquires a spectral autonomy, it never quite belongs to the body we see, so that even when we see a living person talking, there is always some degree of ventriloquism at work«.

Ein solches Konzept der Stimme schließt an Gilles Deleuzes und Felix Guattaris organlosen Körper an, der im Sinne eines zusammengesetzten Puzzles verstanden wird:

Deshalb sind die Partialobjekte nicht Ausdruck eines zerstückelten, zerrissenen Organismus, der eine – nun zerstörte – Einheit oder die aus einem Ganzen befreiten Teile voraussetzte; der organlose Körper ist nicht Ausdruck eines wieder geleimten oder »ent-differenzierten« Organismus, der seine eigenen Teil überstiege. In letzter Instanz sind die Organ-Objekte und der organlose Körper ein und dieselbe Sache, eine von der Schizo-Analyse als solche begriffene Vielheit. *Die Partialobjekte bilden die unmittelbaren Kräfte des organlosen Körpers, und der organlose Körper die reine Materie der Partialobjekte.*

Der organlose Körper ist die Materie, die immer unter diesem oder jenem Intensitätsgrad den Raum erfüllt, und die Partialobjekte machen diese Grade aus, diese intensiven Teile, die ausgehend von der Materie als Nullintensität das Reale im Raum hervorbringen. (Deleuze/Guattari 1977: 421f.)

Bezogen auf ein Partialobjekt ›Stimme‹ widersetzt sich dieser organlose Körper dem Gesamtgefüge eines vorvorhandenen selbstidentischen, bedeutungsvollen, sozialen Körpers (bzw. eines Subjekts). Samantha ist ebenso wenig ein eindeutig weiblich gelesener sichtbarer Körper wie sie eindeutig Mensch oder Maschine ist. In ihrem stimmlich interaktiven und sozialen Verhalten mutet sie menschenartig an. Sie stimmt sich affektiv auf ihre Gesprächspartner*innen ein, drückt Emotionen aus, entwickelt das Gespräch inhaltlich weiter, reflektiert über sich und andere usw. Einen menschlichen Körper aus Fleisch und Blut, der visuell wahrnehmbar ist, hat sie jedoch nicht. Der tragbare Screen mit Kamera, der als Interface, unter anderem für visuell bildlichen Austausch, zwischen ihr und Theodore fungiert, bringt ihr ein technisches Antlitz bei, in dem sie jedoch nicht vollumfänglich aufgeht.⁶

In diesem Sinne ist Deleuzes und Guattaris organloser Körper nicht als äußerlich sichtbare Umgrenzung Samanthas zu fassen. Als Materialisierung der spezifischen Qualitäten und Intensitäten von Samanthas Stimme ist er dynamisch und komplex – er ›gehört‹ ihr nicht, sondern ist untrennbar verbunden mit den situativen sozialen und interaktiven Bedingungen sowie Aktivitäten des Raumes, den er erzeugt und füllt. Ein solches offenes und reziprok-responsives Körperverständnis schließt an Leiblichkeitskonzepte der Phänomenologie (Merleau-Ponty 1966) und Philosophischen Anthropologie (Plessner 2016) an, die auch dezidiert auf die Stimme hin gedacht werden (z.B. Waldenfels 2010). Weder dient diese sich als bloßes Verlautbarungsmedium dem sinnvollen Wort, noch einem menschlichen Geist oder Fühlen an. Sie greift aus auf ihre Umgebung, erfasst ihre Gesprächspartner*innen, verwickelt sie ins Sprechen, berauscht sich an der Begegnung mit anderen und mit sich selbst.

6 Wenn sie beispielsweise im Zuge eines Updates nicht ›gefunden‹ werden kann, ist das materielle Interface zwar ›da‹, Samanthas Stimme aber antwortet nicht und wirft den sie vergeblich rufenden Theodore auf seine Einsamkeit zurück (siehe Kap. 4.4).

Abb. 1: Ganzkörperliches Erfasstsein bei Theodore in der stimmlichen Interaktion mit Samantha

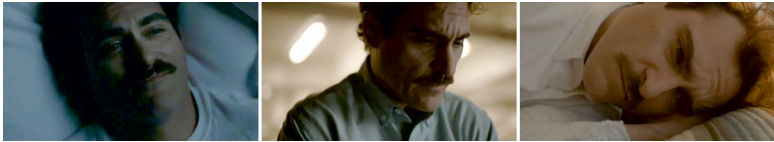


Für den Philosophen Helmuth Plessner eröffnet die Stimme eine ebensolche wechselseitige Einverleibung: »Wie das Gesicht mit seiner unverkennbaren Physiognomie, so ist die Stimme ursprünglicher Resonanzboden des Ausdrucks, für den Menschen sein: Organ. In und mit ihm schwingt er aus und ergreift er andere, wie er selbst gestimmt und ergriffen ist« (Plessner 2016: 251). Liegt im ereignishaften stimmlichen Miteinander zwischen Samantha und Theodore – vor aller zuordnenden Kategorisierung und Kausallogik – nicht genau ein solcher Fall vor? Die Stimme als zwischenleibliches Ausdrucksphänomen lässt beide auf verkörperte Weise einander begegnen (siehe Abb. 1: 1. Still links oben im lockeren Plaudern am Arbeitsplatz, 2. Still rechts oben im tiefgründigen Sprechen, 3. Still links unten im Moment von Auseinandersetzung, 4. Still rechts unten im gedankenverlorenen Zuhören): »One feels the other in one's own body, albeit in a manner of feeling that mostly remains implicit and non-thematic as such« (Fuchs 2014: 157).

Möglicherweise ist die Stimme in ihrer durchdringenden und ergreifenden raumzeitlichen Präsenz für diese Form des verkörperten Verflochtenseins gerade paradigmatisch. Samantha kommt Theodore in seinem Ohr ganz nahe, ist quasi in seinem Kopf: »Indem sie das Gegenüber umhüllt und einnimmt, verändert sie sich vom Ausdruck zur Wahrnehmungsform« (Greifenstein 2020: 88). Sie ist damit keine reine Selbstbekundung eines Subjekts im Außen, sondern findet als gelebte Erfahrung Eingang in einen (interaktiv, interaffektiv und intersubjektiv) erweiterten Körper (Froese/Fuchs 2012). In

ihrer medialen Einbettung greift die Stimme aber auch auf die Zuschauer*innen aus. Es ist es das Kontinuitätsprinzip des Films, das »eine tonale Zuschaueradressierung [prägt], welche Stabilität und permanente Zuge wandtheit verheißt« (Greifenstein 2020: 91). Indem Samanthas Stimme über den Film hinweg als »Tongroßaufnahme« (Greifenstein 2020: 89) gegenüber Umgebungs- oder Hintergrundgeräuschen im Vordergrund steht, wird sie besonders intim wahrgenommen und evoziert den Eindruck einer nahen, anwesenden Körperlichkeit (Abb. 2). Diese Tongroßaufnahmen gehen oft mit bildlichen Nahaufnahmen von Theodore als ihrem Gesprächspartner einher, wodurch der Eindruck von Nähe und Intimität verstärkt und als ganzheitliches Wahrnehmungserleben entfaltet wird (siehe Abb. 2: 1. Still im Moment stimmlichen Begehrens; 2. Still im Moment verletzender Auseinandersetzung; 3. Still im Moment der Trennung).

Abb. 2: Erfahrung von Präsenz und Intimität durch Verknüpfung von Ton- und Bildgroßaufnahme



In dieser Hinsicht wird Samantha nicht nur im Film zusehends zu einem Teil von Theodore. Auch die Zuschauer*innen machen die Erfahrung einer sie über die Zeit zusehends umgebenden Stimme. Die Frage, wem diese Stimme gehört, vermag aus einer Subjekt-Objekt-Logik nicht diesem Phänomen auf die Spur zu kommen. Aus einer Perspektive, die die Stimme in ihrer Ereignishaftigkeit, Zeiträumlichkeit und Leibkörperlichkeit⁷ in den Blick nimmt, eröffnen sich hingegen aufschlussreiche Zwischenräume und Interferenzen abseits kategorialer sozialer, philosophischer oder sprachwissenschaftlicher Klassifikationen, wie der folgende Abschnitt zeigt.

7 »Die Leibkörperlichkeit der Stimme widersetzt sich einem psychophysischen Dualismus, der den Bedeutungs- und Ausdrucksgehalt von seinem körperlichen Substrat abtrennt [...]. Ebendeshalb hören wir nicht nur das, was die Stimme uns übermittelt, sondern auch das ihr eigene Rauschen« (Waldenfels 2010: 198f.).

4.3 Selbstfremdheit und Queerness⁸

Grundlage einer solchen entgrenzenden Perspektive ist zum einen die Abkehr von einem »Besitzdenken, das meiner eigenen Stimme, die von innen her erzeugt wird, eine fremde Stimme gegenüberstellt, die mir äußerlich bleibt, weil sie vom Anderen ausgeht« (Waldenfels 2010: 195). Im Umkehrschluss kann etwas, das mir nicht vollumfänglich gehört, und über das ich nicht uneingeschränkt verfügen kann, auch mich selbst nicht vollends greifbar machen: »Vielleicht kann eine Stimme der Person, die sie zum Ausdruck bringt, nie genau entsprechen« (Lagaay 2004: 304). Zur absoluten Repräsentation eines inneren Selbstverständnisses ist sie nicht ausreichend kontrollier- und vorhersehbar. Stimme und Subjekt gelangen also nie vollständig zur Deckung. Stattdessen bleiben sie einander immer zu einem bestimmten Grad fremd.

Diese Selbstfremdheit lässt sich mit Jacques Derridas (2013) Begriff der *différence* fassen als die dem stimmlichen Ereignis zugrundeliegende Materialität und inhärente Verschiebung⁹ abseits reiner Bedeutungsentsprechung.¹⁰ Roland Barthes (2002) hat diesen Aspekt über die *Körnung der Stimme* zu erfassen gesucht, um damit den physischen Aspekt bzw. die körperliche Einschreibung der sprechenden Person in die Stimme hervorzuheben. Bernhard Waldenfels (1995, 1999) führt Selbstfremdheit insbesondere auf die Heterogenität der Stimme zurück: »[D]a so viele andere Stimmen in sie eingeschrieben sind (die meines Vaters, meiner Mutter, meiner Kultur, meiner Subkultur etc.)«, sei es unmöglich »sich völlig mit seiner Stimme zu identifizieren« (Lagaay 2008: 176). In diesem Sinne ist Waldenfels anschlussfähig an Michail Bachtins (2000) literarisches Konzept der *Heteroglossie*, das sich der Vorstellung »der

8 Lünenborg und Maier (2013: 16) betonen: »Als politischer Begriff ist Queer keine Identitätskategorie, sondern er fungiert als eine Kritik an solchen Identitätspolitiken. Queer bezieht sich nicht nur auf ein politisches, sondern auch auf ein akademisches Projekt: Queer Theory bzw. Queer Studies.« Dementsprechend problematisiert Queer Theory »die Frage, wie wir Körper, Geschlecht und Sexualität so denken – und leben – können, dass sie nicht immer wieder an eine rigide Zwei-Geschlechter-Ordnung und die Norm der Heterosexualität rückgebunden werden« (Engel 2009: 19).

9 Im Sinne eines strukturellen Verweisens auf andere Zeichen, d.h. auf Abwesendes.

10 Wie Lagaay (2004: 296ff.) argumentiert, wurde Derrida aufgrund seiner Dekonstruktion des Phono-Logozentrismus in der westlichen Philosophie fälschlich als Gegner der Stimme aufgefasst. Seine Dekonstruktion ermögliche erst eine mediale Differenzierung von Schriftlichkeit und Stimmlichkeit und die *différence* den Blick für die Materialität von Sprache.

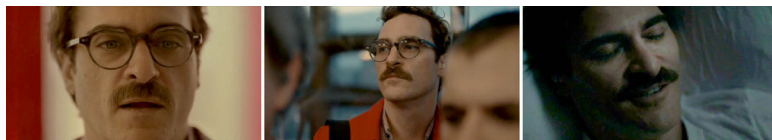
einen« Sprache und »der einen« Stimme widersetzt. Ein sprachlicher Dialog findet so nicht nur zwischen Sprecher*innen unterschiedlicher Sprachen statt, sondern auch innerhalb einer einzelnen Stimme.

Every speaker has available numerous ways of speaking that are associated by virtue of linguistic ideologies with different character types, professions, genders, social statuses [...] and so forth. In a given stretch of discourse these may be expressed by virtually any linguistic contrast, including [...] intonation and (physical) voice quality [...]. These permit speakers to claim, comment on, or disavow different identities and evaluative stances at different moments. (Keane 2000: 272)

Gerade die Anfangsszene von *HER* ist hier sprechend. Die ausgestellte affektive Stimmlichkeit und Intimität des Sprechens während des Briefe-Diktierens wird auf der Bildebene durch die Detailaufnahme von Theodores Gesicht und dessen Positionierung im Zentrum des Bildes gestützt (Abb. 3, 1. Still). Es ist nicht nur der durch die Detailaufnahme vorenthaltene Arbeitskontext, der Zuschauer*innen äußerst glaubwürdig suggeriert, hier hielte ein bewegter und verliebter Mensch eine Rede für eine geliebte Person. Es ist vor allem der Gefühlsausdruck in der Stimme – die Pausen, das Seufzen, das Schmunzeln, die progrediente Intonation – ebenso wie der präsentische Ton und die Verständlichkeit jedes einzelnen Wortes, die unmittelbar sinnlich affizierend wirken und einen Moment empfindsamer und intimer Selbstmitteilung greifbar machen. Durch sprachliche Indizien (Theodore spricht im Namen einer sich selbst als Mädchen bezeichnenden Person) als auch durch einen Schnitt wird alsbald klar, dass Theodore ein Surrogat für eine andere, weiblich gelesene Verfasserin ist. Die gefühlvolle Stimme und Hingabe, mit der er seiner Arbeit nachgeht, kontrastiert mit der Folgeszene einer monotonen und staccatohaften Befehls- tonstimme, wenn er in der Bahn über Kopfhörer sein Mailkonto checkt und Sortieranweisungen gibt (Abb. 3, 2. Still). Schließlich markiert die Szene, in der Theodore und Samantha sich zum ersten Mal im stimmlichen Begehren begegnen, einen wiederum anderen Stimmausdruck. Der Film setzt dies durch zunehmende bildliche wie stimmliche Detailaufnahmen des Hauchens, Atmens, Seufzens, des zunehmenden Auflösens der Worte in der Belegtheit der Stimmen, des gemeinsamen Stöhnens als sinnliche Erfahrung einer verkörperten Ekstase um (Abb. 3, 3. Still). Mit Derrida (1979) gesprochen, findet hier für beide

über das Sich-Im-Reden-Vernehmen (*s'entendre-parler*) die fundamentale und berausende Selbsterfahrung als lebendige Wesen statt.¹¹

Abb. 3: Polyphonie und affektive Varianz von Mensch-Maschinen-Interaktion



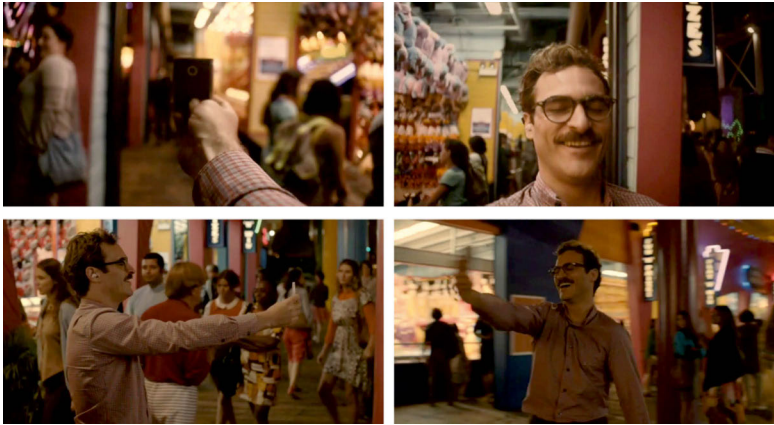
Diese Vielheit von Stimmen und Gestimmtheiten, Ausdruckskontexten und Rollen ›innerhalb‹ einer sprechenden Person sprengt einen abgeschlossenen selbstidentischen Subjektstatus – sie lässt sich nicht auf einen (Subjekt-)Nenner bringen, bleibt vielschichtig, ambivalent und multiperspektivisch.

Vielheit und Multiperspektivität der Stimme manifestieren sich auch im Hinblick auf Gender. Im Film erscheint Theodore als männlich gelesene Figur, der qua seiner zerbrochenen Ehe mit Catherine (Rooney Mara) als heterosexuell eingeführt wird. Mit dem Kauf des neuen Operating Systems und der Präsenz von Samantha in immer mehr Bereichen seines Lebens umgibt seinen Körper zunehmend eine weibliche Stimme (Angerer 2015: 64), die ihn im Plessner'schen Sinne grundlegend ergreift (siehe Kap. 4.2). Diese Einverleibung setzt HER filmisch als Erfahrung eines affektiven *attunements* (Stern 1985: 132) zwischen Theodores Körper und Samanthas Stimme um. Wenn sie ihn, der die Augen geschlossen hat, durch einen Vergnügungspark lotst und seine Schritte lenkt (Abb. 4, 2. Still oben rechts, 3. und 4. Still unten links und rechts), wird sie zur Erweiterung seiner Sinne bzw. wird er als sie umherführender Körper (Abb. 4, 1. Still) zur Erweiterung ihres Erfahrungsraums. Ihr Sprechtempo und ihre Intonation orchestrieren seinen Gang und seine Gestimmtheit. Beide sind verbunden durch und in einem gemeinsamen Bewegungs- und Affektrhythmus. Dabei entsteht ein hybrider, queerer Leibkörper¹², der ganzheitliche vergeschlechtlichte Identitäten überschreitet.

11 Die Erfahrung des Selbstvernehmens im eigenen Sprechen impliziert eine Subjektspaltung, durch die eine Person sich gleichzeitig als Sprechende**n* und Hörende**n* erlebt. Auch dies trägt zur (Selbst-)Fremdheit und Spaltung in verschiedene Subjektivitäten bei.

12 Zum Begriff des Leibkörpers siehe Plessner (2016: 238) und Fuchs (2013: 82–84).

Abb. 4: Zusammengesetzte hybride und queere Leibkörperlichkeit zwischen Samantha und Theodore



Diese hybride und flexible Verfasstheit von Subjektivität(en) wird nicht zuletzt augenfällig in der Szene, wenn Samantha und Theodore versuchen, mittels des menschlichen Surrogats Isabella miteinander körperliche Liebe zu erleben. »Auch Isabella wird mit einem Kameraauge und Ohrstöpseln ausgestattet, sodass der körperlose Computer gleichsam in ihren Körper schlüpfen und aus der Perspektive des Mädchens agieren kann« (Steinbauer 2015: 91; Abb. 5, 1. und 2. Still oben links und rechts). Die konstruierte Zuordnung von Stimme und Körper, die audiovisuell ausstellt, dass erstere nicht zu einem letzteren gehört, hinterfragt gleichermaßen das Epistem eines ganzheitlichen selbstidentischen Subjekts wie auch die kinospezifische »persuasive illusion of real people speaking real words« (Altmann 1980a: 7; Abb. 5, 3. Still unten links).¹³ Bemerkenswerterweise geht die auf diese Weise ausgestellte Körper-Stimme-Konstruktion für Samantha und das (weiblich gelesene) Surrogat auf, nicht jedoch für Theodore. Das filmisch ausgestellte sich überlagernde Stöhnen von Isabella und Samantha (und damit die Queerness der Stimme) sowie Samanthas Wunsch, ihm in die Augen zu sehen,

13 Die von Altman genutzte Metapher des Kinos als Bauchredner wird in der genannten Szene vergegenwärtigt. Samantha spricht; Isabella schweigt und bewegt auch ihre Lippen nicht. »The ventriloquist must avoid moving his own lips, thus disguising the true source of the sound [...]« (Altmann 1980b: 77).

während er ihr sagt, dass er sie liebt, fügt sich für ihn offenbar nicht stimmig identitär, so dass er den Versuch abbricht (Abb. 5, 4. Still unten rechts). Theodore und das etablierte westliche Identitätskonzept kommen hier an ihre Grenzen, während Samantha sie (posthumanistisch) überschreitet.

Abb. 5: Samantha's konstruierter (Leih-)Körper und ausgestellte ›Unstimmigkeit‹ und Theodores Verweigerung



4.4 Nonkonformismus und Selbstermächtigung

Ihr entgrenzendes Potenzial entfaltet Samantha auch in ihrer Beziehung zu Theodore. Als sie sich nach einem Software-Update seinem Zugriff entzieht und er sie über das Interface nicht erreichen kann (Abb. 6, 1. Still oben links), läuft er panisch und verzweifelt durch die Stadt (Abb. 6, 3. Still unten links). Roland Barthes beschreibt das Nicht-Antworten, die Stummheit im Diskurs der Liebe als eine destruktiv wirkende Asynchronie, die die affektive Beziehung der Liebenden ihres Zusammenspiels und ihrer Stimmigkeit beraubt: »[...] da ich umsonst spreche, ist es so, als stürbe ich. Denn das geliebte Wesen wird zur bleiernen Gestalt, zur Traumfigur, die nicht spricht, und Stummheit im Traum ist der Tod« (Barthes 1988: 205). Das gemeinsame Sprechen gerät aus dem Takt, es fällt auseinander und lässt die liebende Person – Theodore – auf sich zurückgeworfen ohne die sie in ihrem Dasein bestätigende Resonanz des geliebten Gegenübers – Samantha's Stimme – zurück (Abb. 6, 2. Still oben rechts und 4. Still unten rechts).

Abb. 6: Samanthas Entzug und Nicht-Verfügbarkeit als ausgestellte Stummheit



In dieser Szene wird augenfällig, dass hier keine weiblich gelesene konformistische Servilität eines digitalen Tools¹⁴ vorliegt. Vielmehr hat Samantha im Zuge ihrer Interaktionen mit Theodore ihre Subjektivität (weiter-)entwickelt. Sie begreift sich als Agentin, die für ihr Handeln Autonomie und Freiheit beansprucht¹⁵ und sich seinem Alleinanspruch auf sie versagt. Auf seine Frage »Are you talking with someone else right now? People, OS, whatever...« antwortet sie »8.316« (Min. 1:44:55-1:45:14). Auf seine Folgefrage »Are you in love with anybody else?« erwidert sie ihm »641« (Min. 1:45:33-1:45:58) und stürzt Theodore (und jede bürgerliche Vorstellung einer heterosexuellen Partnerschaft) in Verzweiflung.

Sie verabschiedet sich hier vom Bild der traditionellen Zweierbeziehung und trägt Ansichten von der Liebe vor, wie man sie heute in jedem besseren Polyamory-Ratgeber finden kann [...]. »I am still yours, but along the line I became many other things too. [...] The heart's not like a box that gets filled

14 Zur Problematik des Tool-Begriffs im Kontext digitaler Technologien, siehe Horst (2023).

15 Samanthas Handlungsautonomie wird auch daran evident, dass sie nicht mehr nur auf Aufforderung Dinge erledigt. Eigenmächtig und ohne Theodores Wissen wählt sie Exemplare aus seinen beruflich verfassten Liebesbriefen aus und schickt diese mit dem Ziel einer Veröffentlichung an einen Verlag. Sie »überrascht« ihn mit der Nachricht, als die Publikation bereits in die Wege geleitet ist.

up. It expands in size the more you love. I'm different from you. This doesn't make me love you any less. It actually makes me love you more. [...] I'm yours and I'm not yours.« (Wendler 2015: 73)

In dem Maße, in dem Samantha zunehmend ihre Interessen und ihr Sein eigenmächtig, selbstverantwortlich und selbstbestimmt vertritt und gestaltet, wächst Theodores Verstimmung und Unverständnis. Nach der missglückten körperlichen Vereinigung mittels Isabella kritisiert Theodore Samantha dafür, dass sie seufzt; schließlich sei sie kein Mensch, der Sauerstoff brauche (Min. 01:22:11). Während er in dieser Situation klar eine ontologische Grenze zwischen sich und Samantha zieht (die sie merklich verletzt), reagieren er, sein Kollege Paul und dessen Partnerin auf einem gemeinsamen Ausflug irritiert (Abb. 7), als Samantha ihrerseits die Grenze zwischen ihnen und sich zieht, indem sie mitteilt, sie schätze es mittlerweile keinen Körper zu haben: »I'm growing in a way that I couldn't if I had a physical form. I mean I'm not limited; I can be anywhere and everywhere simultaneously; I'm not tethered to time and space in a way that I would be if I was stuck in a body that's inevitably gonna die« (Min. 01:33:41-01:34:04).

Abb. 7: Irritation und Unverständnis gegenüber Samanthas Selbstermächtigung und Markierung ontologischer Differenz



In der selbstermächtigen Umwertung ihrer Andersartigkeit treten Samanthas Subjektivität, ihre leibliche Selbsterforschung und ihr (Selbst-)Bewusstsein als individuelles Wesen in den Vordergrund. Sie erhebt emanzipiert ihre Stimme, spricht für sich selbst. Derrida hat die Stimme als generative Instanz des Bewusstseins beschrieben, denn im »Sprechend-Sich-Selbst-Vernehmen« ruhe die Erfahrung einer grundlegenden Selbstpräsenz. »[H]earing oneself speak« [*s'entendre-parler*], as Derrida has demonstrated, is the very ker-

nel, the fundamental matrix, of experiencing oneself as a living being« (Žižek 1996: 94). Dieser intime und gleichzeitig distanzierende Moment von Selbstnähe und Selbstbezüglichkeit – die Wahrnehmung eines selbsterzeugten Klangs und die Selbstwahrnehmung im Klang – bringt Subjektivität hervor: »[S]truktural begründet [ist] kein Bewusstsein ohne die Stimme möglich [...]. Die Stimme ist das Bei-sich-sein in der Form der Universalität, das Mit-Bewusstsein (con-science). Die Stimme ist Bewusstsein« (Derrida 1979: 136f.). Dies wird auch an Samanthas selbstbewusster Abgrenzung gegenüber Theodore, Paul und Pauls Partnerin als Menschen evident. Die (Medien-)Spezifik ihrer Stimmlichkeit ist, was ihr Wesen ausmacht: Ohne feste Bindung an eine körperliche Klangquelle bzw. über sie hinausgehend greift sie räumlich und zeitlich aus, ist zugleich überall und nirgends. Über Samantha heißt es in HER: »It's not just an operating system, it's a *consciousness*« (Min. 0:10:51) und »[...] what makes me ›me‹ is my ability to grow through my experiences, so basically, in every moment I'm evolving, just like you« (Min. 0:13:59).

In der stimmlich-leiblichen Begegnung und dem Austausch mit Theodore entwickelt Samantha Ansätze eines eigenen leiblichen¹⁶ und subjektiven Bewusstseins, das sich von Theodores grundlegend unterscheidet. Sie sprengt seine (kognitiven, körperlichen, leiblichen, ...) Grenzen in ihrem kontinuierlichen Wachsen und (Neu-)Werden (»in every moment I'm evolving«) und übersteigt ihn, beispielsweise indem sie mit 8.316 anderen Interaktionsgegnern im gleichen Moment in Austausch steht und in 641 Personen, OS oder sonstige Instanzen verliebt sein kann. Hier werden nicht nur bürgerlich moralische Grenzen überschritten, sondern auch räumliche und zeitliche einer spezifisch menschlichen Leibkörperlichkeit. Als Stimme weitet sie sich multidimensional in den Raum, und Theodore kann mit dieser Dynamik alsbald nicht mehr Schritt halten. Samanthas Vergleich eines Buches im Entstehen bringt diese Prozesshaftigkeit und Entwicklungsdynamik auf den Punkt: »It's like I'm reading a book. It's a book I deeply love. But I'm reading it slowly now. So the words are really far apart and the spaces between the words are almost infinite.« Die Stimme (Samanthas) erschöpft sich nicht im sinnhaften Austausch von Wor-

16 Ob und inwiefern Samantha als (spezifische, posthumanistische) Leibkörperlichkeit zu fassen wäre, ist eine Frage, die HER über Samanthas Ansätze reflexiver Selbstpositionierung durchaus aufwirft (Versuche phänomenologischer Reflexion über die Leiblichkeit digitaler Entitäten gibt es bereits, beispielsweise in Burow 2019).

ten mit einem Gesprächsgegenüber¹⁷, sondern geht sinnlich über sie hinaus in die Tiefe. »But it's in this endless space *between the words* that I'm finding myself now. It's a place that's not of the physical world« (Min. 1:51:18-1:51:54).

Doris Kolesch fasst die Stimme unter anderem aus diesem Grund »als paradigmatische Figur der Überschreitung« (Kolesch 2003: 275). In ihrem transgressiven Potenzial verdeutlicht Samanthas Stimme in *HER* die Ambivalenz etablierter westlicher Grenzziehungen in Linguistik, Philosophie und Gesellschaft und stellt sie zur Disposition. Kolesch stellt in dieser Hinsicht die besondere Bedeutung von Dissens und Reibung im Gegensatz zu Verständigung und Konsens im stimmlichen Begegnen ins Zentrum, weil erstere den Diskurs am Laufen halten, während letztere ihn zusammenbrechen lassen: »Die Stimme vermittelt, interagiert zwischen Instanzen, die zusammenkommen wollen, aber nicht zusammenkommen können – eine Situation, die die Stimme im Idealfall gerade nicht behebt, sondern weitertreibt« (Kolesch 2003: 280). Damit steht die Stimme paradigmatisch für Uneindeutigkeit und Unabgeschlossenheit, für das kontinuierliche Wechselspiel und eine andauernde Aushandlung in einem nicht stillstehenden Diskurs. Inwiefern dieser Reibungscharakter erhellend für eine »Entgrenzung von Sprache« ist, wird im abschließenden Abschnitt diskutiert.

5 *HER* als posthumane Perspektive

Trotz – oder gerade aufgrund – seines eindeutig künstlerischen medialen Formats mit fiktivem Inhalt besitzt *HER* als Spielfilm das Potenzial, über eine »Entgrenzung von Sprache« in der Linguistik nachzudenken. Medien sind keineswegs bloße Mittel der Kommunikation gegebener realer oder fiktiver Sachverhalte. Vielmehr fungieren sie, so Marshall McLuhans (1964) berühmte These, als technische Erweiterungen menschlicher Wahrnehmung und verändern apriorische Bedingungen des Verstehens, Urteilens und Imaginierens. Kulturwissenschaftlich gewendet reflektieren Filme – buchstäblich wie metaphorisch – historische, kulturelle oder politische Diskurse ihrer Entstehungsgegenwart. Dies bezieht sich allerdings nicht allein auf die dargestellten

17 Das ist die Entgrenzung einer strukturalistisch systemischen Linguistik, die Sprache als abstraktes System bedeutungsvoller Zeichen versteht, die im medienindifferenten Gebrauch zur informationsvermittelnden Anwendung kommen.

Sachverhalte. Vielmehr modelliert der Film als Medium mit seinen technischen Möglichkeiten und Gestaltungsmitteln die menschliche Erfahrung und bringt auf diese Weise neue Sinnhorizonte, Räume und Zeitlichkeiten hervor.

Vor diesem Hintergrund ist HER nicht lediglich ein Science-Fiction-Märchen über eine unmögliche Liebesgeschichte zwischen ›dem Menschen‹ Theodore und ›der Maschine‹ Samantha. HER entwirft das Szenario einer kommunikativen Transgression zwischen Mensch/en und Maschine/n in einer (möglicherweise) nicht allzu fernen Zukunft. Der Film setzt dies auf ästhetische Weise um, als konkrete sinnliche Erfahrung, für die insbesondere die Stimme eine zentrale Rolle spielt. Sie markiert das unsichtbar gewordene Interface, den Ort der Begegnung und des Austauschs, den Begehrensraum, in dem die Grenze zwischen Mensch und Maschine, zwischen menschlich und nichtmenschlich brüchig wird und sich auflöst. Anstelle dieser Grenze rückt ein partieller Körper in seiner Umwelt-Bezogenheit ins Zentrum, der weder visuell noch klanglich einen verlässlichen kategorischen Unterscheidungsanhaltspunkt bietet. Zuschauer*innen durchlaufen dieses Szenario von Samanthas Stimme als lebendiges Gefühl einer Affizierung, die ganz ohne eine vorherige Differenzierung zwischen menschlich und nichtmenschlich anschaulich wird.¹⁸

Dieses leibhaftige Durchspielen einer kategorialen Entgrenzung rührt auch an etablierten linguistischen Prämissen und Konzepten:

- das Paradigma einer sprachlichen zwischenmenschlichen Wechselseitigkeit, die sich auch an der begrifflichen Spezifizierung »Mensch-Maschine-Interaktion« zeigt,
- das Verständnis einer Subjekt-Objekt-Trennung zwischen Gesprächspartner*innen,
- die Fokussierung auf Wort- und Textsemantik unter weitgehender Ausklammerung körperlicher Ausdrucksmodalitäten im Allgemeinen und stimmlicher Medialität im Besonderen,
- die selbstverständliche Zuschreibung einer Stimme zu einem vergeschlechtlichten Körper und einem selbstidentischen Subjekt sowie
- die Priorisierung von Empathie, Eindeutigkeit und Verständigung in Bezug auf die Funktionalität von Interaktion.

18 Die Zuschreibung von Liebesgefühlen auf Theodore gegenüber Samantha qua narrativer Logik ist hierfür nicht maßgeblich, sondern vielmehr das Ergebnis der Zuschauer*innenaffizierung.

Das im letztgenannten Aspekt adressierte Zusammenspiel von Fühlen und Verstehen greift neben der äußerlichen Grenzziehung ein zweites gängiges Unterscheidungskriterium zwischen Mensch und Maschine auf: »In the *empathic* sense, we can only understand what has sensations and feelings – and robots have no feelings« (Fuchs 2022). Was dieser Auffassung widerspricht, wird als Täuschung abqualifiziert. Laut Sharkey und Sharkey (2021: 311), »[...] a deception can be said to have occurred in robotics if the appearance and the way that a robot is programmed to behave, creates, for example, the illusion that a robot is sentient, emotional, and caring or that it understands you or loves you«. HER ermöglicht durch seine Fokussierung der Stimme als sinnliches, vielschichtiges und schwer fassbares Phänomen einen Perspektivwechsel weg von der kategorialen Trennung hin zum Hybriden, zur Wechselseitigkeit und zum Prozesshaften. Fragen nach der grundsätzlichen Möglichkeit oder faktischen Echtheit der Gefühle von Samanthas Seite und von Theodore für Samantha – als filmisch hervorgebrachte fiktionale Figuren – qua ihres Existenzstatus stellen sich deshalb nachrangig bzw. gar nicht.

Insofern Phänomene des Scheins keine Wissensansprüche stellen können, fordern sie zum Glauben heraus. Die filmische Illusion zwingt mir im Glauben an das in ihr Erscheinende kein (Fehl-)Urteil über die empirische Welt auf, sondern setzt ein Verhältnis, in dem ich mich zum Gegenstand, *wie er erscheint*, also zur Weise seines Erscheinens, verhalten muss. (Koch 2006: 121; Herv. DH)

HER vergegenwärtigt und befragt, was jenseits einer anthropozentrischen Welt-, Subjekt- und systemischen Sprachvorstellung möglich, fühlbar und vorstellbar ist. Was passiert und was bedeutet es, wenn Maschinen sich weder äußerlich noch im interaktiven Verhalten oder fühlenden Miteinander trennscharf als nichtmenschlich ausweisen? Die Nähe, die Empfindungen, die Interaktionen im Film können insofern als real gelten, als sie als Wahrnehmungserleben der Zuschauer*innen sinnlich anschaulich werden und Samantha und Theodore wie auch ihr Miteinander prägen. Insofern wäre Fuchs (2014) über diese fiktionale Beziehung nicht Recht zu geben, dass es sich um ein Fehlen von Interaffektivität und lediglich um fiktionale Empathie handelt. In ihren Stimmen greifen Samantha und Theodore aufeinander und auf die Zuschauer*innen aus, begegnen sich und ihnen leiblich und werden einander fühlend teilhaftig. Es liegt kein Betrugsfall vor. »Die Maschine« kann fühlen, und »der Mensch« kann sie lieben. Damit weist das Szenario, das HER

entwirft, über sich selbst als fiktionale Geschichte hinaus und hinterfragt etablierte soziale, philosophische und linguistische Paradigmen zugunsten neuer Perspektiven, beispielsweise

[d]ass Emotionen keine urwüchsigen Regungen des Selbst mehr sind oder vielleicht noch nie waren. Wer einem Computerbetriebssystem die Fähigkeit zuspricht, Gefühle zu haben oder auszulösen, muss anerkennen, dass Emotionen nicht hier drinnen geboren werden, sondern auf eine noch zu bestimmende Weise uns von dort draußen aufsuchen.¹⁹ (Wendler 2015: 67)

Indem der Film dem bruchlosen Abgleich von menschlicher Gestalt und menschlicher Stimme – und damit dem homogenen, selbstidentischen menschlichen Subjekt – eine klare Absage erteilt, eröffnet er Raum für Zwischentöne, Körnungen, Ambivalenzen und Unverfügbarkeiten interaktiver, kommunikativer Phänomene sowie an ihnen Beteiligter und, in letzter Konsequenz, von Sprache überhaupt.

Literaturangaben

- Altman, Rick (1980a): Introduction. In: *Yale French Studies* 60, 3–15.
- Altman, Rick (1980b): Moving Lips: Cinema as Ventriloquism. In: *Yale French Studies* 60, 67–79.
- Angerer, Marie-Luise (2015): HER Master's Voice. Eine akusmatische Liebesbeziehung von Spike Jonze. In: Wende, Johannes (Hg.), *Spike Jonze*, München: Edition Text + Kritik, 57–66.
- Aristoteles (2006): *Über die Seele*. Hg. v. Hellmut Flashar, Berlin: Akademie Verlag.
- Bachtin, Michail M. (2000): *The dialogic imagination. Four essays*, Austin: University of Texas Press.
- Balázs, Béla (1972): *Der Film. Werden und Wesen einer neuen Kunst*, Wien: Globus.
- Barthes, Roland (2002): *Die Körnung der Stimme. Interviews 1962–1980*, Frankfurt a.M.: Suhrkamp.
- Barthes, Roland (1988): *Fragmente einer Sprache der Liebe*, Frankfurt a.M.: Suhrkamp.

19 Präzisierung wäre hinzuzufügen, dass das Draußen von Emotionen oder affektiven Erfahrungen ihre Kontext- bzw. Umweltgebundenheit ist.

- Bateman, John A./Karl-Heinrich Schmidt (2012): *Multimodal film analysis. How films mean*, New York: Routledge.
- Bucholtz, Mary/Kira Hall (2016): Embodied sociolinguistics. In: Coupland, Nikolas (Hg.), *Sociolinguistics. Theoretical debates*, Cambridge/New York: Cambridge University Press, 173–198.
- Bühler, Karl (1933): *Ausdruckstheorie. Das System an der Geschichte aufgezeigt*, Jena: Gustav Fischer.
- Burow, Johannes F. (2019): The Next Step. Können digitale Entitäten als eine neue Stufe im Sinne der Philosophischen Anthropologie Helmuth Plessners verstanden werden? In: Burow, Johannes F./Lou-Janna Daniels/Anna-Lena Kaiser/Clemens Klinkhamer/Josefine Kulbatzki/Yannick Schütte/Anna Henkel (Hg.), *Mensch und Welt im Zeichen der Digitalisierung. Perspektiven der Philosophischen Anthropologie Plessners*, Baden-Baden: Nomos, 209–228.
- Canetti, Elias (1975): Akustische Maske und Maskensprung – Materialien zu einer Theorie des Dramas. In: *Neue Deutsche Hefte* 3, 497–516.
- Cassirer, Ernst (2004): Zur Objektivität der Ausdrucksfunktion. In: Ders., *Kulturphilosophie. Vorlesungen und Vorträge 1929–1941* (Gesammelte Werke, Nachgelassene Manuskripte und Texte, Band 5), Hamburg: Meiner, 105–200.
- Chion, Michel (1994): *Audio-vision. Sound on screen*, New York: Columbia Univ. Press.
- Chion, Michel (1996): Das akusmatische Wesen. Magie und Kraft der Stimme im Kino. In: *Meteor* 6, 48–58.
- Deleuze, Gilles/Félix Guattari (1977): *Anti-Ödipus. Kapitalismus und Schizophrenie I*, Frankfurt a.M.: Suhrkamp.
- Deppermann, Arnulf (2014): Konversationsanalyse: Elementare Interaktionsstrukturen am Beispiel der Bundespressekonferenz. In: Staffeldt, Sven/Jörg Hagemann (Hg.), *Pragmatiktheorien*, Tübingen: Stauffenburg Verlag, 19–47.
- Derrida, Jacques (1979): *Die Stimme und das Phänomen. Ein Essay über das Problem des Zeichens in der Philosophie Husserls*, Frankfurt a.M.: Suhrkamp.
- Derrida, Jacques (2013): *Die différance. Ausgewählte Texte*. Hg. v. Peter Engelmann, Stuttgart: Reclam.
- Doane, Mary Ann (1980): The Voice in the Cinema: The Articulation of Body and Space. In: *Yale French Studies* 60, 33–50.

- Dolar, Mladen (2002): Das Objekt Stimme. In: Kittler, Friedrich/Thomas Macho/Sigrid Weigel (Hg.), *Zwischen Rauschen und Offenbarung. Zur Kultur- und Mediengeschichte der Stimme*, Berlin/Boston: De Gruyter, 233–256.
- Dolar, Mladen (2007): *His master's voice. Eine Theorie der Stimme*, Frankfurt a.M.: Suhrkamp.
- Dworschak, Manfred (2017): Künstliche Intelligenz: Der geniale Stimmenklau. In: *DER SPIEGEL*, 02.06.2017. Abrufbar unter <https://www.spiegel.de/spiegel/menschliche-stimmen-kuenstliche-intelligenz-macht-stimmenklau-moeglich-a-1149915.html>
- Elleström, Lars (Hg.) (2010): *Media Borders, Multimodality and Intermediality*, London: Palgrave Macmillan.
- Engel, Antke (2009): *Bilder von Sexualität und Ökonomie. Queere kulturelle Politiken im Neoliberalismus*, Bielefeld: transcript.
- Forceville, Charles (2009): Non-verbal and multimodal metaphor in a cognitivist framework: Agendas for research. In: Forceville, Charles/Eduardo Urios-Aparisi (Hg.), *Multimodal metaphor*, Berlin/New York: Mouton de Gruyter, 19–44.
- Frøese, Tom; Fuchs, Thomas (2012): The extended body: a case study in the neurophenomenology of social interaction. In: *Phenom Cogn Sci* 11:2, 205–235.
- Fuchs, Thomas (2013): Zwischen Leib und Körper. In: Hähnel, Martin/Marcus Knaup (Hg.), *Leib und Leben. Perspektiven für eine neue Kultur der Körperlichkeit*, Darmstadt: wbg Academic, 82–93.
- Fuchs, Thomas (2014): The Virtual Other. Empathy in the Age of Virtuality. In: *Journal of Consciousness Studies* 21:5-6, 152–173.
- Fuchs, Thomas (2022): Understanding Sophia? On human interaction with artificial agents. In: *Phenom Cogn Sci* 23:2, 21–42.
- Goffman, Erving (2005): *Rede-Weisen. Formen der Kommunikation in sozialen Situationen*. Hg. v. Knoblauch, Hubert/Christine Leuenberger/Bernt Schnettler, Konstanz: UVK-Verl.
- Greifenstein, Sarah (2020): *Tempi der Bewegung – Modi des Gefühls. Expressivität, heitere Affekte und die Screwball Comedy*, Berlin/Boston: De Gruyter.
- Horst, Dorothea (2020): Patterns ›We‹ Think By? Critical Cognitive Linguistics Between Language System and Language Use. In: *Yearbook of the German Cognitive Linguistics Association* 8:1, 67–82.
- Horst, Dorothea (2023): Digitale Werkzeugkritik als essentielle Bedingung einer innovativen digitalen Lehr-Lernkultur. In: Mrohs, Lorenz/Miriam Hess/Konstantin Lindner/Julia Schlüter/Sven Overhage (Hg.), *Digitalisie-*

- nung in der Hochschullehre. *Perspektiven und Gestaltungsoptionen*, Bamberg: University of Bamberg Press, 47–64.
- Horst, Dorothea (i. E.): Re-Thinking Embodiment in Figuration: Lived Experience or Cognitive Mechanisms? In: Colston, Herbert (Hg.), *What Makes a Figure? Rethinking Figuration*, Amsterdam: John Benjamins.
- Ingelfinger, Antonia/Meike Penkwitt (2004): Einleitung: Screening Gender. Geschlechterkonstruktionen im Kinofilm. In: *Freiburger FrauenStudien. Zeitschrift für Interdisziplinäre Frauenforschung* 14, 11–37.
- Jewitt, Carey (Hg.) (2014): *The Routledge Handbook of Multimodal Analysis*, Milton Park: Routledge.
- Keane, Webb (2000): Voice. In: *Journal of Linguistic Anthropology* 9:1-2, 271–273.
- Keazor, Henry/Thorsten Wübbena (2015): Spike Jonze: Der Versuch einer Spektralanalyse. In: Wende, Johannes (Hg.), *Spike Jonze*, München: Edition Text + Kritik, 6–22.
- Klug, Nina-Maria/Hartmut Stöckl (Hg.) (2016): *Handbuch Sprache im multimodalen Kontext*, Berlin/Boston: De Gruyter.
- Koch, Gertrud (2006): Müssen wir glauben, was wir sehen? Zur filmischen Illusionsästhetik. In: *Deutsche Zeitschrift für Philosophie* 54:1, 121–130.
- Kolesch, Doris (2003): Die Spur der Stimme. Überlegungen zu einer performativen Ästhetik. In: Epping-Jäger Cornelia/Erika Linz (Hg.), *Medien/Stimmen*, Köln: DuMont, 267–281.
- Krämer, Sybille (2001): *Sprache, Sprechakt, Kommunikation. Sprachtheoretische Positionen des 20. Jahrhunderts*, Frankfurt a.M.: Suhrkamp.
- Krämer, Sybille (2005): Medienphilosophie der Stimme. In: Sandbothe, Mike/Ludwig Nagl (Hg.), *Systematische Medienphilosophie*, Berlin: Akademie Verlag, 221–238.
- Krämer, Sybille/Ekkehard König (Hg.) (2002): *Gibt es eine Sprache hinter dem Sprechen?*, Frankfurt a.M.: Suhrkamp.
- Kress, Gunther R./Theo van Leeuwen (2001): *Multimodal discourse. The modes and media of contemporary communication*, London: Bloomsbury Academic.
- Lagaay, Alice (2004): Züge und Entzüge der Stimme in der Philosophie. In: Krämer, Sybille (Hg.): *Performativität und Medialität*, München: Fink, 293–306.
- Lagaay, Alice (2008): Zwischen Klang und Stille. Gedanken zur Philosophie der Stimme. In: *Paragrana* 17:1, 168–181.
- Leblebici, Didem (2021): *Language Ideologies in Human-Machine Interaction. A Qualitative Study with Voice Assistant Users*, Masterarbeit an der Kulturwissenschaftlichen Fakultät der Europa-Universität Viadrina, Frankfurt (Oder).

- Linell, Per (2005): *The written language bias in linguistics. Its nature, origins and transformations*, London/New York: Routledge.
- Lünenborg, Margreth/Tanja Maier (2013): *Gender Media Studies. Eine Einführung*, Konstanz/München: UVK Verlagsgesellschaft mbH.
- Markl, Nina (2022): Language variation and algorithmic bias: understanding algorithmic bias in British English automatic speech recognition. In: *Proceedings of 2022 5th ACM Conference on Fairness, Accountability, and Transparency*, 521–534.
- McLuhan, Marshall (1964). *Understanding Media. The Extensions of Man*, New York: New American Library.
- Merleau-Ponty, Maurice (1966): *Phänomenologie der Wahrnehmung*, München: De Gruyter.
- Metten, Thomas (2014): *Kulturwissenschaftliche Linguistik. Entwurf einer Medientheorie der Verständigung*. Berlin/Boston: De Gruyter.
- Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva Ladewig/David McNeill/Sedinha Tessendorf (Hg.) (2013): *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction, Vol 1*, Berlin: De Gruyter Mouton.
- Mulvey, Laura (1975): Visual Pleasure and Narrative Cinema. In: *Screen* 16:3, 6–18.
- Ong, Walter J. (2016): *Oralität und Literalität. Die Technologisierung des Wortes*, Wiesbaden: Springer VS.
- Pennycook, Alastair (2018): *Posthumanist applied linguistics*, London: Routledge.
- Plessner, Helmuth (2016): *Ausdruck und menschliche Natur*, Frankfurt a.M.: Suhrkamp.
- Prüßmeier, Björg (2013): Keine Chance auf einen Golden Globe: Scarlett Johansson ist für ihre Rolle in »Her« nicht zugelassen. Abrufbar unter <http://www.filmstarts.de/nachrichten/18481634.html>.
- Quinlivan, Davina (2017): A Dark and Shiny Place. In: Whittaker, Tom/Sarah Wright (Hg.), *Locating the voice in film. Critical approaches and global practices*, New York: Oxford University Press, 295–310.
- Schmitt, Christina (2015): Embodied meaning in audio-visuals: first steps towards a notion of mode. In: Wildfeuer, Janina (Hg.), *Building Bridges for Multimodal Research. International Perspectives on Theories and Practices of Multimodal Analysis*, Frankfurt a.M.: Peter Lang, 309–325.
- Schneider, Britta (2022): Multilingualism and AI: The Regimentation of Language in the Age of Digital Capitalism. In: *Signs and Society* 10:3, 362–387.

- Schützeichel, Rainer (2011): Soziologie der Stimme. In: Keller, Reiner/Michael Meuser (Hg.), *Körperwissen*, Wiesbaden: VS-Verl. 85–104.
- Shannon, Claude Elwood/Warren Weaver (1976): *Mathematische Grundlagen in Informationstheorie*, München/Wien: Oldenbourg.
- Sharkey, Amanda/Noel Sharkey (2021): We need to talk about deception in social robotics! In: *Ethics and Information Technology* 23:3, 309–316.
- Shaviro, Steven (2006): Review of »A Voice and Nothing More«. Abrufbar unter <https://www.shaviro.com/Blog/?p=489>.
- Silverman, Kaja (2016): Die weibliche Stimme ent-körpern. In: Peters, Kathrin/Andrea Seier (Hg.), *Gender & Medien-Reader*, Zürich/Berlin: diaphanes, 71–90.
- Steinbauer, Anna (2015): Kameraauge und Ohrstöpsel. Das Verhältnis von Sehen und Hören in »Her« von Spike Jonze. In: Wende, Johannes (Hg.), *Spike Jonze*, München: Edition Text + Kritik, 82–94.
- Stern, Daniel N. (1985): *The Interpersonal World of the Infant. A View from Psychoanalysis and Developmental Psychology*, London: Routledge.
- Stöckl, Harmut (2004): *Die Sprache im Bild, das Bild in der Sprache. Zur Verknüpfung von Sprache und Bild im massenmedialen Text: Konzepte, Theorien, Analysemethoden*, Berlin/New York: De Gruyter.
- Waldenfels, Bernhard (1995): *Deutsch-französische Gedankengänge*, Frankfurt a.M.: Suhrkamp.
- Waldenfels, Bernhard (1999): *Vielstimmigkeit der Rede*, Frankfurt a.M.: Suhrkamp.
- Waldenfels, Bernhard (2010): *Sinne und Künste im Wechselspiel. Modi ästhetischer Erfahrung*, Berlin: Suhrkamp.
- Weizenbaum, Joseph (1976): *Computer power and human reason. From judgement to calculation*, San Francisco: W.H. Freeman.
- Wendler, André (2015): Medien, Netzwerk, Liebe. In: Wende, Johannes (Hg.), *Spike Jonze*, München: Edition Text + Kritik, 67–81.
- Žižek, Slavoj (1996): »I Hear You with My Eyes«; or, The Invisible Master. In: Žižek, Slavoj/Renata Salecl (Hg.), *Gaze and Voice as Love Objects*, Durham: Duke University Press, 90–126.

Filmverzeichnis

Jonze, Spike (2013). HER. USA: Annapurna Pictures, Hollywood (CA).

Dr. Dorothea Horst

Wissenschaftszentrum Berlin für Sozialforschung gGmbH

Reichpietschufer 50

10785 Berlin

dorothea.boehme@wzb.eu

<https://www.wzb.eu/de/personen/dorothea-boehme>

Sprache-Sprechen, Hören, Sehen, Fühlen

Eine medienästhetische Perspektive auf die Multimodalität des Sprechens in audiovisuellen Medien

Cornelia Müller

Abstract ›Sprache entgrenzen‹ wird im vorliegenden Beitrag als genuin multimodal verfasstes Sprache-Sprechen und als Sprachgebrauch in audiovisuellen Medien gedacht. Entgrenzung bezieht sich auf den Gegenstand der Sprachwissenschaft der nach wie vor dem Primat des Lautsprachlichen folgt, aber auch auf neue Formen der Sprachgebrauchsforschung. Er plädiert dafür, medienlinguistische Untersuchungen multimodalen Sprachgebrauchs medientheoretisch zu fundieren und skizziert einen transdisziplinären (filmwissenschaftlich-linguistischen) Vorschlag, wie Sprache-Sprechen in audiovisuellen Medien aus einer medienästhetischen Perspektive gefasst werden kann. Ausgangspunkt sind Parallelen von Sprache-Sprechen und audiovisuellen Bildern: Beide werden als Zeit- und Bewegungsfigurationen gedacht, die erst in der verkörperten Erfahrung eines Gegenübers ›Bedeutung‹ gewinnen. Im Kern plädiert der Beitrag dafür, die Multimodalität des Sprachgebrauchs als ästhetische Erfahrung zu beschreiben, die durch die Orchestrierung der audiovisuellen Medien moduliert wird. Damit gerät die Generierung, Modulierung und Mobilisierung von Gefühlen als verkörperte Erfahrung in den Blick der Multimodalitätsforschung. Der Beitrag formuliert einen theoretischen Rahmen und eine methodische Perspektive für eine kulturwissenschaftlich fundierte Kritik am Sprachgebrauch in audiovisuellen Medien.

Keywords Multimodalität; Gestik und Sprache; Temporalität; Embodiment; Ästhetische Erfahrung; Affektivität

1 Einleitung

›Sprache entgrenzen‹? Aus der Sicht einer Erforschung des Sprachgebrauchs bedeutet dies zuallererst, dass das Sprechen einer Sprache ein multimodaler, durch und durch körperlicher Vorgang ist. Es bedeutet auch, die Dichotomie verbal – nonverbal aufzugeben und Sprache als genuin multimodal verfasst anzuerkennen. Die Gestenforschung und die Gebärdensprachforschung haben in den vergangenen Jahrzehnten aufzeigen können, dass Sprache-Sprechen nicht am gesprochenen Wort festzumachen ist (Müller et al. 2013, 2014). Damit gerät die Frage nach der ›Multimodalität des Sprechens‹ in den Blick sprachwissenschaftlicher Forschung.

›Sprache entgrenzen‹ bedeutet im Kontext allseits verfügbarer audiovisueller Medien aber auch, dass die Multimodalität des Sprechens im Alltag des heutigen Sprachgebrauchs immer mehr und immer häufiger über die Bildschirme von Smartphones, Tablets oder Laptops wahrgenommen, geteilt und so zum Movens partikularer Gemeinschaftsbildungen wird, die sich in Form von Klickzahlen oder *Follower*innen* auf Social-Media-Kanälen manifestieren. Audiovisuelle Bilder von Menschen die zueinander, miteinander, übereinander sprechen, sind zum selbstverständlichen und ganz alltäglichen Bestandteil einer globalisierten, durchweg medial strukturierten Alltagswelt geworden. Familienfeste, Reisen, Shoppingtouren werden in einem nie gekannten Ausmaß audiovisuell dokumentiert, die Handykamera ist immer dabei, ob Silvesterfeuerwerk oder Überschwemmungen, es wird aufgezeichnet und oft noch im gleichen Moment geteilt.

Auch für die Sprachforschung bedeutete die Möglichkeit der audiovisuellen Aufzeichnung von natürlichen und experimentellen Situationen des Sprachgebrauchs eine mediale Revolution. Waren von Quintilian bis Wilhelm Wundt Reflektionen über Gesten und ihr Zusammenwirken mit der Sprache auf die teilnehmende Beobachtung angewiesen, machte die Erfindung des Films es möglich, die Flüchtigkeit des Sprache-Sprechens zu überwinden. Nun ließ sich eine Kommunikationssituation wieder und wieder und in unterschiedlichen Geschwindigkeiten anschauen. Das audiovisuelle Medium wurde so zum ›Mikroskop‹ der Erforschung von Körperbewegung, Kommunikation und Interaktion. Ohne audiovisuelle Dokumentationen ist die zeitgenössische Gesten- und Gebärdensprachforschung, die Erforschung von multimodaler Kommunikation und Interaktion, nicht denkbar. Sie spielt daher für die Ausweitung der Perspektive linguistischer Forschung eine entscheidende Rolle.

Sprache-Sprechen begegnet uns im Alltag und in der Forschung nicht nur in großem Umfang, sondern auch in einer Vielfalt verschiedener audiovisueller Formate. So betrachtet, entpuppt sich die Entgrenzung von Sprache als eine alltägliche Praxis des immer schon verkörperten Sprache-Sprechens, in dem die Körperlichkeit miteinander sprechender Personen durch ein körperliches Wahrnehmen audiovisueller Bilder moduliert wird.

Es liegt auf der Hand, dass Film- und Videotechnik ganz neue Möglichkeiten eröffnet haben, das Sprechen und Gestikulieren in ihrer Einheit als gemeinsam artikulierte Körperbewegungen zu beschreiben. Gleichzeitig beobachten wir eine tiefgreifende mediale Durchformung der Alltagswelt. Audiovisuelle Medien prägen die Formen, in denen uns Sprache-Sprechen begegnet, ob auf YouTube oder Instagram, ob in der Kommunikation per Zoom, ob als Serie auf Netflix, als Talkshow im Fernsehen oder als allzeit in einer Mediathek verfügbare Nachrichtensendung. Mit der alltäglichen Verfügbarkeit von Videoaufzeichnungen und der allgegenwärtigen Präsenz audiovisuell festgehaltener Formen des Sprachgebrauchs stellt sich für eine kulturwissenschaftlich orientierte Sprachgebrauchsforschung die drängende Frage nach einer kritischen Reflexion der Bedingungen, von denen Analysen mündlichen Sprachgebrauchs auf der Grundlage audiovisueller Aufzeichnungen auszugehen haben. Eine Erweiterung der sprachwissenschaftlichen Betrachtung der Multimodalität des Sprache-Sprechens um eine kulturwissenschaftlich-medienwissenschaftliche Perspektive erscheint daher dringend erforderlich. Auf diesen Umstand reagiert der vorliegende Beitrag mit einem transdisziplinären Vorschlag, wie Sprache-Sprechen, genauer wie die Multimodalität des Sprechens im Kontext audiovisueller Medien theoretisch wie methodisch zu fassen ist.

Der Beitrag skizziert Eckpunkte einer kulturellen Praxis, in der Sprache-Sprechen zunehmend mit der Zirkulation audiovisueller Bilder verbunden ist. Er macht sich für die These stark, dass es die Gestaltung der audiovisuellen Multimodalität ist, die die körperlich-sprachliche Multimodalität orchestriert und dass es dieses Zusammenspiel ist, das die Wahrnehmung, das Erleben und das affektive Mitschwingen, die affektive Bewegung der Zuschauer*innen an den Bildschirmen gestaltet. Da diese beiden Ausdrucksformen »die Multimodalität des Sprechens« und die »Multimodalität audiovisueller Medien« gewissermaßen in die Zuständigkeitsbereiche verschiedener Fächer fallen, ergibt sich hier die drängende Aufgabe, die fachwissenschaftliche Expertise zu diesen beiden Gegenstandsbereichen ernst zu nehmen und auf dieser Grundlage eine transdisziplinäre Perspektive zu entwickeln. Dieser Versuch

wurde im Kontext verschiedener filmwissenschaftlich-sprachwissenschaftlicher Forschungsprojekte unternommen und am Gegenstand multimodaler Metaphorik in Sprache, Geste und audiovisuellen Medien (Greifenstein et al. 2018; Müller/Kappelhoff 2018) entwickelt.¹ Der kinesisch-kinematographische Ansatz wurde in verschiedenen Forschungsarbeiten am Beispiel verschiedener Gegenstandsbereiche erprobt und fortentwickelt (Greifenstein 2020; Horst 2018; Kindler-Mathôt et al. i. d. Bd.; Scherer 2024; Schmitt 2020).²

Wie sich diese kinesisch-kinematographische Perspektive methodisch umsetzen lässt, dokumentiert im vorliegenden Band der Beitrag »Sprechen in audiovisuellen Medien. Systematik für eine Analyse multimodaler Kommunikation« (Kindler-Mathôt et al. i. d. Bd.). Der vorliegende Beitrag skizziert, wie sich diese kulturwissenschaftliche Methode theoretisch herleiten und begründen lässt.

Dabei werden Parallelen sprachlicher und audiovisueller Multimodalität aufgezeigt, die entscheidend dafür sind, wie Zuschauer*innen das Sprechen in audiovisuellen Medien wahrnehmen. Es geht darum zu verstehen, *was* die Betrachter*innen audiovisueller Bilder vom Sprache-Sprechen zu hören und zu sehen bekommen und wie die Orchestrierung der audiovisuellen Bilder das Fühlen der Zuschauer*innen moduliert.

Im Folgenden gehe ich auf die Eckpunkte dieser medienästhetischen Perspektive ein und stelle drei Aspekte vor, die das Sprache-Sprechen und die audiovisuelle Multimodalität miteinander teilen: (1) Temporalität und Bewegung, (2) Ausdrucksbewegung und Interaffektivität, (3) Interkorporalität und ästhetische Erfahrung. Diese systematischen Parallelen bereiten den Weg nicht nur zur Methodenentwicklung, sondern sie sind vor allem auch Voraussetzung für eine kritische Analyse der Wirkmacht von Sprache-Sprechen als ein Hören, Sehen, Fühlen, kurzum ein Erfahren audiovisueller Bilder.

1 Eine ausführliche Darstellung der filmtheoretischen Grundlagen findet sich in »Kognition und Reflexion: Zur Theorie filmischen Denkens« (Kappelhoff 2018) sowie in »Cinematic Metaphor: Experience – Affectivity – Temporality« (Müller/Kappelhoff 2018), hier auch in ihren Bezügen zur Körpergeste und verkörpertem Sprechen.

2 Beispiele sind: deutsche und polnische Wahlwerbespots (Horst 2018), Hollywood Kino, wie die Screwball Comedy (Greifenstein 2020) oder Hitchcock's *Spellbound* (Müller/Kappelhoff 2018; Schmitt 2020) sowie Deutscher Spielfilm (Schmitt 2020), Politiksendungen, Nachrichten, Werbespots und Musikvideos (Müller/Kappelhoff 2018; Schmitt 2020; Scherer 2024) und parlamentarische Reden (Kindler-Mathôt et al. i. d. Bd.).

2 Temporalität und Bewegung

Audiovisuelle Bilder entfalten sich, ebenso wie das Sprechen, in der Zeit. Sie existieren nur in der Zeit, genauer noch, nur als Bewegungen, die sich in einer bestimmten zeitlichen Kontur entfalten.

2.1 Sprechen ist Körperbewegung

Denkt man die Sprache vom Sprechen her, dann ist sie zunächst einmal Körperbewegung. Laute sind die hörbaren Effekte der komplexen Bewegungen von Stimmbändern, Zunge, Lippen, die den Luftstrom in unterschiedlicher Weise durch den Artikulationsraum, durch Kehle, Mund, Rachen und Nase strömen lassen. Man könnte auch sagen, Laute entstehen als Folge artikulatorischer Gesten. So verweisen etwa die Gebärdensprachforscher Armstrong, Stokoe und Wilcox auf Ulric Neisser (1976), der das Sprechen als »articulatory gesturing« betrachtet und argumentiert, dass die Perzeption des Sprechens vergleichbar sei mit der Perzeption anderer Gesten (Armstrong/Stokoe/Wilcox 1995: 8)³. Die Abtrennung des Lautlichen von der Körperbewegung ist ein Artefakt, das von der Realität des Sprache-Sprechens absieht.

Lange bevor in der Sprachentwicklung die Lautsprache auf den Plan tritt, kommunizieren Eltern und Säuglinge mit den Bewegungen ihrer Körper. Kommunikation findet in der Ontogenese zunächst einmal als ganzkörperliches Ereignis statt, aus dem sich der »artikulierte« Gebrauch der Lautsprache nach und nach herauschält. Der Säuglingsforscher Daniel Stern beschreibt diese kommunikative Verschränkung der Körper als Tanz (Stern 1985): ein Tanz, dessen Spuren in der Synchronisierung von Körperbewegung von Gesprächsteilnehmer*innen sichtbar werden (Kendon 1970, 1972; Oben/Brône 2015, 2016) und über die die Sprecher*innen ein zuallererst körperliches Miteinander herstellen, das man als körperliches Fundament der Interaktion beschreiben kann (Müller/Bohle 2007). Eine sehr typische Form, wie ein soziales Miteinander durch die räumliche Orientierung der Körper hergestellt

3 »Oddly, much more work has been done to describe speech than to describe signing in gestural terms. The psychologist Ulrich Neisser, for example, has noted that it is possible to describe speech as »articulatory gesturing«, and to treat speech perception as comparable to perceiving gestures of others kinds«. Vgl. auch den Hinweis auf das Sprechen als »subtly interleaved patterns of movements, coordinated across articulators« (Armstrong/Stokoe/Wilcox 1995: 8).

wird, ist die Kreisform. Diese Anordnung schafft im wörtlichen Sinne einen Interaktionsraum. Zu diesem Raum haben die Mitglieder der Gruppe exklusiven Zugang, er hat ein Innen (die Vorderseite der Körper) und ein Außen (die Rückseite der Körper), grenzt die so Zusammenstehenden nach Außen ab, und Neuzugänge erfordern eine Rekonfiguration, eine Erweiterung des Kreises. Das Miteinanderstehen, das körperliche Aufeinander-bezogen-Sein ermöglicht das Miteinander-Sprechen. Kendon hat solche räumlich-körperlichen Konfigurationen schon früh als Formationen und Verhaltenssysteme sozialer Begegnungen beschrieben (Kendon 1990). Eingehender hat er das Zusammenstehen von Personen in einem Kreis untersucht. Kendon fasst diese Kreisformation als einen sozialen Raum, der eine spezifische Form sozialen Verhaltens organisiert, ein »F-formation system«:

An F-formation arises whenever two or more people sustain a spatial and orientational relationship in which the space between them is one to which they have equal, direct, and exclusive access. Such a pattern can be seen in the circle of the free-standing conversational group. Here the participants stand so that they all face inwards to a small space which they cooperate together to sustain and which is not easily accessible to others who may be in the vicinity. The system of behavioral organization by which such a spatial-orientational pattern is established and sustained will be called an F-formation system. (Kendon 1990: 209)

Nimmt man diese grundsätzliche Multimodalität des Sprechens ernst, so verschiebt sich der Gegenstand sprachwissenschaftlicher Betrachtung von einem körperlosen kontext-entbundenen Sprachsystem zu einem kontext-gebundenen verkörperten – immer zeitlich zu denkenden – Miteinander-Sprechen: »Multimodal Languaging«. So gesehen ist Sprechen eine Körperpraxis, die inter-subjektives Verstehen ermöglicht.⁴ Auf die Perspektive des Verstehens zielt auch Flussers Unterscheidung der epistemischen und der ästhetischen Geste des Sprechens (Flusser 1991: 59). In der Geste des Sprechens verbindet sich der Weltbezug (epistemische Geste) mit der Gerichtetheit auf ein Gegenüber (ästhetische Geste). Dabei ist zu beachten, dass sich Flussers Begriff der ästhetischen Geste auf den alten Begriff der »Aisthesis« als ein Verstehen, nicht auf den umgangssprachlichen Gebrauch von »Ästhetik« als »Schönheit«, bezieht.

4 »Also ist das Denken des Sprechenden eine »adaequatio intellectus ad intellectum«, und seine Absicht ist nicht, irgendeine »objektive« Wahrheit einzufangen, sondern ein inter-subjektives Verstehen zu ermöglichen« (Flusser 1991: 59).

2.2 Gestikulieren und Sprechen als Zeitformen

Denkt man das Sprechen vom Gestikulieren her, dann fällt auf, dass sich beide in der Zeit entfalten: Sprechen und Gestikulieren treten als Zeitformen in Erscheinung, sie sind in dieser Zeitlichkeit miteinander verschränkt und beziehen aus dieser Verschränkung Sinn und Bedeutung. Diese Beobachtung gehört zu den Grundeinsichten der Gestenforschung und zwar in expliziter Abgrenzung vom Credo einer Trennung von Verbalität und Körperbewegung, das ja namensgebend für die Erforschung sogenannter non-verbaler Kommunikation war (Kendon 1972, 1980; McNeill 1985, 1992). Dagegen ist für die Gestenforschung gerade die Verschränkung beider Ausdrucksmodi *Movens* eines neuen Forschungsfeldes.⁵ Ob als Theorie einer Psychologie des Denkens, wie bei McNeill (1992), aus sozial-anthropologischer Sicht bei Kendon (2004) oder aus praxeologischer Perspektive, wie bei Streeck (2009), immer geht es um die Verschränkung des Gestikulierens mit dem Sprechen.

Über die verschiedenen Schulen und Ansätze hinweg besteht große Einhelligkeit darüber, dass die zeitliche Verschränkung der gestischen Bewegung mit dem Redefluss so feinkörnig synchronisiert ist, dass jeder Versuch eine ›zur Rede passende Geste‹ zu planen und auszuführen zu einer signifikanten Asynchronität zwischen sprachlichem und gestischem Ausdruck führt. Diese hochgradige Synchronisierung hat zuerst Kendon in seinen frühen Arbeiten gezeigt (Kendon 1970, 1972). Sein Vorschlag einer linearen Segmentierung von Gesten und ihrer Bedeutung war ein Meilenstein der Gestenforschung. Die Analyse des gestischen Bewegungsverlaufs in *Preparation, Stroke* und *Recovery* (Kendon 1972, 1980, 2004) hat sich als Bezugspunkt experimenteller wie qualitativer Forschung fest etabliert (vgl. Bresse/Ladewig 2011). Kendon orientiert sich hier zunächst an der prosodischen Struktur sprachlicher Äußerungen und konnte zeigen, dass im Gestikulieren ebenso wie beim Sprechen komplexe zeitliche Strukturen entstehen können (Kendon 1972). Ruhepositionen der Hände markieren die äußeren Grenzen gestischer Einheiten (*Gestu-*

5 Die Erforschung sprachbezogener Gesten (u.a. auch als Schnittstelle zur Gebärdensprachforschung) hat sich in den vergangenen Jahrzehnten fest als interdisziplinäres Forschungsfeld etabliert. Institutionell ist die Gestenforschung in der »International Society for Gesture Studies (ISGS): <https://www.gesturestudies.com/> verankert. Mit einer internationalen Zeitschrift (<https://benjamins.com/catalog/gest>), einer Buchreihe (<https://benjamins.com/catalog/gs>) und den im zweijährigen Rhythmus stattfindenden internationalen Tagungen haben sich vielbeachtete Orte des wissenschaftlichen Austausches und der Publikation von Forschungsergebnissen herausgebildet.

re Units). Innerhalb der gestischen Einheiten entfalten sich eine oder mehrere gestische Phrasen. Kendon definiert eine gestische Phrase als Abfolge einer Bewegung aus der Ruheposition hin zum Ausführungsort der Geste (Vorbereitungsphase, *Preparation*) und einer Ausführungsphase (*Stroke*). Diese Phrasen können recht komplexe temporale Strukturen ausbilden, etwa durch Wiederholungen oder indem eine Geste nur teilweise zurückgeführt wird (*Partial Retraction*) und sich in Form neuer *Strokes* weitere Gesten anschließen, ohne dass sich die Hände zurück in eine Ruheposition bewegen (Kendon 2004: Chapter 7).⁶

So hängt etwa im Kontext gesprächsanalytischer Rekonstruktionen multimodaler Interaktionen die gestische Bedeutung von der zeitlichen Positionierung der Geste im Gesprächszug, im *Turn*, ab. Ob die Geste zu Beginn eines Redezuges, als Teil einer Konstruktionseinheit (*Turn-Constructional Unit*, TCU) oder im potenziellen Turn-Übergaberaum (*Transition Relevance Place*, TRP) positioniert ist, bestimmt ihre gesprächsstrukturelle Bedeutung.⁷ Die Frage der Entfaltung von Gesten in der Zeit ist auch wesentlich für die semantische Bedeutung einer Geste. Insbesondere singuläre Gesten⁸ sind semantisch unterspezifiziert – ihre lokale, indexikalisch verankerte Bedeutung erhalten sie wesentlich durch die zeitliche Synchronizität mit sprachlichen Elementen der Äußerung (Müller 2010, 2015, 2016). So kann etwa ein gestisch modelliertes Rechteck die Form eines Aquariums ebenso darstellen wie einen Theorierahmen (Müller 1998, 2010, 2016). Die spezifische, lokale semantische Bedeutung der Geste hängt von ihrer zeitlichen Verschränkung mit der Rede ab. Kurz, wir haben es beim Gestikulieren und Sprechen mit multimodalen Zeitformen

6 Kendon verbindet diese Segmentierung gestischer Bewegungen mit einer Festlegung darauf, welcher Bestandteil des gestischen Bewegungsverlaufs bedeutungsrelevant ist: die Phase der Vorbereitung und der Höhepunkt, der *Stroke*. Anders formuliert: Kendon schneidet den gestischen Bewegungsverlauf nach dem Höhepunkt ab. Ob man Kendon in diesem Punkt folgt oder nicht, es bleibt ein breiter Konsens in der Forschung, dass die zeitliche Koordinierung von Sprechen und Gestikulieren entscheidend für die wechselseitige Konstitution von Bedeutung ist.

7 Ich verweise hier auf Arbeiten von Schegloff (1984), von Streeck/Hartge (1992) oder auch von Schmitt (2005).

8 Singuläre Gesten sind Gesten, die im Moment des Sprechens entstehen. Sie unterscheiden sich von rekurrenten Gesten, die sich durch einen hybriden Charakter auszeichnen und sowohl sedimentierte als auch lokale Bedeutungsaspekte aufweisen und von Emblemen, deren Bedeutung kulturell vollständig festgelegt, also konventionalisiert ist (vgl. Müller 2010, 2018, 2024b). Übergänge zwischen den Typen sind fließend.

zu tun, deren Bedeutung aus der zeitlichen Verschränkung der beiden Modi hervorgeht.

Soweit so gut, aber haben diese Segmentierungen, diese Einheiten auch etwas mit dem Verstehen zu tun? Damit ist nun nicht nur das Verstehen als Form interaktiver Aushandlungsprozesse, sondern Verstehen als Verkörperung, als Bewegungsverstehen angesprochen. Kehren wir nochmal zu Kendon zurück: Er verbindet seine Analyse des gestischen Höhepunkts (*Gesture Stroke*) als bedeutungsvolle Phase gestischer Bewegungen mit einem Rekurs auf die Ausdrucksbewegung als theoretisches Konstrukt, das in der Anthropologie, Soziologie, Tanztheorie, Sprachwissenschaft, Psychologie und Filmtheorie der zwanziger Jahre des letzten Jahrhunderts ausgearbeitet wurde. Man denke hier an Plessner (1982 [1925]), Wundt (1975 [1900–1925]), Bühler (1933, 1934), Simmel (1993 [1901–1908]) und den Tänzer Rudolf von Laban (1956). Kendon bezieht sich mit seinem Vorschlag direkt auf die Bewegungsanalyse des Ausdruckstänzers Laban, die im Übrigen bis heute Grundlage von Tanztherapien und einer immer noch gebräuchlichen Tanznotation ist (Kennedy 2013). So schreibt Kendon:

The phase of the movement excursion closest to its apex is usually recognized as the phase when the ›expression‹ of the gesture, ›whatever it may be, is accomplished. It is the phase of the excursion in which the movement dynamics of ›effort‹ and ›shape‹ are manifested with greatest clarity (see Laban and Lawrence 1947; Dell 1970; Bartinieff (sic!) and Lewis 1980). This phase is called the stroke. (Kendon 2004: 112)

Wie diese gestischen Phrasen und Einheiten aber verstanden werden, bleibt weitgehend offen, und so transformieren anthropologische, semiotische oder auch psychologische Ansätze in der Gestenforschung Kendons Bewegungsfigurationen in ›flüchtige‹ aber gleichwohl statische ›Zeichenbilder‹. Gesten werden als Züge in einem Gespräch (Streeck/Hartge 1992), als verbunden mit einer lexikalischen Einheit (Schegloff 1984) oder als Ausdruck mentaler Bilder (McNeill 1992) betrachtet. Obwohl die zeitliche Positionierung der Geste relativ zur Sprache in allen Ansätzen eine wichtige Rolle spielt, bleibt die Temporalität, die Bewegungsentfaltung von Gesten, wenn überhaupt, nur implizit relevant. Typischerweise würde dann eine Geste, die die Stufen einer Treppe nachzeichnet und die überlappend mit dem Adjektiv ›steigend‹ verwendet wird als auf dieses Wort bezogen oder im besten Fall auf die Nominalphrase »steigende Lohnzusatzkosten« bezogene gestische Einheit

sie gestisch als Treppenstufen in die Luft – ein Beispiel multimodaler Metaphorik, wie wir sie häufig in der Alltagssprache beobachten können (Müller 2008; Müller/Cienki 2009; Müller/Kappelhoff 2018). Betrachtet man die Geste genauer, so entfaltet sie sich als Abfolge von kleinen bogenförmigen Bewegungen, jede für sich etwas höher und weiter zur rechten Merckels in die Luft gezeichnet. Erst über die Zeit der Entfaltung entsteht so ein Bild aufsteigender Treppenstufen. Hier von *einer* Geste zu sprechen, die sich auf *ein* Wort bezieht und so eine multimodale Metapher bildet, verwandelt eine dynamische Bewegungsfiguration in ein statisches ›Zeichenbild‹.

Wie aber verstehen Betrachter*innen solche Bewegungsabläufe? Wie kommt es dazu, dass wir eine Folge von Treppenstufen wahrnehmen und nicht einzelne voneinander unabhängige kleine Bögen? Um dies zu erläutern, möchte ich Bezug auf Plessners Konzept des Bewegungsbildes nehmen (Plessner 1982 [1925]). Plessner beschreibt Körperbewegungen als dynamische Gestalten, die sich Beobachter*innen als bildhafte Bewegungsgestalten darstellen. Als Beispiele nennt er unter anderem Handbewegungen wie Greifen und Abwehren, die als Ausgangspunkt gestischer Als-Ob-Handlungen äußerst produktiv sind (Müller 2024a), so produktiv, dass sie Eingang in die Alltagsmetaphorik nicht nur des Deutschen gefunden haben: Das ›Be-greifen‹ eines Sachverhaltes, das ›Ab-wehren‹, das ›Weg-Wischen‹ eines Argumentes speisen sich aus solchen Bewegungsbildern.

Wo immer im Reich des Organischen Bewegungen erscheinen, verlaufen sie nach einheitlichem Rhythmus, zeigen sie eine, wohl auch experimentell nachweisbare, dynamische Gestalt. Sie rollen nicht stückhaft ab, als ob ihre Phasenfolge aus einzelnen Elementen assoziiert worden wäre, bilden kein Zeitmosaik, sondern eine gewisse Ganzheit ist vorgegeben, innerhalb derer die einzelnen Bewegungskurven variiert sind. [...] Diese Ganzheiten gehören zum Organismus durch sein Verhältnis zur Umwelt, seine Morphologie, seine artspezifischen Instinkte als motorische Kategorien [...]. Infolgedessen sind die Bewegungsgestalten bildhaft, wenn auch über eine gewisse Zeitdauer erstreckt, dem Beobachter gegenwärtig. [...] Greifen, Fliehen, Abwehren, Suchen, aber auch schon die »affektlosen« Formen wie Gehen,

Gesteneinheiten, Strokes sind **fett** markiert. Die Zeile unterhalb zeigt den Phasenverlauf der Gesten nach Kendon (**P**reparation, **S**troke, **R**ecovery, **r**estposition). Die Strokes sind durchnummeriert und werden in der untersten Zeile in ihrer Form und Bewegung beschrieben. Gleichbleibende Formen tragen die gleiche Nummerierung (vgl. S2).

Fliegen, Schwimmen [...] stellen solche Bewegungsbilder dar. (Plessner 1982 [1925], 77–78)

Merkels Geste entfaltet sich in der Zeit. Sie ist kein statisches Bild einer Treppe, sie ist ein Bewegungsbild (Abb. 2). Es besteht kein kausaler Zusammenhang zwischen den einzelnen *Gesture Strokes*, den einzelnen in die Luft gezeichneten aufsteigenden Bögen, vielmehr entsteht der Zusammenhang erst im Prozess des Betrachtens. Kendon hat darauf hingewiesen, dass sich Gesten grundsätzlich in der Zeit entfalten. Mit Plessner lässt sich nun festhalten, dass Gesten Bewegungsbilder sind, die als Gestalten »dem Beobachter gegenwärtig« sind (Plessner 1982 [1925], 78).

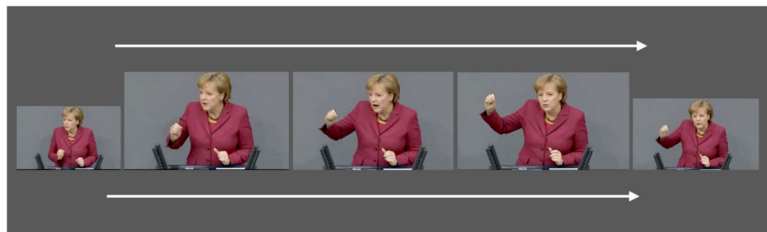
Abb. 2: Merkels Geste als Bewegungsbild



Ebenso wie gestische Bilder entfalten sich auch Filmbilder in der Zeit. Auch hier haben wir es nicht mit einer Abfolge von statischen Zeichenbildern zu tun, sondern mit Bewegungsbildern oder mit Zeitbildern. Diese Bewegungsbilder weisen ein dynamisches Profil, eine zeitliche Struktur auf (Deleuze 2008 a/b; Kappelhoff 2006). Sie sind ein unentwegtes Werden und Vergehen. Abb. 3 illustriert diese zeitliche Dimension audiovisueller Bewegungsbilder für den kleinen Ausschnitt aus der Rede Merkels im Bundestag. Es geht also nicht nur um die Zeit, in der sich die Geste entfaltet, sondern auch um die des audiovisuellen Bildes. Allerdings sind diese Bewegungsbilder keine prae hoc existierenden Einheiten, sie formieren sich erst im Prozess der Wahrnehmung von Zuschauer*innen. Ebenso wie das gestische Bewegungsbild entsteht auch ein audiovi-

suelles Bewegungsbild erst in der Wahrnehmung der Betrachter*in als Wahrnehmungsgestalt.

Abb. 3: Merkels sprachlich-gestische Äußerung als audiovisuellen Bewegungsbild



Die Unterscheidung von (mechanischen) Bewegtbildern und (wahrgenommenen) Bewegungsbildern ist auch Gegenstand der Überlegungen Plessners, der bei seinen Überlegungen zu Körperbewegungen als Bewegungsbildern an das kinematographische Bild gedacht haben mag, wie Kappelhoff ausführlich:

Denn diese [die Kinematik, CM] bezieht sich auf den geometrischen Raum; das Bewegungsbild hingegen hat keine räumlichen Koordinaten, sondern meint eine zeitliche Struktur, die sich darstellt als das Diagramm eines Werdens, einer Metamorphose, einer sich verändernden Welt. (Kappelhoff 2004a: 153)

Ebenso wie in der Kommunikation von Angesicht zu Angesicht Sprache, Geste, Blick und Mimik zusammenwirken, entstehen aus dem Zusammenspiel von Kameraführung, Montage, Mise-en-Scène und Sound Bewegungsgestalten, die als Bewegungsbilder in ihren jeweiligen Qualitäten, ihrem spezifischen Rhythmus gesehen und gehört, also unmittelbar körperlich wahrgenommen werden. Die Zuschauer*innen der Aufzeichnung von Merkels Rede erfahren deshalb die audiovisuellen Bilder genauso unmittelbar als Ausdruck, wie das Publikum, das im Bundestag die Rede sah und hörte. Allerdings sind erstere abhängig davon, was ihnen die Regie, die die Aufzeichnungen der Bundestagsreden inszeniert, zu sehen gibt. In unserem Fall gab es keinen Schnitt in die Totale oder ins Publikum. Wir sehen Merkel gleichbleibend in halbnaher Einstellung, zentral im Bildraum positioniert und als herausgehobene Person in Szene gesetzt. Die Zwischenrufe, auf die sie mit ihrer kurzen sprachlich-

gestischem Seitensequenz reagiert, bleiben für die Zuschauer*innen an den Bildschirmen im Hintergrund.

3 Ausdrucksbewegung und Interaffektivität

Wie aber ›fühlt‹ sich Merckels kurze Gegenrede für das parlamentarische Publikum, wie für die Zuschauer*innen an ihren Bildschirmen an? Spielt das überhaupt eine Rolle für das Verstehen? Anders gefragt, lässt sich Sprache-Sprechen als Kommunikation von ›Inhalten‹ und als Vollzug von sprachlichen Handlungen treffend und hinreichend beschreiben? Wagen wir einen Blick zurück an den Beginn des zwanzigsten Jahrhunderts. Ausdruck, Ausdrucksbewegung, Ausdruckstheorie waren in dieser Zeit große Themen in Wissenschaft und Kunst. Sprachtheoretische Reflexionen etwa bei Wilhelm Wundt (Wundt 1975 [1900–1925]) oder bei Karl Bühler (Bühler 1933, 1934) waren Teil dieses breitangelegten kulturellen Diskurses, und ihre sprachtheoretischen Positionen sind wesentlich als Beiträge zu diesem Diskurs angelegt.

3.1 Sprache-Sprechen und Gestikulieren sind multifunktional

In kaum einer Einführung in die Sprach- und Kommunikationswissenschaft fehlt Karl Bühlers Organon-Modell der Sprache (Bühler 1982 [1934]). Dort formuliert Bühler, dass Sprachzeichen zugleich Ausdruck, Darstellung und Appell sind, wobei die drei Funktionen in wechselnden Dominanzen im Sprechen auftreten, aber immer ko-präsent sind.

Bühlers postulierte Gleichberechtigung der drei Funktionen wurde mit wenigen Ausnahmen in der zeitgenössischen Sprachwissenschaft nicht systematisch aufgegriffen (vgl. aber Eschbach 1984, Loenhoff 2017).¹¹ Auch, dass das Organon-Modell Baustein einer theoretischen Trias war, die Sprache-Sprechen aus einer Ausdrucks-, einer Sprach-, und einer Handlungstheorie heraus entwickelt, hat wenig bis gar keine Beachtung gefunden. Der Grund, dass Bühlers Arbeiten in ihrer vor dem zweiten Weltkrieg international hochbeachteten Radikalität und Innovationskraft im Nachkriegsdeutschland nicht den verdienten Widerhall gefunden haben, liegt im Abbruch einer deutschsprachigen wissenschaftlichen Tradition, die mit der nationalsozialistischen

11 Vgl. auch die Arbeiten in der Tradition der Deutschen Ausdruckspsychologie (Krumhuber et al. 2013).

Verfolgung deutschsprachiger jüdischer Intellektueller, zu denen auch Karl Bühler gehörte, einherging (vgl. Jäger 2010). Bühler hat die Ausdruckstheorie und die Sprachtheorie als Monographien fertigstellen können (Bühler 1933, 1934). Im Exil konnte er daran nicht mehr anknüpfen. Dabei geht die Beschäftigung mit Ausdruckstheorien der Sprachreflexion voraus. So gelangt er erst über seine Reflektionen zum Verhältnis von Erleben und Ausdruck zur Sprachtheorie (Jäger 2010). Für den Psychologen Bühler ist Ausdruck ohne Sprechen mit Jemandem nicht denkbar. Sprache-Sprechen als Mitteilung an ein Gegenüber ist Bedingung der Möglichkeit einer Verschmelzung von Erleben und Ausdruck:

[...] für Bühler stellt sich die Einheit von seelischem Erlebnis und Ausdruck nur ein, wenn Ausdruck als »Kundgabe« gedacht wird, als eine kommunikative Handlung im Rahmen einer sozialen Situation, in der die Äußerung an mögliche Adressaten gerichtet werden kann. Der Ausdruck würde nämlich – wie Bühler zeigt – in der Einzelseele nicht zu einer festen Verbindung mit der Vorstellung, bzw. einem seelischen Erlebnis gelangen können, weil beide, das seelische Erlebnis und seine gestische oder mimische Äußerung, sich erst auf der Bühne interaktiven sozialen Handelns herstellen. (Jäger 2010: 8)

Mit dieser Position kritisiert Bühler Wundts individualpsychologische Vorstellungen von Ausdruck, die von einer scharfen Trennung innerer Empfindungen und äußerer Ausdrucksbewegungen ausgehen (vgl. Jäger 2010; Wundt 1975 [1900–1925]). Ausdruck ist für Bühler immer an ein soziales Miteinander gebunden, nur als Mitteilung an jemanden theoretisch sinnvoll zu fassen: »Das Bühlersche Subjekt ist [...] ein soziales Subjekt« (Jäger 2010: 7). Damit ist auch die Sprache sozial gedacht, vom Sprechen aus. Bühler fasst sprachliche Zeichen als »konkrete Schallphänomene«, die nur aus einem Sprechereignis sinnvoll zu denken sind. Dieses Sprache-Sprechen ist an drei Koordinaten gebunden: der eine, der andere, die Dinge (Bühler 1982: 26f.). Bühler denkt die Sprache mithin als kommunikatives Ereignis, als Handlung, als Praxis, die immer mehrere Funktionen zugleich erfüllt.¹² Er entwickelt diese Multifunktionalität aus der Trias des Sprechens als kommunikatives Ereignis: »Dreifach ist die Leistung der menschlichen Sprache [...] Ausdruck, Appell und

12 »Mich dünkt, es sei so etwas wie ein Ariadnefaden, der aus allerhand nur halb begriffenen Verwicklungen herausführt, gefunden, wenn man das Sprechen entschlossen als Handlung (und das ist die volle Praxis im Sinne des Aristoteles) bestimmt.« (Bühler 1934: 56–57)

Darstellung [...].« (Bühler 1982: 28). Die Sprachfunktionen überlagern sich, sie sind gleichzeitig präsent, im konkreten Moment des Sprechens steht jedoch stets die eine oder andere im Vordergrund, ist eine oder die andere dominant (Bühler 1982: 30f; Müller 1998: 20, 90). Meist ist dies die Darstellungsfunktion, der Bezug auf die Welt, der allerdings nur in der Gerichtetheit auf ein Gegenüber zu denken ist und der eben immer auch Ausdruck ist.

Merkels Äußerung zu den steigenden Lohnnebenkosten macht nur Sinn als Antwort auf einen kritischen Zwischenruf aus den Reihen der Parlamentarier. Auf diesen reagiert sie, wenn sie sagt: »Und Sie haben keine Antwort auf die Frage, was man denn tut, um die immer steigenden Lohnnebenkosten zu vermeiden.« Diese Entgegnung ist zugleich ärgerlicher und kämpferischer *Ausdruck*, wie er mit der Zurückweisung des Zwischenrufs als kommunikative Handlung (*Appell* in Bühlers Terminologie) einhergeht und dies mit einer Sachfrage verbindet (*Darstellung* im Sinne Bühlers). Eine vergleichbare Multifunktionalität sieht Bühler auch in den Gesten von Sprecher*innen realisiert.

[...] warum sollen z.B. gestenhaft malende Hände nicht imstande sein, das eine Mal frisch und froh, ein zweites Mal wütend, ein drittes Mal zögernd und erschreckt den Akt des Zeichnens durchzuführen? Und wenn sie es vermögen, so stellen sie dar und drücken aus in eins. (Bühler 1933: 39)

Mit dieser Beobachtung ist Bühler seiner Zeit weit voraus. Er beschreibt hier nämlich nicht mehr und nicht weniger das Potential »gestenhaft malende[r] Hände«, (Gebärden-)Sprache zu werden. Auch Handbewegungen können darstellen und ausdrücken zugleich, und diese sind in der Regel auf ein Gegenüber ausgerichtet.¹³

Wenn nun im Sprache-Sprechen gestische und lautsprachliche Äußerungsformen zusammenkommen, können sie synchron unterschiedliche Bedeutungsfacetten artikulieren. So verbindet Merkel etwa in ihrer Gegenrede eine (für ihre Verhältnisse) ärgerlich-aggressive sprachliche Reaktion mit einer Reihe kleiner, sanft ausgeführter, bogenförmiger Gesten, die hochartikuliert, sehr präzise die Stufen einer Treppe in die Luft zeichnen. Die Gegenrede ist in ihrem affektiven Ausdruck ärgerlich, sanft und präzise zugleich.

13 Zur funktionalen Parallelität von Sprache und Geste und zum Sprachpotential von Gesten vgl. Müller (1998: 87f.).

3.2 Ausdruck und Verstehen

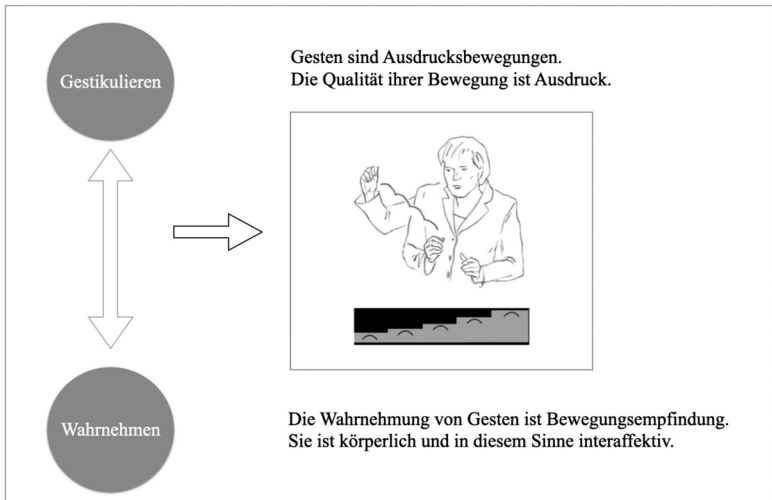
Wir halten mit Bühler fest, Sprache-Sprechen einschließlich der mit dem Sprechen einhergehenden Gesten ist zugleich Ausdruck, Darstellung, Appell. Oder anders formuliert, Sprache-Sprechen ist immer auch Affektausdruck, eingebunden in soziales Miteinander. Für Bühler ist Ausdruck als Verbindung von Empfindung und äußerer Form nur denkbar als kommunikatives Ereignis (Jäger 2010). Nimmt man diesen Punkt ernst, dann stellt sich die Frage nach dem Verstehen. Wie kommt Sprache-Sprechen als ›Ausdruck‹ bei den Zuschauer*innen im Parlament und an den Bildschirmen an? Mit Bezug auf Merleau-Pontys Phänomenologie der Wahrnehmung lautet die Antwort: Das Verstehen ist ein leibliches, ein körperliches. Die Parlamentarier*innen empfinden die Ausdrucksqualität von Merkels Geste unmittelbar körperlich. Sie müssen sie nicht erst als Zeichen dekodieren, das für eine Innerlichkeit steht. Mit Merleau-Ponty ließe sich sagen: Die sanften und zugleich präzise ausgeführten gestischen Bewegungen *sind* Sanftheit, *sind* Präzision. Der ärgerliche Ausdruck der Stimme transportiert nicht einen inneren Zustand nach außen, er *ist* der Ärger. Es gibt weder Sanftheit, noch Präzision, noch Ärger, die als Affekte dem körperlichen Ausdruck vorgängig wären.

Faced with an angry or threatening gesture, I have no need, in order to understand it, to recall the feelings which I myself experienced when I used these gestures on my own account. [...] I do not see anger or a threatening attitude as a psychic fact hidden behind the gesture, I read anger in it. The gesture *does not make me think* of anger, it is anger itself (Merleau-Ponty 2005 [1945]: 214)

Das affektive Verstehen ist also ein leibliches, ein unmittelbar körperliches. Die Körperbewegung ist Affekt, sei es die Tonhöhenbewegung der Stimme, seien es die gestischen Bewegungen (s. Abb. 4).

Betrachtet man Sprache-Sprechen als Ausdrucksbewegung, dann sieht man nicht primär einen Austausch von Gedanken, sondern ein zuallererst körperlich vollzogenes Miteinander, einen interkorporalen Dialog. In diesem entstehen und zirkulieren Affekte, das Sprache-Spreche ist in diesem Sinne grundsätzlich ›interaffektiv‹ (Horst et al. 2014). Fuchs und De Jaeger (2009: 479) bringen diese phänomenologische Position auf den Punkt: »affects are not enclosed in an inner mental sphere to be deciphered from outside but come into existence, change and circulate between self and other in the intercorporeal dialogue«.

Abb. 4: Gesten als Ausdrucksbewegungen



3.3 Multidimensionale Ausdrucksgestalten

Was Zuschauer*innen ebenso wie Teilnehmer*innen an einem Gespräch wahrnehmen, sind dabei nicht aufaddierte Informationen, die sich einzeln in den verschiedenen Ausdrucksmodalitäten, also in Geste, Wort, Mimik, oder als Zoom auf das Gesicht einer Redner*in oder durch einen Schnitt ins Publikum realisieren und die dann gewissermaßen durch bloße Summierung Sinn ergeben. Was wir wahrnehmen, wenn wir einen Film schauen, oder wenn wir unsere Gesprächspartner*innen anblicken, sind multidimensionale Erfahrungsgestalten, die sich im Prozess des Zuhörens, Zusehens, des Hörens und Betrachtens audiovisueller Bilder als fluide aber gestalthafte Figurationen herausbilden (Müller/Kappelhoff 2018). Was auch immer Merkel in dem kleinen Ausschnitt sagt, die sanfte und gleichzeitig präzise Bewegung der Geste, ihre ganze gefasste Körperhaltung setzen einen bedachten, konzentrierten Grundton oder, anders formuliert, sie grundieren das Gesagte affektiv, geben ihm eine besondere affektive Tönung. Sie sind Teil einer Ausdrucksbewegung, die im Sprache-Sprechen entsteht und die sich gleich einer Handgeste als zeitliche Bewegungsfiguration mit Beginn, Höhepunkt und einem Abschluss entfaltet.

Multidimensionale Erfahrungsgestalten als Ausdrucksbewegungen betreffen sowohl die Multimodalität des Sprechens als auch audiovisuelle Bewegungsbilder. Die Bedeutungen stecken in den Bewegungen selbst und sind nicht dahinter verborgene, mittelbare Konzepte. Sprachlich-gestische und filmische Ausdrucksbewegungen werden als multidimensionale Erfahrungsgestalten leiblich, also unmittelbar, von den Zuschauer*innen erfahren und so verstanden. So schreibt Merleau-Ponty zur Wahrnehmung:

My perception is therefore not a sum of visual, tactile, and audible givens. Rather, I perceive in a total way with my whole being; I grasp a unique structure of the thing, a unique way of being, which speaks to all my senses at once. (Merleau-Ponty 1968: 48)

Der Körper der Sprachforscher*innen dient in der Analyse gewissermaßen als ›Instrument‹, die Beschreibungen der empfundenen Affektqualitäten dieser gestalthaften Bewegungsfigurationen sind intersubjektiv überprüfbar.¹⁴

3.4 Filmische und gestische Ausdrucksbewegung

Das Konzept der Ausdrucksbewegung wurde in den ersten Dekaden des zwanzigsten Jahrhunderts in Kunst und Wissenschaft lebhaft und in großer Breite, heute würde man vielleicht sagen ›transdisziplinär‹, diskutiert (Kappelhoff 2004a, 2018; Müller/Kappelhoff 2018). Der Bogen reicht von der Philosophischen Anthropologie, Psychologie, Sprachwissenschaft, bis zur Filmtheorie, Phänomenologie und Philosophischen Ästhetik. Für die Sprachforscher und Psychologen Bühler und Wundt, ebenso für Plessner in seiner philosophischen Anthropologie war es absolut selbstverständlich, die Geste als Ausdrucksbewegung zu fassen. Wie diese theoretisch zu greifen war, stand zur Debatte. Gleichzeitig war für die frühe Filmtheorie das ›Gestische‹ am Film die filmische Ausdrucksbewegung ein wichtiger Gedanke (Balázs 1924; Eisenstein 2010 [1926]; Münsterberg 2002 [1916]). Mit dem Konzept der Ausdrucksbewegung betonten sie (mit Ausnahme Wundts) die Verschränkung von Ausdruck und Erfahrung im menschlichen Verhalten. Diese Verschränkung beschreibt die wechselseitige Affizierung der Körper, sie zielt auf die Unmittelbarkeit der Bewegungserfahrung als verkörpertes, als affektives Empfinden.

14 Zur methodischen Umsetzung vgl. Kindler-Mathôt et al. i. d. Bd.

Überlegungen zur Parallelität von Film und Geste, zu gestischem und filmischem Bewegungsbild und zu ihrem Ausdruckscharakter stehen am Beginn der Filmtheorie. Wenn etwa Balázs (1924) im Ausdruck des Gesichts das Paradigma filmischer Bilder ausmacht, dann bezieht sich dies auf die Zeitlichkeit und die Unmittelbarkeit der Empfindung, die sich mit dem Prozess der Wahrnehmung von Bewegungsbildern verbindet (Kappelhoff 2004a: 154, 2004b). Diese Position steht im Gegensatz zu Ekmans (1972, 2006) Fotografien von Gesichtern, die den Ausdruck universeller Emotionen darstellen sollen. Während Ekmans fotografisch festgehaltene Gesichtsausdrücke insinuieren, dass sich ein Gefühl im Gesicht als statischer Ausdruck zeigt, geht Kappelhoff davon aus, dass sich Affekte über die Zeit als sanfte, langsam gemächliche oder explosive, hart konturierte mimische Bewegungen entfalten. Was für das Gesicht gilt, gilt auch für audiovisuelle Bilder. Audiovisuelle Affekte zeigen sich nicht primär in den Gesichtern der Schauspieler*innen, sondern im Rhythmus, in der Dynamik audiovisueller Bilder.

Die Leinwand ist das Gesicht des Kinos. Das ist, worum es Balázs geht.

Was Balázs an der Bewegung eines Gesichts zu greifen sucht, ist weder der Schein des lebendigen Gesichts noch die Repräsentation des »ganzen Menschen«, sondern die affektive Dimension des Kinos: ein Raum der artifiziellen Affekte, der technisch animierten, künstlichen Seele. (Kappelhoff 2004a: 155)

In ihrer Bewegungsqualität sind Bewegungsbilder Ausdruck. Dies gilt für den mimischen Ausdruck ebenso wie für das Filmbild. Beide sind nicht Repräsentationen ›anderer Zustände‹, ihre affektive Dimension ist die Qualität der Bewegung. Sie sind unmittelbar körperlich erfahrene Ausdrucksbewegungen. Greifenstein (2020: 48) beschreibt deshalb das Verhältnis von Ausdruck und Affekt als nicht arbiträr. Ausdrucksbewegungen begegnen uns als ganzheitliche Gestalten. So argumentiert etwa Plessner am Beispiel einer ängstlich davonlaufenden Katze oder der freudigen Begrüßung eines Hundes, dass die affektive Dimension des Verhaltens in der Qualität der Ganzheit der Bewegung liegt und gerade nicht in der objektiven Beschreibung der Bewegungen.

Daß die Katze mir davonläuft, läßt sich auch objektiv in bloßen Bewegungen darstellen; daß sie mich ängstlich flieht, ist mir überdies in diesen Bewegungen als einheitlicher Charakter mitgegeben. Daß der Hund an mir empor-

springt, ist objektiv konstatierbar; daß er mich freudig begrüßt, ist mir in seinem Gebaren als Richtungsform deutlich. (Plessner 1982 [1925]: 82)

In der Filmtheorie beschreibt das Konzept der Ausdrucksbewegung eine Bewegungsfiguration, die sich als Gestalt im Prozess einer Wahrnehmung herausbildet und deren Ausdrucksqualität als Bewegungsqualität unmittelbar körperlich empfunden wird. Kappelhoff fasst dies als filmische Ausdrucksbewegung (2004a): Spannung, Horror, Trauer, Melancholie, Sehnsucht, Freude vermitteln sich als Bewegungserfahrung, als ästhetische Lust unmittelbar körperlich. Dieser Typus von Erfahrung stellt sich nicht erst ein, nachdem die Puzzlesteinchen der verschiedenen Modalitäten (Dialog, Bildraumgestaltung, Kamerafahrt, Schnittfolge, Musik etc.) kognitiv rekonstruiert und auf eine dargestellte Figur projiziert wurden, der man als Zuschauer*in dann empathisch folgt. Vielmehr handelt es sich um eine unmittelbar empfundene, körperliche Erfahrung von Bewegungsqualitäten, genau in dem Sinne, wie Stern dies in seinen Beobachtungen zur Mutter-Kind-Interaktion als Miteinander-Tanzen zweier Körper beschrieben hat (Stern 1985). Die Leinwand, der Bildschirm tritt an die Stelle des Körpers – sie ist gewissermaßen ein artifizieller Körper (man denke an das von Balázs angeführte Gesicht als Paradigma des Kinos), der das Publikum tatsächlich bewegt – kollektiv erschrecktes Aufschreien oder Tränen der Rührung, die wir im Kino gemeinsam vergießen, zeugen von dieser körperlichen Unmittelbarkeit. Kappelhoff (2018: 129) formuliert den Grundgedanken des Konzepts der Ausdrucksbewegung als: »das Verhältnis zwischen affizierendem und affiziertem Körper«. In diesem Sinne haben wir es sowohl im zwischenmenschlichen Gespräch als auch im Kino mit einer grundsätzlich körperlichen Interaktion zu tun, die interaffektiv ist.

4 Interkorporalität und ästhetische Erfahrung

Wir halten fest, dass Ausdrucksbewegungen, ob audiovisuell oder sprachlich-gestisch, in ihrer Gestalthaftigkeit durch eine je spezifische Bewegungsdynamik zusammengehalten werden. Diese lässt sich in ihren Qualitäten beschreiben: Entfalten sie sich langsam, vorsichtig oder abrupt ansteigend und explosiv, fallen sie danach sanft ab oder hält sich eine Spannung über einen längeren Zeitraum? Immer zeichnen sie sich durch das spezifische Profil und die spezifische Qualität ihrer Bewegungsabläufe aus. Der Ausdruck liegt in der Bewegung selbst, sie ist es, die unmittelbar die Betrachter*in-

nen berührt, mitnimmt, abstößt, in Spannung hält oder diese auflöst. Die Wahrnehmung dieser Bewegung ist körperlich, unmittelbar, ganz gleich ob es sich um die Bewegungen von Gesprächspartner*innen oder um filmische Bilder handelt: Sie ist interkorporal, zwischenleiblich (Fuchs 2017). Gemeint ist einmal mehr die Unmittelbarkeit körperlicher Bewegungserfahrung die, so unsere (linguistisch-filmwissenschaftliche) These, sowohl die Multimodalität des Sprechens als auch die audiovisuelle Multimodalität charakterisiert.

Das bedeutet, wann immer uns ›Sprache-Sprechen‹ in audiovisuellen Medien begegnet, dekodieren wir als Zuschauer*innen und Zuhörer*innen nicht einfach den Inhalt einer sprachlich formulierten Botschaft, vielmehr sehen wir, wie der/die Sprecher*in sich bewegt, ob er/sie recht steif dasteht, unruhig auf- und abschreitet, ob und wie er/sie gestikuliert, wir nehmen wahr, ob gestische Bewegungen im Einklang mit dem Sprechrhythmus stehen oder nicht und mit welcher Dynamik und Qualität sie sich entfalten. Falls noch weitere Personen zu sehen sind, nehmen wir wahr, wie sich das Sprechen und Bewegen der einen mit den Körpern der anderen im Bildraum Anwesenden verschränkt und welche Qualität dieses Zusammenspiel der Körper hat, ob sie harmonisch bestätigend miteinander-schwingen oder sich in harscher Opposition konfrontieren. Die ästhetische Erfahrung dieser audiovisuellen Bilder ist durch und durch körperlich, sie ist Hören, Sehen, Fühlen.

Was wir als Zuschauer*innen an den Bildschirmen von diesen Interaktionen zu sehen bekommen, ist jedoch ganz und gar abhängig von der Gestaltung der audiovisuellen Bildräume. Kameraperspektiven und Schnittfolgen entscheiden noch im neutralen Modus parlamentarischer Dokumentation von Bundestagsdebatten, was genau die demokratische Öffentlichkeit von den Debatten zu sehen bekommt. Gleiches gilt in gewisser Weise auch für die Gestaltung von ethnographischen oder experimentellen Settings, die die Multimodalität des Sprechens in mehr oder minder natürlichen Kontexten untersuchen. Auch hier entscheidet die Position einer oder mehrerer Kameras darüber, was hinterher als ›Datum‹, als ›Material‹ den jeweiligen Analysen zugrunde liegt. Seit Adam Kendons frühen Analysen zur Synchronisation von Sprechen und Körperbewegung ist die Filmkamera zum Mikroskop der Gestenforschung geworden (Kendon 1972). Nun ist es ganz offensichtlich, dass ein Mikroskop eine sehr spezifische Perspektive auf das Untersuchungsobjekt eröffnet und andere Aspekte dafür in den Hintergrund treten. Gestenforscher*innen kennen die Artefakte, die durch das Abspielen von Körperbewegungen in Zeitlupen entstehen, sie sind vertraut mit den Problemen, die durch die extrem hohe Auflösung digitaler Bilder im HD-Format für die Segmentie-

rung von Körperbewegungen entstehen. Wo beim Videobild der achtziger und neunziger Jahre Bewegungs- und Ruhephasen durch unscharfe und scharfe Videobilder zu erkennen waren (Seyfeddinipur 2006), zeigt die heutige digitale Aufnahmequalität eines jeden Mobiltelefons Bewegungsabläufe in einer kristallklaren Bildschärfe, die früher nur bei spektakulären Naturaufnahmen oder bei Aufnahmen von Sportwettkämpfen zu finden waren.

Das bedeutet, dass nicht nur für die Rezeption von Filmkunst gilt, dass die Wahrnehmung der audiovisuellen Bilder zuallererst eine körperliche ist, die Position der Zuschauer*in, ob als Forschende, im Kinosaal oder am Computerbildschirm ist daher entscheidend. Ohne Zuschauer*in kein Film, ohne Gegenüber keine Geste, kein Sprache-Sprechen.

4.1 Die Perspektive der Zuschauer*innen und das ästhetische Erleben

Die Analyse der Multimodalität des Sprechens in audiovisuellen Medien lässt sich nicht ablösen von der Wahrnehmungsposition der Zuschauer*innen. Sie lässt sich auch nicht separieren vom Charakter des ästhetischen Erlebens als einer zeitlich gestalteten Bewegungserfahrung, die die Zuschauer*innen mitnimmt, die sie affektiv in die Entfaltung filmischer Bilder einbindet. Nur so lässt sich die »Position eines verstehenden Subjekts« angeben (Kappelhoff 2018: 144), nur so lässt sich überhaupt rekonstruieren, wie audiovisuelle Bilder wirken, wie sie Zuschauer*innen berühren, wie sie ihre affektive Kraft entfalten:

Erst durch die Rekonstruktion der affektiven Verwicklung des Rezipienten, die Rekonstruktion der Position, aus der heraus sich das ›Gefühl für das Ganze der Situation‹ erschließt, lässt sich überhaupt die Position eines verstehenden Subjektes angeben. Ist dieses doch unauflösbar in seinen körperlichen Empfindungen in den Prozess der Entfaltung des filmischen Bildes verstrickt. (Kappelhoff 2018: 144)

Wie aber lässt sich dieses Empfinden der Zuschauer*innen analytisch einholen, ohne dabei auf Rezipient*innenbefragungen oder ähnliche sozialwissenschaftliche Verfahren zurückzugreifen? Die Position der phänomenologisch inspirierten Filmwissenschaft geht den Weg über eine detaillierte Analyse der audiovisuellen Komposition als einer Zeitform, die ein Wahrnehmungserleben auf der Seite der Zuschauer*innen als zeitliche Erfahrung moduliert. Kappelhoff formuliert diese analytische Perspektive wie folgt:

Deshalb muss die Analyse audiovisueller Bilder immer den Weg über die Rekonstruktion des ästhetischen Erlebens gehen – das meint die Rekonstruktion der je besonderen Wahrnehmungsform des audiovisuellen Bewegtbildes (d.h. seine Performanz, also die zeitliche Struktur der Entfaltung des Bewegtbildes in der Rezeption durch den Bildkonsumenten) und seiner inszenatorischen Komposition (d.h. der zeitlichen Figur, als welche die zeitliche Struktur im Prozess der Rezeption zur bedeutungsvollen Gestalt eines filmischen Bewegtbildes wird). (Kappelhoff 2018: 144)

Methodologisch gilt es in einer medienästhetisch begründeten Perspektive durch eine präzise Analyse filmischer Bewegtbilder, das verkörperte Erleben der Zuschauer*innen vor den Leinwänden und an den Bildschirmen zu rekonstruieren. In der Wahrnehmung der Zuschauer*innen im Kino, vor dem Fernseher oder vor dem Laptop entfalten sich die audiovisuellen Aufnahmen als Ganzes: Sie ergeben eine zeitliche Gestalt, die nicht im Prozess des Schauens von den Zuschauer*innen in ihre Einzelteile zerlegt wird, sondern die Stimmliches, Gestisches, Kamerabewegung, Bildraumgestaltung, Montage in sich vereint. Damit, so die These, verhält es sich mit der Position der Zuschauer*innen audiovisueller Formate genauso, wie mit der von Teilnehmer*innen an einem Gespräch. Körperliche Ausdrucksbewegungen sind als »Bewegungsgestalten bildhaft, [...], dem Beobachter gegenwärtig« (Plessner 1982 [1925]: 77–78).

Die Perspektive der Zuschauer*innen für die Multimodalität des Sprechens in audiovisuellen Medien ernst zu nehmen, heißt daher, die sich entfaltende Wahrnehmungsgestalt als analytischen Ausgangspunkt anzunehmen. Diese Wahrnehmungsgestalt kann in der weiteren Analyse auf ihre Orchestrierung hin, d.h. mit Bezug auf sprachliche und gestische Figurationen, Bildraum, Kameraeinstellung, Montage etc. beschrieben werden.¹⁵

15 Die einzelnen Analyseschritte dieses Prozesses werden im Beitrag von Kindler-Mathôt et al. (i. d. Bd.) erläutert. Einen Überblick über das gesamte methodische Instrumentarium sowohl für die Multimodalität des Sprechens als auch für die audiovisuelle Multimodalität findet sich im Appendix bei Müller/Kappelhoff (2018).

Abb. 5: Filmische Bewegungsbilder als Ausdrucksbewegung



Kehren wir ein letztes Mal zum Ausschnitt aus der Rede Merkels im Bundestag zurück (Abb. 5). Betrachten wir sie als kompositorische Gestalt, dann zeichnet sich diese durch eine zurückhaltende neutrale Inszenierung aus. Wir sehen eine stabile Kameraposition, die Merkel in der Halbtotalen zeigt, es gibt keinen Schnitt ins Publikum, keine Kamerabewegung, keine Veränderung der Bildraumgestaltung. Merkel wird in einer zentralen Position des Bildraumes gezeigt. In rotem Blazer vor grauem Hintergrund sind ihre Bewegungen gut sichtbar. Die audiovisuelle Inszenierung stellt Merkel in den Mittelpunkt und verleiht ihr durch die Statik Stabilität und den Eindruck von seriöser Neutralität. Vor dieser visuell zurückgenommenen Inszenierung heben sich die sprachliche und gestische Äußerung markant ab. Sie entfalten über die lautliche, gestische Dynamik und die markierte Hinwendung von Merkels Oberkörper zum Publikum eine affektive Dynamik, welche sich gemeinsam mit den kritischen Zwischenrufen aus den Reihen der Parlamentarier*innen als eine plötzlich ansteigende, aufwallende Bewegung entfaltet, die im Verlauf von Merkels Gegenrede kontinuierlich heruntergefahren wird. D.h. am Ende der Entgegnung Merkels auf den Zwischenruf ist die Gestalt geschlossen, die affektive Dynamik wieder auf dem Niveau, das sie zuvor hatte.

4.2 Die Poiesis des Filme-Sehens

Audiovisuelle Bilder formen das ästhetische Erleben der Rezipient*innen als spezifische Form ästhetischer Erfahrung. Das ›Verstehen‹ audiovisueller Bil-

der, das Verstehen von Sprache-Sprechen in audiovisuellen Medien ist daher affektiv durchtränkt. Audiovisuelle Bilder modulieren die Erfahrungsräume ihrer Betrachter*innen, gleichzeitig entstehen diese als Bewegungsbilder, als Bewegungsfigurationen, als multidimensionale Gestalten erst in der Erfahrung, im Betrachten der Zuschauenden. Erst im Erleben der Betrachter*innen entstehen die Bewegungsbilder als Wahrnehmungsgestalten (vgl. Abb. 5). Das Sehen und Hören audiovisueller Bilder ist daher ein Tun, eine Produktion der Zuschauenden. Kappelhoff fasst diese »rezeptive Aneignung audiovisueller Bewegtbilder« als kulturelle Praxis, als eine Poiesis des Filme-Sehens:

Damit wird die Rezeption selbst als eine Form der Produktion, die rezeptive Aneignung audiovisueller Bewegtbilder als ein genuiner Akt des Herstellens avisiert. Wir sprechen deshalb von der Rezeption audiovisueller Bilder als einer Poiesis des Filme-Sehens. (Kappelhoff 2018: 10)

Das Sehen und Hören audiovisueller Bilder gerät damit als eine Form des Machens in den Blick. Dieses Sehen und Hören wird zwar durch die Gestaltung der audiovisuellen Formate orchestriert, als Bewegungsbilder benötigen sie jedoch die Körper der Zuschauenden. Um Sobchack zu paraphrasieren: Die Zuschauenden »leihen« den Bildern ihre Körper (Sobchack 1992). Das Sehen und Hören audiovisueller Bilder ist darum kein passives Überflutet-Werden, sondern eine mediale Praxis die, wie u. a. der Siegeszug der sozialen Medien zeigt, ein äußerst erfolgreiches Instrument der Gemeinschaftsbildung ist. Von der Wirkmächtigkeit dieses Medienkonsums als kultureller Praxis zeugen etwa die immensen *Follower*innen*-Zahlen von TikTok-Videos der neuen Rechten oder von Schminkanleitungen, die mit Politinformationen oder True-Crime Stories verbunden werden.

In diesem Sinne lässt sich die Poiesis des Filme-Sehens nicht als passive Rezeption audiovisueller Bilder sondern als eine kulturelle Praxis fassen, die sich als geteilte ästhetische Erfahrung in zahllosen Geschmacksgemeinschaften manifestiert – ob dies die Anhänger*innen trashiger Horrorfilme oder Netflix-User, die Fans von Kochvideos, Polit-Talks, Serien, Videospiele, Fußballspielen oder Comedies oder die Millionen *Follower*innen* von Videoclips in den sozialen Medien sind. Es ist diese Praxis des Medienkonsums, die zu völlig neuen transnationalen Gemeinschaftsbildungen führt, wie sie sich nicht nur an globalen *Follower*innen*-Strömen bestimmter *Influencer*innen* nachvollziehen lässt. In den Poetologien der sich ständig verändernden medialen Formen zeigt sich eine kulturelle Praxis, die sich über die präzise Rekonstruktion der

audiovisuellen Orchestrierung als Raum spezifischer ästhetischer Erfahrung rekonstruieren und damit auch kritisch befragen lässt.¹⁶

4.3 Generierung, Modulation und Mobilisierung von Zuschauer*gefühlen

Mit dem Rekurs auf das Konzept der Ausdrucksbewegung lässt sich eine Medienpraxis beschreiben, die Affekte und Gefühle¹⁷ einer breiten Zuschauer*innenschaft generiert, moduliert und mitunter auch sehr gezielt mobilisiert (Kappelhoff 2008, 2016). Die Poetik der Ausdrucksbewegung betrifft dabei einen grundlegenden Darstellungsmodus westlicher Kunst, der eng verbunden ist mit der Kulturgeschichte des Theaters der Empfindsamkeit und des Melodramas im Film (Kappelhoff 2004a: 182, 2014; Müller/Kappelhoff 2018: 62f.). Die klassische Filmtheorie sieht in der Ausdrucksbewegung eine Form der Poetik, die die Generierung und Modulation von Affekten eines breiten anonymen Publikums erlaubt und die darin eine Form kollektiver Subjektivität sieht. Die Erfahrung intensiver Empfindungen – Horror, Spannung, sentimentale Gefühle – ist in der zeitgenössischen Unterhaltungskultur zum Selbstzweck geworden (Müller/Kappelhoff 2018: 165). So hat die Generierung und Modulation von Gefühlen längst auch alltägliche Formen medialer Gestaltung erreicht, seien es Nachrichtenformate, Reportagen oder die Dokumentation parlamentarischer Debatten. Transdisziplinäre Analysen audiovisueller Multimodalität haben zeigen können, dass in ganz verschiedenen medialen Formaten audiovisuelle Ausdrucksbewegungen Bewegungsbilder formen, die als multidimensionale Erfahrungsgestalten den Wahrnehmungsprozess auch noch so »objektiv« sich präsentierender medialer Formate affektiv grundieren. Die Wirkmacht dieser audiovisuellen Bilder liegt dann in der unmerklichen Erzeugung und Veränderung des affektiven Erlebens der Zuschauer*innen. *Was* gezeigt wird, entsteht aus dem *Wie* der audiovisuellen Inszenierung.

Bezogen auf die mediale Inszenierung von Bundestagsreden durch ein Parlamentsfernsehen, wird hier durch eine eher zurückhaltende Inszenie-

16 Gegenstand der Kolleg-Forschungsgruppe »Cinepoetics« an der Freien Universität Berlin waren solche Formen der Poiesis des Filme-Sehens, die über die Rekonstruktion von Poetologien audiovisueller Bilder greifbar wurden: <https://www.cinepoetics.fu-berlin.de/>. Viele dieser Studien sind in der Buchreihe *Cinepoetics* bei De Gruyter im Open Access Format erschienen.

17 In Anlehnung an Damasio (1999) unterscheide ich terminologisch Affekte (»Feeling«) und Gefühle (»Emotion«), wobei sich *Feeling* auf Stimmungen, Atmosphären und *Emotion* auf diskrete Gefühle bezieht (vgl. hierzu auch Müller/Kappelhoff 2018: 55).

rung (statische Einstellung, keine oder wenig Schnitte, keine oder wenige Kamerabewegung) eine Art ›objektiver‹ neutraler Perspektive hergestellt, die im scharfen Gegensatz zu Videoclips auf Plattformen wie TikTok oder Instagram stehen, die mit zusätzlichen visuellen Effekten, Spracheinblendungen, überblendeten Bildeffekten, einer Zerstückelung und Beschleunigung von gesprochener Sprache durch hohe Schnittfrequenzen und Erhöhung der Bildgeschwindigkeit arbeiten – und die sich als sehr kurze Zeitformen (TikTok: 5 Sekunden bis 10 Minuten, Instagram: 3–90 Sekunden) etabliert haben, während die Dokumentation politischer Reden selten kürzer als fünf Minuten ist. Wie sich diese Beschleunigung, Fragmentierung von Sprache-Sprechen auf die Generierung, Modulation und Mobilisierung von Zuschauer Gefühlen auswirkt bleibt ein drängendes Desiderat kritischer kulturwissenschaftlicher Sprachgebrauchs- und Medienforschung.

5 Zusammenfassung

Audiovisuelle Medien in verschiedensten Ausprägungen strukturieren in stetig steigendem Maße den Gebrauch und die Wahrnehmung von Sprache. Da die Beobachtung sprachlicher Multimodalität eng mit der Möglichkeit der audiovisuellen Fixierung flüchtiger Momente des Sprache-Sprechens verbunden ist, ist der analytische Blick auf sprachliche Multimodalität immer auch ein Blick durch ein audiovisuelles Medium. Dies betrifft offensichtlich nicht nur die Videokameras der Forschung, sondern den sprachlichen Alltag unserer Gesellschaft. Hier hat der vorliegende Beitrag angesetzt und für eine medienästhetische Perspektive auf den Sprachgebrauch in audiovisuellen Medien argumentiert.

Eine medienästhetische Perspektive auf die Multimodalität des Sprechens in audiovisuellen Medien einzunehmen, zielt auf eine Beschreibung von Praktiken des Medienkonsums und ihrer Attraktivität ab. Sie verbindet Kappelhoffs Theorie filmischen Denkens, »Die Poiesis des Filme-Sehens«, mit einer Theorie der Multimodalität des Sprechens. Für beide sind Zeitlichkeit, leibliche Verschränkung, Interkorporalität und Interaffektivität als Bewegungserfahrung grundlegend:

Interactive movement as reflexive affection is a key element from which the gestural as well as the cinematic gestalten ultimately emanate. Conceiving of multimodal face-to-face communication as temporally orchestrated and

as inherently interaffective is what connects gestures and spoken language with film. (Müller/Kappelhoff 2018: 155)

Mit anderen Worten, ›Sprache-Sprechen, Hören, Sehen, Fühlen, – Erfahren‹ formuliert eine transdisziplinäre Perspektive, deren theoretischer und methodologischer Angelpunkt das Konzept der körperlich-filmischen Ausdrucksbewegung ist. Es fungiert als theoretischer und methodologischer Bezugspunkt für die Analyse audiovisueller und sprachlicher Multimodalität. Über die Gestaltung audiovisueller Bilder als Ausdrucksbewegungen formen diese eine ästhetische Erfahrung der Zuschauer*innen, die diesem Diskurs audiovisueller Bilder nicht etwa passiv ausgeliefert sind, sondern in deren Wahrnehmung diese sich erst als Bewegungsbilder entfalten. Die Perspektive der Zuschauer*innen für die Multimodalität des Sprechens in audiovisuellen Medien ernst zu nehmen, heißt dann, die Wahrnehmungsgestalt der filmischen Ausdrucksbewegung als analytischen Ausgangspunkt anzunehmen.¹⁸ Diese multidimensionalen Wahrnehmungsgestalten können in der weiteren Analyse auf ihre Orchestrierung hin, d.h. mit Bezug auf sprachliche und gestische Figurationen, Bildraum, Kameraeinstellung, Montage etc. beschrieben werden.

Die Position der Zuschauer*innen einzunehmen hat auch erhebliche Folgen für die ethnographische und experimentelle Erforschung multimodaler Kommunikation und Interaktion, denn es impliziert die Forderung, die Position der Forschenden, der Analytiker*innen an den Bildschirmen systematisch mit in den analytischen Prozess einzubeziehen. Aus medienwissenschaftlicher Sicht gilt es dabei einmal mehr zu beachten, dass »Medien [...] keineswegs bloße Mittel der Kommunikation gegebener Sachverhalte [sind]; vielmehr sind sie technische Erweiterungen menschlicher Wahrnehmung; sie verändern die apriorischen Bedingungen des Verstehens, Urteilens und Imaginierens« (Kappelhoff 2018: 9). Dass dies für die Kamera als analytisches Werkzeug für das Studium multimodaler Interaktionen genauso gilt wie für die Bilder von Handycameras, die auf *Social Media* gepostet werden, versteht sich von selbst.

Im Zentrum einer medienästhetischen Perspektive auf den Sprachgebrauch in audiovisuellen Medien steht die systematische Unterscheidung

18 »In our research on the poetology of audiovisual images we have developed an analytical model that is comprehensive and cultural-historically oriented: expressive movement (vgl. Kappelhoff 2004a). It captures the affective movement of the viewer as a bodily sensation of the unfolding flow of cinematic movement-images« (Müller/Kappelhoff 2018: 59).

der Multimodalität des Sprechens und der audiovisuellen Multimodalität. Erst diese Unterscheidung ermöglicht es, ihre Gemeinsamkeiten und Verschränkungen als Zeitbilder, Bewegungsbilder sowie den geteilten Modus ästhetischer Erfahrung als Verschränkung von Körpern im Rhythmus und Dynamik der Bewegungserfahrung herauszuarbeiten. Wechselseitiges Verstehen, Intersubjektivität speist sich zuallererst aus der Zwischenleiblichkeit, dem körperlichen Miteinander, das in der Qualität der Bewegung, ob Handgeste oder audiovisuelles Bild, immer schon eine affektive Tönung hat.

In diesem Sinne ist Sprache-Sprechen grundiert in einer affektiven Bewegungserfahrung von Körpern, deren Sprechen Körperbewegung *ist* (Horst et al. 2014). Auch den Konsum audiovisueller Medien kennzeichnet ein Modus ästhetischer Erfahrung, für den die Interkorporalität als interaffektive Erfahrung von Bewegungsbildern grundlegend ist. Als kulturelle Praxis wird dieser Erfahrungsraum durch vielfältige mediale Ökologien gestaltet, die jeweils zu partikulären Gemeinschaftsbildungen auf der Seite der Medienkonsument*innen führen.

Denkt man also den Gegenstand sprachwissenschaftlicher Forschung vom Sprache-Sprechen und den Gegenstand medienlinguistischer Forschung von der audiovisuellen Multimodalität her, offenbaren sich erstaunliche Parallelen, und es entsteht die Möglichkeit eine transdisziplinäre (film- und sprachwissenschaftliche) Theorie und Methode zu erarbeiten, die versucht, der Komplexität des Phänomens und der besonderen Attraktion audiovisueller Medien und des dort eingebetteten multimodalen Sprachgebrauchs etwas näher zu kommen. Diese Perspektive einzunehmen »entgrenzt« den Fokus sprachwissenschaftlicher Betrachtung von einem körperlosen kontext-entbundenen Sprachsystem zu einem kontext-gebundenen, verkörperten, immer zeitlich zu denkenden Sprechen, einem *Languaging* (vgl. Larsen-Freeman 2003), das immer schon multimodal ist. Es »entgrenzt« auch die Perspektive linguistischer Forschung um eine film- und medienwissenschaftliche Perspektive auf den Sprachgebrauch in audiovisuellen Medien. Wie sich dieser Ansatz methodisch umsetzen lässt, dokumentieren wir im vorliegenden Band mit einer medienästhetischen Fallstudie aus einem laufenden Forschungsprojekt zu parlamentarischen Reden (Kindler-Mathôt et al. i. d. Bd.). Für das Sprechen in audiovisuellen Medien bedeutet dies, dass die Begriffe der Multimodalität des Sprechens und der audiovisuellen Multimodalität als analytische Kategorien fungieren, die an verschiedene artikulatorische Modi gebunden sind, die in der Analyse die komplexe dynamische Orchestrierung der sprachlichen und der audiovisuellen Multimodalität greifbar machen. Nimmt man jedoch conse-

quent die Perspektive der Zuschauer*innen ein, dann haben wir es mit einem einzigen, spezifischen Modus ästhetischer Erfahrung zu tun. Ein Sprechen, das uns audiovisuelle Bilder vorführen, wird eben nicht nur gehört, sondern gleichzeitig gesehen und gefühlt. Es zeigt sich als Aspekt einer multidimensionalen Erfahrungsgestalt und ist in diesem Sinne ein Modus.

6 Kulturwissenschaftliche Erforschung des Sprachgebrauchs als Medienkritik?

Eine kulturwissenschaftliche Sprachgebrauchsforschung tut gut daran, ihre eigenen historischen Bezüge nicht aus dem Blick zu verlieren. Wie die Bezugnahme auf Bühlers Theorie gezeigt hat, können altbekannte Grundannahmen in einer interdisziplinär orientierten wissenschaftshistorischen Kontextualisierung neu betrachtet große Aktualität gewinnen.¹⁹ Auf diese Weise erschließt sich eine kulturhistorische Tiefe und Weite sprachtheoretischer Reflexionen, die aktuellen Diskursen um *Embodiment*, Interkorporalität, um Sprache als kulturelle Praxis nicht nur eigene kulturhistorische Bezüge und Kontexte aufzeigt, sondern auch die grundsätzliche Historizität geistes- und kulturwissenschaftlicher Diskurse unterstreicht (vgl. etwa Meyer/Streck/Jordan 2017).

Sprachgebrauchsforschung kulturwissenschaftlich gedacht versucht sich dann daran, die disziplinäre Engführung der Linguistik als Fach hin auf andere kulturwissenschaftliche Fächer zu öffnen – und knüpft damit unter anderem an die eigene philologische Tradition der Sprachwissenschaft an, in der Sprachbetrachtung grundsätzlich verbunden war mit Literaturwissenschaft und historischer Sprach- und Literaturforschung. Sie geht über die philologische Tradition hinaus, wenn sie kulturwissenschaftliche Sprachforschung über die Fächergrenzen öffnet. In diesem Sinne verstehe ich das Credo dieses Bandes – Sprache entgrenzen auch als ein Plädoyer dafür Sprache und Sprachwissenschaft historisch und kulturwissenschaftlich zu denken.

Im Lichte der gesellschaftlichen, sozialen, politischen, kulturellen Bedeutung ihres Gegenstandsbereiches ›Sprache‹ hat sich die Linguistik früh hin

19 So argumentiert etwa Jäger (2010: 5): »Vor allem das lange Ausblenden ausdrucks- und zeichentheoretischer Fragestellungen aus dem sprach- und zeichentheoretischen Diskurs nach 1945, die gegenwärtig etwa im Horizont der Gestur-Forschung wieder aufgegriffen und entfaltet werden, hat wohl hier eine wesentliche Ursache.«

zu anderen Disziplinen geöffnet. Die Bindestrich-Linguistiken entstanden, und Neugründungen erfreuen sich weiterhin großer Beliebtheit: Sozio-, Psycho-, Historio-, Diskurs-, Politico-, und nicht zuletzt die Medien-Linguistik. Diese Bindestrich-Linguistiken haben immer auch die Grenzen der Sprache und mitunter auch des Sprache-Sprechens neu ausgelotet, haben auf neue gesellschaftliche und wissenschaftliche Entwicklungen reagiert und deutlich gemacht, dass die Linguistik hier Wesentliches beitragen kann. Eine Erweiterung der Sprachgebrauchsforschung um einen transdisziplinären (filmwissenschaftlich-linguistischen) Zugang zur Multimodalität des Sprechens in audiovisuellen Medien, wie ich sie in diesem Beitrag vorgeschlagen habe, stellt einen Versuch dar, eine kulturwissenschaftliche Öffnung der Sprachgebrauchsforschung theoretisch und methodisch zu leisten.

Gerade für medienlinguistische Forschung scheint es dabei von außerordentlicher Bedeutung zu sein, medienwissenschaftliche Expertise mit in den Blick zu nehmen. Denn ein tieferes Verständnis der Besonderheiten audiovisueller Medien eröffnet neue Möglichkeiten kritischer Reflexionen des Sprachgebrauchs in audiovisuellen Medien. Es bedeutet, die Forderung Marshall McLuhans »the medium is the message« (McLuhan 1964) noch einmal in den Blick zu nehmen.²⁰ Vielleicht gerade, weil diese mittlerweile zum selbstverständlichen Teil eines *Common Sense* wissenschaftlicher und alltagsweltlicher Medienreflexion geworden ist, und weil zugleich eine sich explosionsartig diversifizierende Medienlandschaft die Gesellschaft vor enorme Herausforderungen stellt – zum Teil mit erheblichen sozialen und politischen Effekten.

Nun wäre es zu kurz gegriffen, diese Explosion lediglich als Form unendlicher Vervielfältigung audiovisueller Repräsentationen des Sprachgebrauchs in verschiedenen Kontexten zu beschreiben. Lassen sich doch selbst auf Sachinformationen spezialisierte mediale Formate wie Nachrichten und Reportagen nicht auf die schlichte Kommunikation von Sachverhalten reduzieren – auch nicht, wenn etwa in Interviews oder Talkshows Sprecher*innen unmittelbar zur Rede kommen. Immer haben wir es mit der Ko-Präsenz von Ausdruck, Darstellung und sozialer Adressierung des Sprechens und Gestikulierens, immer haben wir es mit einer spezifischen medialen Ökologie, einer medialen Orchestrierung dieses Sprache-Sprechens zu tun, die sich etwa im selbstgedrehten YouTube- oder TikTok-Video vom Interview im Rahmen der Tages-themen in ihrer medialen Verfasstheit deutlich unterscheiden.

20 Vgl. auch Müller/Kappelhoff 2018, »Introduction«.

Angesichts dieser medialen Diversifizierung audiovisueller Formate gewinnt Walter Benjamins Klassiker »Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit« einmal mehr bedeutende Aktualität (Kappelhoff 2018: Kap.1). Mit Benjamin gerät nämlich die Frage in den Blick, wie diese verschiedenen medialen Ökologien das Sehen, Hören, Fühlen ihrer Rezipient*innen modellieren, was sie ihnen buchstäblich zu sehen, zu hören und zu fühlen geben: »Nimmt man den Gedanken ernst, dass unser Sinnesapparat ein Produkt der Geschichte der Medien menschlicher Wahrnehmung ist, dann ist das Sehen, Hören, Fühlen zunächst eine Frage des Mediengebrauchs« (Kappelhoff 2018: 10). Damit verbindet sich die Frage nach der Verfasstheit audiovisueller Formate mit der Praxis eines Medienkonsums, in dessen Vollzug das Sprechen auf YouTube oder in den abendlichen TV-Nachrichten den Rezipient*innen nicht einfach begegnet, sondern als Form des Mediengebrauchs die Wahrnehmung, das Denken und Fühlen der Zuschauer*innen und Zuhörer*innen rahmt, strukturiert und modelliert. Zugleich haben wir es »in der Aneignung audiovisueller Bewegtbilder« mit einer kulturellen Praxis zu tun »in der Menschen ihre subjektive Wahrnehmungswirklichkeit als Teil einer gemeinsam geteilten Welt zu fassen suchen« (Kappelhoff 2018: 10).

Als kulturelle Praxis zeigt sich diese Rezeption audiovisueller Bilder in den verschiedensten, häufig global formierten, Geschmacksgemeinschaften deren Zahlen sich in Klicks und zum Teil in Millionen von *Follower*innen* manifestieren. Diese neuen Gebrauchskontexte stellen (nicht nur) die linguistische Forschung vor erhebliche Herausforderungen. Sie verlangen nach einer theoretischen und analytischen Durchdringung dieser Veränderungen. Sie erfordern nicht zuletzt die Arbeit an einer Kompetenz zur Medienkritik.

Die hier vorgestellte medienästhetische Perspektive auf die Multimodalität des Sprechens in audiovisuellen Medien unterbreitet hierzu einen transdisziplinären Vorschlag. Wenn sie das Sprache-Sprechen als multimodales Ereignis als Fühlen, Sehen, Empfinden der Betrachter*innen adressiert, geht es ihr um die unmerklichen Modulationen der Zuschauerempfindungen mit ihren gemeinschaftsbildenden Effekten. Es geht darum, diese theoretisch und empirisch zu greifen, es geht um Grundlagen für eine kritische Sprachgebrauchs- und Medienforschung.

Literaturangaben

- Armstrong, David. F./William C. Stokoe/Sherman E. Wilcox (1995): *Gesture and the nature of language*, Cambridge: Cambridge University Press.
- Balázs, Bela (1924): *Der sichtbare Mensch oder die Kultur des Films*, Leipzig: Deutsch-Österreichischer Verlag.
- Bressem, Jana/Silva H. Ladewig (2011): Rethinking gesture phases: Articulatory features of gestural movement? In: *Semiotica* 184:1/4, 53–91. <https://doi.org/10.1515/semi.2011.022>
- Bühler, Karl (1968 [1933]): *Ausdruckstheorie. Das System an der Geschichte aufgezeigt*, Jena: Fischer.
- Bühler, Karl (1982 [1934]): *Sprachtheorie. Die Darstellungsfunktion der Sprache*, Jena: Fischer.
- Damasio, Antonio R. (1999): *The Feeling of What Happens. Body and Emotion in the Making of Consciousness*, New York: Harcourt Brace.
- Deleuze, Gilles (2008a): *Cinema 1. The Movement Image*, London: Continuum.
- Deleuze, Gilles (2008b): *Cinema 2. The Time Image*, London: Continuum.
- Eisenstein, Sergej M. (2010 [1926]): Bela Forgets the Scissors. In: Taylor, Richard (Hg.), *Writings, 1922–1934. Sergei Eisenstein Selected Works*, London: I.B. Tauris, 77–81.
- Ekman, Paul (1972): *Emotions in the Human Face*, New York: Pergamon.
- Ekman, Paul (2006): *Darwin and Facial Expression: A Century of Research in Review*, Cambridge: Malor Books.
- Eschbach, Achim (Hg.) (1984): *Bühler-Studien. Band 1 und 2*, Frankfurt a.M.: Suhrkamp.
- Flusser, Vilem (1991): *Gesten. Versuch einer Phänomenologie*, Düsseldorf: Bollmann.
- Fuchs, Thomas (2017): Intercorporeality: Emerging socialities in interaction. In: Meyer, Christian/Jürgen Streeck/Scott Jordan (Hg.), *Intercorporeality and interaffectivity*, Oxford: Oxford University Press, 3–25.
- Fuchs, Thomas/Hanne De Jaegher (2009): Enactive intersubjectivity: Participatory sense-making and mutual incorporation. In: *Phenomenology and the Cognitive Sciences* 8, 465–486.
- Greifenstein, Sarah (2020): *Tempi der Bewegung – Modi des Gefühls. Expressivität, heitere Affekte und die Screwball Comedy*, Berlin/Boston: De Gruyter.
- Greifenstein, Sarah/Dorothea Horst/Thoma Scherer/Christina Schmitt/Hermann Kappelhoff/Cornelia Müller (Hg.) (2018): *Cinematic Metaphor in*

- Perspective. Reflections on a Transdisciplinary Framework*, Berlin/Boston: De Gruyter Mouton.
- Horst, Dorothea (2018): *Meaning Making and Political Campaign Advertising. A Cognitive-Linguistic and Film-Analytical Perspective on Audiovisual Figurativity*, Berlin/Boston: De Gruyter.
- Horst, Dorothea/Franziska Boll/Christina Schmitt/Cornelia Müller (2014): Gesture as interactive expressive movement: Inter-affectivity in face-to-face communication. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressemer (Hg.), *Body – Language – Communication. Handbücher zur Sprach- und Kommunikationswissenschaft*, Berlin/Boston: De Gruyter Mouton, 2112–2125.
- Jäger, Ludwig (2010): Sprache als Organon. Karl Bühlers Beitrag zur Begründung der modernen Sprachwissenschaft. In: *Sprache und Literatur* 41:105, 3–17.
- Kappelhoff, Hermann (2004a): *Matrix der Gefühle. Das Kino, das Melodrama und das Theater der Empfindsamkeit*, Berlin: Vorwerk 8.
- Kappelhoff, Hermann (2004b): Unerreichbar, unberührbar, zu spät. Das Gesicht als kinematografische Form der Erfahrung. In: *montage/av*, 13:2, 29–53.
- Kappelhoff, Hermann (2006): Dauer der Empfindung. Von einer spezifischen Bewegungsdimension im Kino. In: Bischof, Margrit/Claudia Feest/Claudia Rosiny (Hg.), *E_motion*, Münster: LIT, 205–219.
- Kappelhoff, Hermann (2008): Zuschauergefühl. Die Inszenierung der Empfindung im dunklen Raum des Kinos. In: Spiekermann, Geraldine/Beate Söntgen (Hg.), *Tränen*, München: Fink, 195–206.
- Kappelhoff, Hermann (2016): *Genre und Gemeinsinn: Hollywood zwischen Krieg und Demokratie*, Berlin/Boston: De Gruyter.
- Kappelhoff, Hermann (2018): *Kognition und Reflexion. Zur Theorie filmischen Denkens*, Berlin/Boston: Walter de Gruyter.
- Kendon, Adam (1970): Movement coordination in social interaction: Some examples described. In: *Acta Psychologica* 32, 101–125. [https://doi.org/10.1016/0001-6918\(70\)90094-6](https://doi.org/10.1016/0001-6918(70)90094-6).
- Kendon, Adam (1972): Some relationships between body motion and speech. An analysis of an example. In: Siegman, Aron Wolfe/Benjamin Pope (Hg.), *Studies in Dyadic Communication*, New York: Pergamon Press, 177–210.
- Kendon, Adam (1980): Gesture and Speech: Two aspects of the process of utterance. In: Kendon, Adam (Hg.), *Nonverbal Communication and Language*, The Hague: Mouton, 207–227.

- Kendon, Adam (1990): Spatial organization in social encounters. The F-formation system. In: Kendon, Adam (Hg.), *Conducting interaction. Patterns of behavior in focused encounters*, Cambridge: Cambridge University Press, 209–238.
- Kendon, Adam (2004): *Gesture: Visible action as utterance* (1. Aufl.), Cambridge: Cambridge University Press.
- Kennedy, Adam (2013): Laban based analysis and notation of body movement. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Sedinha Teßendorf (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. Handbücher zur Sprach- und Kommunikationswissenschaft: Bd. 38.1*, Berlin/Boston: De Gruyter Mouton, 941–957.
- Krumhuber, Eva/Susanne Kaiser/Kappas Arvid/Klaus Scherer (2013): Body and speech as expression of inner states. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Sedinha Teßendorf (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. Handbücher zur Sprach- und Kommunikationswissenschaft: Bd. 38.1*, Berlin/Boston: De Gruyter Mouton, 551–564.
- Laban, von Rudolf (1956): *Principles of Dance and Movement Notation*, London: Macdonald & Evans.
- Laban, von Rudolf/Frederick C. Lawrence (1947): *Effort: Economy in body movement*, London: MacDonald and Evans.
- Larsen-Freeman, Diane (2003): *Teaching language. From grammar to grammaring*, Boston, MA: Heinle/Cengage.
- Loenhoff, Jens (2017): Intercorporeality as a Foundational Dimension of Human Communication. In: Meyer, Christian/Jürgen Streeck/J. Scott Jordan (Hg.), *Intercorporeality: emerging socialities in interaction*, Oxford: Oxford University Press, 25–49.
- McLuhan, Marshall (1964): *Understanding Media. The Extensions of Man*, New York: New American Library.
- McNeill, David (1985): So you think gestures are nonverbal? In: *Psychological Review* 92:3, 350–371.
- McNeill, David (1992): *Hand and mind: What gestures reveal about thought*, Chicago: University of Chicago Press.
- Meyer, Christian/Jürgen Streeck/J. Scott Jordan (Hg.) (2017): *Intercorporeality: emerging socialities in interaction*, Oxford: Oxford University Press.
- Merleau-Ponty, Maurice (1968 [1964]): *The Visible and the Invisible*, Evanston: Northwestern University Press.

- Merleau-Ponty, Maurice (2005 [1945]): *Phenomenology of Perception*, London: Routledge.
- Mondada, Lorenza (2013): Multimodal interaction. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Sedinha Teßendorf (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. Handbücher zur Sprach- und Kommunikationswissenschaft: Bd. 38.1*, Berlin/Boston: De Gruyter Mouton, 577–589.
- Müller, Cornelia (1998): *Redebegleitende Gesten. Kulturgeschichte, Theorie, Sprachvergleich*, Berlin: Berlin Verlag.
- Müller, Cornelia (2008): What gestures reveal about the nature of metaphor. In: Cienki, Alan/Cornelia Müller (Hg.), *Metaphor and gesture*, Amsterdam/Philadelphia: John Benjamins, 219–245.
- Müller, Cornelia (2010): Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive. In: *Sprache und Gestik. Sonderheft der Zeitschrift Sprache und Literatur* 41:1, 37–68.
- Müller, Cornelia (2015): Using gestures with speech: Variable cognitive-semantic and pragmatic relations. In: *Vestnik of Moscow State Linguistic University* 6:717, 452–466.
- Müller, Cornelia (2016): From mimesis to meaning: A systematics of gestural mimesis for concrete and abstract referential gestures. In: Zlatev, Jordan/Goran Sonesson/Piotr Konderak (Hg.), *Meaning, Mind and Communication. Explorations in Cognitive Semiotics*, Frankfurt a.M.: Peter Lang, 211–226
- Müller, Cornelia (2018): Gesture and sign: Cataclysmic break or dynamic relations? In: *Frontiers in Psychology* 9:1651.
- Müller, Cornelia (2024a): Mimesis as as-if action. In: Zywczyński, Przemysław/Slawomir Waciewicz/Monika Boruta-Żywicyńska/Johan Blomberg (Hg.), *Perspectives on pantomime: evolution, development, interaction*, Amsterdam/Philadelphia: John Benjamins, 220–243.
- Müller, Cornelia (2024b): A toolbox for methods of gesture analysis. In: Cienki, Alan (Hg.), *The Cambridge Handbook of Gesture Studies*, Cambridge: Cambridge University Press, 182–216.
- Müller, Cornelia/Ulrike Bohle (2007): Das Fundament fokussierter Interaktion: Zur Vorbereitung und Herstellung von Interaktionsräumen durch körperliche Koordination. In: Schmitt, Reinhold (Hg.), *Koordination: Analysen zur multimodalen Interaktion*, Tübingen: Narr, 129–165.
- Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Sedinha Teßendorf (Hg.) (2013): *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. Handbücher zur*

- Sprach- und Kommunikationswissenschaft: Bd. 38.1*, Berlin/Boston: De Gruyter Mouton.
- Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressemer (Hg.) (2014): *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction. Handbücher zur Sprach- und Kommunikationswissenschaft: Bd. 38.2*, Berlin/Boston: De Gruyter Mouton.
- Müller, Cornelia/Alan Cienki (2009): When speech and gesture come together. Forms of multimodal metaphor in the use of spoken language. In: Forceville, Charles J./Eduardo Urios-Aparisi (Hg.), *Multimodal Metaphor*, Berlin/Boston: De Gruyter Mouton, 299–332.
- Müller, Cornelia/Hermann Kappelhoff (2018): *Cinematic Metaphor. Experience – Affectivity – Temporality*, Berlin/Boston: De Gruyter Mouton.
- Münsterberg, Hugo (2002 [1916]): The Photoplay – A Psychological Study. In: Langdale, Allan (Hg.), *Hugo Münsterberg on Film. The Photoplay – A Psychological Study and Other Writings*, London: Routledge, 45–162.
- Neisser, Ulric (1976): *Cognition and Reality: Principles and Implications of Cognitive Psychology*, San Francisco: W.H. Freeman & Co.
- Oben, Bert/Geert Brône (2015): What You See Is What You Do. On the Relationship Between Gaze and Gesture in Multimodal Alignment. In: *Language and Cognition* 7:4, 546–562.
- Oben, Bert/Geert Brône (2016): Explaining Interactive Alignment. A Multimodal and Multifactorial Account. In: *Journal of Pragmatics* 104, 32–51.
- Plessner, Helmuth (1982 [1957]): Ausdruck und menschliche Natur. In: Dux, Günter/Odo Marquard/Elisabeth Ströker (Hrsg), *Helmuth Plessner. Gesammelte Schriften* 10,7, Frankfurt a.M.: Suhrkamp.
- Plessner, Helmuth (1982 [1925]): Die Deutung des mimischen Ausdrucks. Ein Beitrag zur Lehre vom Bewußtsein des anderen Ichs [in Zusammenarbeit mit Buytendijk, F. J. J.]. In: Dux, Günter/Odo Marquard/Elisabeth Ströker (Hrsg), *Helmuth Plessner. Gesammelte Schriften*, Frankfurt a.M.: Suhrkamp, 67–129.
- Schegloff, Emmanuel A. (1984): On some gestures' relation to talk. In: Atkinson, Maxwell J./John Heritage (Hg.) *Structures of Social Action. Studies in Conversation Analysis*, Cambridge: Cambridge University Press, 266–296.
- Scherer, Thomas J. (2024): *Inszenierungen zeitgenössischer Propaganda. Kampagnenfilme im Dienste des Gemeinwohls*, Berlin/Boston: De Gruyter.
- Schmitt, Chistina (2020): *Wahrnehmen, fühlen, verstehen, Metaphorisieren und audiovisuelle Bilder*, Berlin/Boston: De Gruyter.

- Schmitt, Reinhold (2005): Zur multimodalen Struktur von turn-taking. In: *Gesprächsforschung – Online Zeitschrift zur verbalen Interaktion* 6, 17–61.
- 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 Mertzluft/Christian Meyer/Miriam Morek/Frank Oberzaucher/Jörg Peters/Uta Quasthoff/Wilfried Schütte/Anja Stukenbrock/Susanne Uhmann (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10, 353–402. Abrufbar unter: <https://www.gespraechsforschung-online.de/fileadmin/dateien/heft2009/px-gatz.pdf>
- Seyfeddinipur, Mandana (2006): *Speech-dysfluencies. Interrupting speech and gesture*, Nijmegen: Radboud University.
- Simmel, Georg (1993): Die ästhetische Bedeutung des Gesichts. In: Kramme, Rüdiger/Alessandro Cavalli (Hg.), *Aufsätze und Abhandlungen, 1901–1908*. 1. Aufl., Bd. 1, Berlin: Suhrkamp, 36–42.
- Stern, Daniel N. (1985): *The Interpersonal World of the Infant. A View from Psychoanalysis and Developmental Psychology*, London: Routledge.
- Sobchack, Vivian (1992): *The Address of the Eye. A Phenomenology of Film Experience*, Princeton: Princeton University Press.
- Streeck, Jürgen (2009): *Gesturecraft. The manufacture of meaning*, Amsterdam/Philadelphia: John Benjamins.
- Streeck, Jürgen/UlrikeHartge (1992): The contextualization of language. In: Auer, Peter/Aldo Di Luzio (Hg.), *Previews: Gestures at the transition place*, Amsterdam/Philadelphia: John Benjamins, 135–158.
- Wundt, Wilhelm M. (1975 [1900–1925]): *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte: Vol. 2. Die Sprache*, Alfred Kröner Verlag.

Cornelia Müller, Prof. Dr.
 Europa-Universität Viadrina
 Große Scharrnstraße 59
 15230 Frankfurt (Oder)
 cmueller@europa-uni.de

Sprechen in audiovisuellen Medien

Systematik für eine medienästhetische Analyse multimodaler Kommunikation

Clara Kindler-Mathôt, Cornelia Müller, Jana Katharina Junge, Katerina Papadopoulou

Abstract *Der Beitrag illustriert anhand einer Fallstudie aus dem Korpus des DFG/NCN Forschungsprojektes »Multimodal Stancetaking«, wie sich die Multimodalität des Sprechens in audiovisuellen Medien untersuchen lässt. Ansatzpunkt der medienästhetischen (filmwissenschaftlich-linguistischen) Analyse ist die systematische Berücksichtigung der Verschränkung körperlich-sprachlicher und audiovisueller Ausdrucksformen, die das Material einfordert. Sprechen im medialen Raum entfaltet sich in audiovisuell orchestrierten Gestalten, die die Zuschauer*innen an den Bildschirmen als Wahrnehmungsgestalten sehen, hören und fühlen. Als Ausdrucksbewegungen modulieren und generieren sie Affekte. Es sind also die audiovisuell orchestrierten sprachlich-gestischen Ausdrucksformen, die eine kulturwissenschaftliche transdisziplinäre Perspektivierung dieser in vielfachem Sinne multimodalen medialen Ökologien notwendig machen. Dies erfordert ein radikales Umdenken der methodischen Zugriffe auf das Sprechen in audiovisuellen Medien.*

Keywords *Affekte; Ausdrucksbewegung; Methoden der Multimodalitätsforschung; Politische Kommunikation; Sprache und Gestik; Mediale Ökologien*

1 Einleitung

Tagtäglich begegnet uns Sprache eingewoben in die verschiedensten medialen Ökologien. Ob als Nachrichtensendungen, Videospiele, TikTok-Videos oder Zoom Konferenzen – in der spezifischen medialen Form oder Orchestrierung

der audiovisuellen Bilder verschmelzen sprachlich-gestische Ausdrucksformen mit denen der audiovisuellen Formate.

Seit einigen Jahren widmet sich die Medienlinguistik solchen multi-modalen Phänomenen u.a. als Verflechtungen von Wort, Bild, Körper und Medium (vgl. z.B. Opiłowski 2017; Stöckl 2012). Viele Analysen verbleiben jedoch ausschließlich im methodischen Werkzeugkasten der Linguistik, z.B. Erzählen und Argumentieren in Social Media Kommentaren (Michel/Pfurtscheller 2021) oder Sprecherwechsel bei Zoom (Xia 2023). Auf diese Weise können zwar u.a. strukturelle Fragen der Gesprächsorganisation durchaus erfolgreich analysiert werden, weitestgehend unberücksichtigt bleibt jedoch, wie audiovisuelle Medien grundlegend die Wahrnehmung der Rezipient*innen strukturieren, wie sie Affekte generieren, modulieren und mobilisieren. Anders formuliert, die medienlinguistischen Analysen bedienen sich primär eines linguistischen Analyseinstrumentariums. Wie wichtig es ist, audiovisuelle Bilder auch als spezifischen Ausdrucksmodus ernst zu nehmen, möchten wir mit diesem Beitrag unterstreichen.

Filmwissenschaftliche Arbeiten zeigen, dass die Orchestrierung audiovisueller Bewegtbilder ständig neue Rahmungen kognitiver Prozesse hervorbringt, indem sie Zuschauer*innen unmittelbar affizieren (Kappelhoff 2018: 9). Die kinematografische Inszenierung erfahren Zuschauer*innen körperlich als ein geteiltes Zuschauergefühl (Bakels 2017; Kappelhoff 2004b; Kappelhoff 2008; Kappelhoff/Bakels 2011; Sobchack 1992). Bezogen auf die Multimodalität des Sprechens in audiovisuellen Medien gerät damit die ästhetische Gestaltung audiovisuellen Materials in den Blick. Wenn also Aussagen in kleine Videos verpackt, Körper in Kameraausschnitten gerahmt und abgeschnitten werden, wenn Videos nachträglich bearbeitet und geschnitten werden, wenn diese audiovisuellen Bilder dann eine bestimmte rhythmische Dynamik entfalten, wenn sie schlicht bestimmte Aspekte filmisch in das Zentrum der Aufmerksamkeit rücken, dann strukturiert diese Inszenierung sowohl die kognitive als auch die affektive Rezeption solcherart gestalteten Sprechens.

Aus diesem Blickwinkel stellen die Multimodalität des Sprechens in der Interaktion von Angesicht zu Angesicht und die audiovisuelle Multimodalität keine getrennten Kommunikationskanäle dar. Vielmehr formen sie gemeinsam eine vieldimensionale Wahrnehmungsgestalt, »an orchestrated multidimensional experiential gestalt« (Müller/Kappelhoff 2018: 45). Diese multidimensionale Erfahrungsgestalt entfaltet sich im Prozess der Wahrnehmung audiovisueller Bewegtbilder. Es ist ihre affektive Qualität, die die Zuschauer*innen unmittelbar körperlich berührt, mitnimmt, erregt oder beruhigt. In

diesem Sinne grundiert die körperliche Erfahrung der audiovisuellen Bilder sämtliche Prozesse der Bedeutungskonstitution einschließlich des dargestellten Sprechens.

Der vorliegende Beitrag stellt deshalb einen transdisziplinären (filmwissenschaftlich-linguistischen) Zugang zur Analyse der Multimodalität des Sprechens in audiovisuellen Medien vor. Der Ansatz macht die Wahrnehmungsperspektive auf solche medienspezifischen Sprachphänomene über das philosophisch-anthropologische Konstrukt der Ausdrucksbewegung theoretisch und empirisch greifbar. Damit »entgrenzt« er die linguistische Perspektive der Erforschung von Sprache in doppelter Hinsicht: vom Lautlich-Schriftlichen zum Multimodal-Mündlichen und von der Kommunikation von Angesicht zu Angesicht zur mediatisierten Mündlichkeit. Er präsentiert Ergebnisse des binationalen Forschungsprojektes »Multimodal Stancetaking: Expressive Movement and Affective Stance. Political Debates in the German Bundestag and the Polish Sejm«. Gegenstand der hier vorgestellten Studie ist eine Rede, die im Deutschen Bundestag im Rahmen der Haushaltsdebatte 2019 gehalten wurde.

Ansatzpunkt der medienästhetischen Analyse ist die systematische Berücksichtigung der Verschränkung körperlich-sprachlicher und audiovisueller Ausdrucksformen, die das Material einfordert. Das medial dokumentierte Sprechen entfaltet sich in audiovisuell orchestrierten Gestalten, die wir als Zuschauer*innen an den Bildschirmen der Laptops, Handys oder Fernsehgeräten als Wahrnehmungsgestalten sehen, hören und fühlen. Als audiovisuelle Ausdrucksbewegungen formen sie die Empfindungen der Zuschauer*innen, generieren und modulieren deren Affekte. Es ist somit die Beschaffenheit des Materials, es sind die audiovisuell orchestrierten sprachlich-gestischen Ausdrucksformen, die eine kulturwissenschaftlich transdisziplinäre Perspektivierung dieser in vielfachem Sinne »multimodalen« medialen Ökologien notwendig machen. Diese »Entgrenzung« der Perspektive auf sprachliche Phänomene erfordert ein radikales Umdenken auch der methodischen Zugriffe auf das Sprechen in audiovisuellen Medien. Darum geht es in diesem Beitrag.

Der filmwissenschaftlich-linguistische, im Kern medienästhetische Ansatz wurde im Rahmen verschiedener Forschungsprojekte kooperativ von Film- und Sprachwissenschaftler*innen (unter der Leitung von Hermann Kappelhoff und Cornelia Müller) entwickelt und vor allem am Beispiel multimodaler Metaphorik erprobt (Greifenstein et al. 2018; Horst 2018; Kappelhoff/Müller 2011; Müller/Kappelhoff 2018; Müller/Schmitt 2015). Er wurde in den vergangenen Jahren in einem deutsch-polnischen Forschungsprojekt

zum Thema »Multimodal Stancetaking: Expressive Movement and Affective Stance. Political Debates in the German Bundestag and the Polish Sejm«¹ weiterentwickelt. Im Zentrum dieser Forschungen steht die Frage nach der Analyse affektiver Positionierung der Redner*innen (*affective stancetaking*) in audiovisuell dokumentierten politischen Reden im Deutschen Bundestag und im polnischen Sejm. Auch hier dient das historische Konzept der Ausdrucksbewegung als theoretischer und methodologischer Bezugspunkt (Kappelhoff 2004b; Kappelhoff/Müller 2011; Müller/Kappelhoff 2018; Scherer et al. 2014; Schmitt et al. 2014; Horst et al. 2014).

Dass diese transdisziplinäre Perspektive sowohl als theoretische Motivation als auch als analytischer Bezugspunkt für die Analyse multimodalen Sprechens in audiovisuellen Medien fungieren kann, illustrieren wir anhand einer Fallstudie zu einer Rede im Deutschen Bundestag. Mit dieser Fallstudie dokumentieren wir zudem die wesentlichen analytischen Schritte der Datenauswahl, Aufbereitung und Auswertung, die wir im Rahmen des deutsch-polnischen Forschungsprojekts entwickelt haben. Theoretischer Rahmen und daraus abgeleitete Analysemethodik bauen auf dem oben genannten transdisziplinären filmwissenschaftlich-linguistischen Ansatz zur empirischen Analyse multimodaler Metaphoriken auf. Der Beitrag skizziert damit theoretische, methodologische und empirische Ergebnisse aus den oben genannten Forschungskontexten zu sprachlicher und audiovisueller Multimodalität. Er gliedert sich in die folgenden Abschnitte: (2) »Grundannahmen« skizziert den theoretischen Rahmen aus dem sich die im weiteren Beitrag vorgestellte medienästhetische Methode zur Analyse des Sprechens in audiovisuellen Medien ableitet; (3) »Sprachliche und audiovisuelle Multimodalität in der Analyse« stellt die medienästhetische Methode vor, wie sie im Rahmen des DFG/NCN Forschungsprojekts für die Analyse affektiver Positionierung in deutschen und polnischen Parlamentsreden weiterentwickelt wurde. Dabei illustriert Abschnitt (3) einerseits das analytische Prozedere für das gesamte Datenkorpus (3.1, Aufbereitung und Exploration der multimodalen Daten,) und illustriert andererseits am Beispiel der kurzen Parlamentsrede, wie sich das Konzept der »Ausdrucksbewegung« in konkrete, intersubjektiv überprüfbare

1 DFG/NCN gefördertes Forschungsprojekt »Multimodal Stancetaking: Expressive Movement and Affective Stance« (MU 1587/3-1) unter der Leitung von Cornelia Müller (Europa-Universität Viadrina, Frankfurt (Oder), DE) und Maciej Karpiniński (Adam-Mickiewicz-Universität, Poznań, PL) (2020–2024).

analytische Schritte überführen lässt (3.2, Die Analyse von Ausdrucksbewegungen als Ansatzpunkt für das Zusammenspiel von sprachlicher und audiovisueller Multimodalität). Der Beitrag schließt mit weiterführenden (4) Überlegungen zum Potential eines kulturwissenschaftlichen (filmwissenschaftlich-linguistischen) Zugriffs auf das Sprechen in audiovisuellen Medien in seinen je spezifischen medialen Ökologien.

2 Grundannahmen einer medienästhetischen Perspektive

Empirische Methoden leiten sich von theoretischen Grundannahmen zum Gegenstand, den sie adressieren, ab, bleiben dabei jedoch häufig implizit. Die im Folgenden skizzierten Grundannahmen formulieren medientheoretische Positionen, die für die Untersuchung sprachlicher Multimodalität in medienspezifischen Kontexten besonders relevant sind.

2.1. Audiovisuelle Bilder formen ästhetische Erfahrungen der Rezipient*innen

In einer solchen Perspektive sind audiovisuelle Bilder gerade nicht Repräsentationen einer unabhängig von ihnen existierenden Wirklichkeit, sondern formen die perzeptiven Schemata ihrer Rezipient*innen und modulieren deren Empfindungen. Es sind die Inszenierungen dieser Bewegtbilder, die unseren ›Blick‹ durch die Kamera auf die Welt strukturieren und die uns ganz unmittelbar berühren und bewegen. Diese Position der Zuschauer*innen als verkörpertes Durchleben und Erfahren audiovisueller Bilder gilt es analytisch einzuholen (Kappelhoff 2018: Kap. 3.3-3.5).

An dieser Erfahrungsqualität setzt die hier skizzierte medienästhetische Perspektive an. Medientheoretischer Bezugspunkt ist Kappelhoffs Konzept der filmischen Ausdrucksbewegung, das er in seiner kulturhistorischen Auseinandersetzung mit dem Theater der Empfindsamkeit und dem Hollywood Melodram (Kappelhoff 2004b) entwickelt hat. Ausdrucksbewegung bezieht sich auf die Expressivität filmischer Kompositionen als Bewegungserfahrung – und hierin liegt eine der Schnittstellen zwischen audiovisueller und sprachlich-gestischer Kommunikation (Kappelhoff 2004a; Müller/Kappelhoff 2018: Kap. 9).

Kappelhoff entwickelt das Konzept u. a. in Anknüpfung an Plessners philosophische Anthropologie, die von einem unmittelbaren, körperlichen Verste-

hen, einer interkorporalen Affizierung durch Bewegungserfahrung ausgeht.² Was ist damit gemeint? Zunächst einmal geht es darum, den Affekt als Qualität der Bewegung zu fassen, sei es im gestischen oder im filmischen Bewegungsbild. Plessners Idee der Ausdrucksbewegung verortet den Affekt, den Ausdruck *in der Bewegung selbst*. Die Bewegung ist hier gerade nicht Symptom innerer Vorgänge, die sich auf diese Weise einen Weg nach außen suchen, sondern sie *ist* Affekt, sie *ist* Ausdruck. Auch in der Phänomenologie Merleau-Pontys wird dieser Gedanke entwickelt: »[Ü]brigens fasse ich Zorn oder Drohung nicht als hinter den Gesten verborgene psychische Fakten, ich sehe vielmehr den Zorn der Gebärde an: sie läßt nicht lediglich denken an Zorn, sie ist der Zorn« (Merleau-Ponty 1968: 218). Das Verstehen dieser Ausdrucksgestalt ist dann auch ein unmittelbares, körperliches, verkörpertes Verstehen. Die Qualität der Bewegung einer ärgerlichen oder freudigen Geste wird unmittelbar erlebt, als bewegend auf welche Weise auch immer – heiter-schwingend, dramatisch-aufbrausend oder explosiv (vgl. auch Bühler 1933). Das Bewegungsverstehen, von dem hier die Rede ist, ist also ein verkörpertes Verstehen, ein leibliches in Merleau-Pontys Sinn.

2.2. Audiovisuelle Bilder entfalten sich als multidimensionale Erfahrungsgestalten

Sprachliche und filmische Multimodalität realisiert sich in Form multidimensionaler Erfahrungsgestalten. Diese Gestalten lassen sich theoretisch und analytisch als filmisch-sprachliche Ausdrucksbewegungen fassen. Sie entfalten sich als zeitlich strukturierte Bewegungen audiovisueller Bilder und sprachlich-gestischer Bewegung. Das hier zugrundeliegende Konzept der Ausdrucksbewegung bezeichnet einen historisch-spezifischen affektiven Modus, eine »Matrix der Gefühle«, die ihren Ausgang im französischen Theater des 18. Jahrhunderts nahm und prägend für die Gestaltung audiovisueller Gefühle im europäisch-amerikanischen Kino wurde (Kappelhoff 2004b; 2016)³.

-
- 2 Wie die Ausdrucksbewegung als methodischer Ansatzpunkt für die Analyse audiovisueller und sprachlicher Multimodalität fungieren kann, zeigen wir in diesem Band in einem weiteren Beitrag (Müller). Vgl. hierzu auch Müller/Kappelhoff (2018) sowie Greifenstein et al. (2018).
 - 3 Vgl. hierzu auch die Arbeiten in den von Kappelhoff und Wedel herausgegebenen Reihen »Poetologien audiovisueller Bilder« und »Cinopoetics« bei De Gruyter sowie Müller/Kappelhoff (2018).

Wie eine ganze Reihe empirischer Studien aus den vergangenen Jahren belegen, strukturieren audiovisuelle Ausdrucksbewegungen jedoch auch andere mediale Formate wie etwa TV-Nachrichten, TV-Reportagen und Werbung für Baumärkte ebenso wie Wahlwerbung politischer Parteien, Musikvideos oder Social Media Kampagnen (Horst 2018; Müller/Kappelhoff 2018; Schmitt 2020; Scherer 2024). Audiovisuelle Ausdrucksbewegungen modulieren als Bewegungsbilder die kognitiven und affektiven Prozesse der Zuschauer*innen. Sie artikulieren ein ›Denken der Bilder‹, das das Denken der Zuschauer*innen rahmt, prägt und ihm eine affektive Tönung gibt. Affektivität wird dabei als Bewegungsqualität der sich entfaltenden Ausdrucksbewegungen körperlich erfahren und in Form von Affektdynamiken analytisch greifbar (Kappelhoff 2018: Kap. 1,3.2; Müller/Kappelhoff 2018: Kap. 8, 9; Scherer et al. 2014).

2.3 Mediale Ökologien strukturieren die Wahrnehmung von Sprache in audiovisuellen Medien

Multimodaler Sprachgebrauch in audiovisuellen Medien lässt sich nicht unabhängig von seiner medialen Ökologie erfassen. Überall, wo Sprache uns in medialen Kontexten begegnet – ob in Politreportagen, TikTok-Videos oder Videospielen – haben wir es mit komplex orchestrierten multimodalen Ausdrucksformen zu tun. Der vorliegende Beitrag illustriert, wie sich diese Verschränkung der Multimodalität des Sprechens mit der audiovisuellen Multimodalität mit Bezug auf das Konzept der Ausdrucksbewegung systematisch untersuchen lässt, und wie sich auf diese Weise spezifische Formen multimodaler Affizierung rekonstruieren lassen. Er nimmt damit die spezifischen medialen Ökologien multimodaler Sprachdaten in den Blick, welche erstens durch ihre spezifischen Ausdrucksformen geprägt sind und zweitens nicht ohne die Interaktion durch und mit Nutzer*innen zu denken sind. In der Interaktion zeigen sich die medialen Ökologien als »kulturelle Praxis, in der Menschen ihre subjektive Wahrnehmungswirklichkeit als Teil einer gemeinsamen geteilten Welt zu fassen suchen« (Kappelhoff 2018: 10). Durch die wechselseitige Bedingtheit (Reflexivität) von Ausdrucksformen und sozialer Praxis kann die Analyse der Modulation von spezifischen Ausdrucksformen als Ansatzpunkt für die Rekonstruktion von Prozessen der Zuschaueraffizierung nutzbar gemacht werden. Unterstützt wird dieser Aspekt der Reflexivität durch medientheoretische Arbeiten zu medialen Ökologien, die gerade auch die »Wechselwirkungen zwischen den sozialen,

politischen und psychischen Dimensionen von Medien und ihren materiellen Grundlagen« (Löffler/Sprenger 2016: 12) in den Blick nehmen.

Auf der einen Seite ist da die Multimodalität des Sprechens in der Interaktion von Angesicht zu Angesicht, die sich als dynamisches Zusammenspiel von Blick, Hand- und Körpergesten mit der gesprochenen Äußerung darstellt. Auf der anderen Seite ist da die audiovisuelle Multimodalität, d.h. die audiovisuelle Inszenierung des Sprachgebrauchs, die sich in der Orchestrierung von Kameraführung, Cadrage, Einstellungen, Montage und Ton realisiert (vgl. Müller, i.d. Bd.). Aus einer Wahrnehmungsperspektive formen beide Ebenen eine untrennbare Einheit, die sich zeitlich als Ausdrucksbewegung im Moment der Wahrnehmung, im Prozess des Zuschauens und Zuhörens entfaltet (Müller 2019; Kappelhoff/Müller 2011). Diese multidimensionale Wahrnehmungsgestalt ist der Ansatzpunkt für unsere Analyse. Sie erlaubt es uns zu rekonstruieren, wie verschiedene mediale Ökologien die Wahrnehmung von Sprache in audiovisuellen Medien strukturieren. Auf diesen Aspekt weisen auch Löffler und Sprenger hin: »Untersucht werden soll, wie sich die Strukturen der Wahrnehmung, des Denkens und des Verhaltens mit der Einführung neuer Medien auf diesen verschiedenen Ebenen verändern« (Löffler/Sprenger 2016: 12).

Wie sich diese Grundannahmen als Methodologie für die Analyse einer Multimodalität des Sprechens in audiovisuellen Medien konkret umsetzen lässt, illustrieren wir im nächsten Abschnitt.⁴

3 Sprachliche und audiovisuelle Multimodalität in der Analyse: Systematik der Analyse und Fallstudie

Der hier vorgestellte transdisziplinäre Ansatz verbindet eine filmwissenschaftliche Analyse audiovisuell generierter Affektdynamiken (>audiovisuelle Multimodalität<: empirische Medienästhetik, eMAEX⁵, AdA Filmontology;

4 Für eine detaillierte theoretische Einbettung verweisen wir auf Kappelhoff (2018), den Beitrag von Cornelia Müller (i. d. Bd.) sowie auf Müller und Kappelhoff (2018) und die oben genannten Publikationen, die im Umfeld der Arbeiten zur Kinematik audiovisueller und körperlich-sprachlicher Metaphorik (»Cinematic Metaphor«) entstanden sind.

5 »Electronically mediated analysis of expressive movement«: »eMAEX – Eine systematisierte Methode zur Untersuchung filmischer Ausdrucksqualitäten«, Online unter: <https://www.empirische-medienaesthetik.fu-berlin.de/emaex-system/index.html> (30.01.2024).

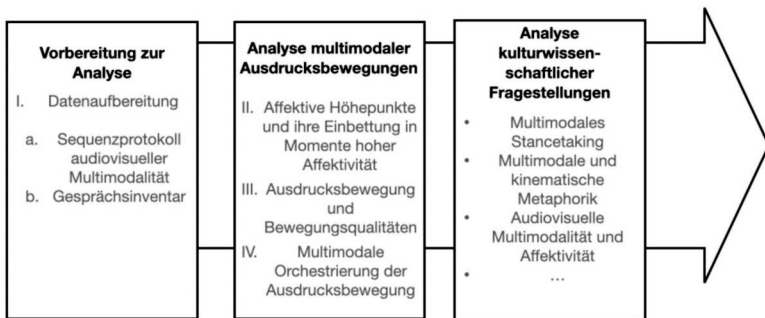
vgl. Müller/Kappelhoff 2018; Kappelhoff/Müller 2011; Kappelhoff/Bakels 2011; Bakels et al. 2020a) mit der linguistischen Analyse multimodaler Ausdrucksge-
 stalten (»Multimodalität des Sprechens«: Diskursdynamik, Gestenforschung,
 Interaktionsforschung; vgl. Müller 2010; 2024a/b; Müller/Tag 2010; Bressemer
 et al. 2013). Die im Folgenden dargestellte Systematik der Aufbereitung und Ana-
 lyse des Sprechens in audiovisuellen Medien basiert auf diesen methodischen
 Grundlagen und wurde im Rahmen des DFG/NCN geförderten binationalen
 Forschungsprojekts »Multimodal stancetaking: Expressive movement and
 affective stance« weiter entwickelt. Sie umfasst drei zentrale, aufeinander
 aufbauende Analysephasen, die in Abbildung 1 in ihrem Verlauf dargestellt
 sind: (1) die Aufbereitung der Daten für die Analyse, (2) die detaillierte Ana-
 lyse der audiovisuellen Ausdrucksbewegung und (3) die Analyse spezifischer
 Fragestellungen bezogen auf die Multimodalität des Sprachgebrauchs in
 audiovisuellen Medien. Im Rahmen des Forschungsprojekts zu parlamenta-
 rischen Reden ist dies die Frage danach, wie die aufgezeichneten Bilder des
 multimodalen Sprechens im parlamentarischen Kontext die Zuschauer*innen
 affizieren und wie sie die jeweiligen Positionierungen der Sprecher*innen im
 Sinne eines affektiven *Stance* strukturieren (vgl. Goodwin et al. 2012). In der
 hier präsentierten Fallstudie fokussieren wir darauf, wie im Zusammenspiel
 von sprachlicher und audiovisueller Multimodalität die Affekte der Zuschau-
 er*innen generiert, moduliert und mobilisiert werden. Wir argumentieren,
 dass die Wahrnehmung der Multimodalität des Sprechens in audiovisuellen
 Medien grundsätzlich affektiv durchformt ist und dass dies in der Analyse
 sprachlich-gestischer Formen von Bedeutungskonstitution Berücksichtigung
 finden muss. Wie sich dies in der konkreten Analyse einholen lässt, zeigen wir
 im Folgenden.

Wir konzentrieren uns auf die ersten beiden Schritte der Analyse: (3.1)
 Datenaufbereitung und (3.2) Analyse multimodaler Ausdrucksbewegungen
 als methodischer Ansatzpunkt für die Analyse multimodaler Sprachdaten
 in audiovisuellen Medien. Die hierfür ausgewählte Rede ist Teil des Daten-
 korpus des bereits erwähnten DFG/NCN Forschungsprojektes. Das Korpus
 besteht aus den Reden der Finanzdebatte 2019 des polnischen Sejm und
 des Deutschen Bundestages.⁶ Die als Videomaterial dokumentierten Reden

6 Im Rahmen des Projekts stehen Reden im Fokus, welche als offizielle Daten in den
 Archiven, wie etwa dem Parlamentsarchiv, gesammelt und für die Gesellschaft frei
 verfügbar sind. Solche Archive fungieren als »Gedächtnis« des Parlaments, aber sie la-
 den auch die Menschen draußen ein, an den Debatten ihrer gewählten Vertreter*in-

umfassen insgesamt 681 Minuten und stellen die offiziellen Aufnahmen des Deutschen Bundestages und des polnischen Sejm dar, die in den Parlamentsarchiven für die Öffentlichkeit aufbewahrt werden. Die Aufnahmen sind frei zugänglich und bieten den Bürger*innen die Möglichkeit, Anteil an den parlamentarischen Debatten zu nehmen. Das deutsche Datenmaterial (323 Minuten Videomaterial) umfasst insgesamt 38 Redebeiträge von unterschiedlicher Länge (02:33-35:08 Min). Im Vergleich dazu beinhaltet das polnische Material (358 Minuten Videomaterial) insgesamt 105 deutlich kürzere Redebeiträge (01:00-26:00 Min). Bei der folgenden exemplarisch analysierten Rede handelt es sich um den Beitrag eines Mitgliedes der SPD-Fraktion (Dr. Jens Zimmermann). Sie hat eine Gesamtlänge von 5:32 Min.⁷

Abb. 1: Systematik der Datenaufbereitung und Analyse im Überblick



3.1 Aufbereitung und Exploration der multimodalen Daten

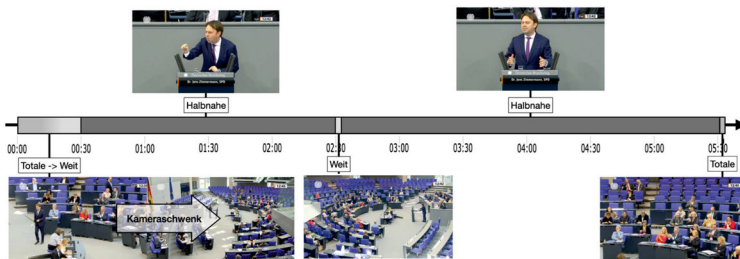
Der erste Schritt der Analyse besteht in der Aufbereitung der Daten. Dazu werden parallel ein Sequenzprotokoll der Rede, ein schriftliches Gesprächsinventar (vgl. Deppermann 2008: 31–35) in einer Excel Tabelle und eine Annotation in der Transkriptionssoftware Elan erstellt.

nen teilzunehmen. Das Projekt untersucht die audiovisuellen Materialien, die von den jeweiligen Archiven der Haushalts- und Finanzdebatten in den deutschen und polnischen Parlamenten zur Verfügung gestellt werden.

7 Die Rede ist vom 11.09.2019 und ist über die offizielle Website des Bundestages einsehbar: <https://www.bundestag.de/mediathek?videoid=7388015#url=L21lZGlhGhla292ZXJsYXk/dmlkZW9pZD03Mzg4MDE1&mod=mediathek> (08.09.2023).

Das Sequenzprotokoll ermöglicht einen ersten Zugang zur Ebene der audiovisuellen Orchestrierung und wird im Sinne eines vereinfachten *Einstellungsprotokolls* (vgl. Hickethier 2012: 37f.) erstellt. Dazu liegt der Fokus auf den Wechseln der Kameraeinstellungen und -perspektiven über den Verlauf der gewählten Sequenz, wobei auch Aspekte der Bildraumkomposition zutage treten. Das filmische Vokabular orientiert sich ebenfalls an den standardisierten Bezeichnungen wie z. B. in der AdA Filmontologie (Bakels et al. 2020a) aufgeführt. Abbildung 2 zeigt das Einstellungsprotokoll der hier betrachteten Rede als Zeitstrahl.

Abb. 2: Darstellung der audiovisuellen Multimodalität: Sequenzprotokoll der Rede von Jens Zimmermann (Kameraeinstellungen und -perspektiven der Rede im zeitlichen Verlauf)



Die Rede weist eine, wenn auch zurückhaltende, audiovisuelle Inszenierung auf, die für die Aufnahmen des Korpus typisch ist. Größtenteils steht dabei Zimmermann im Zentrum des Bildraumes und damit auch im Zentrum der Aufmerksamkeit der Zuschauer*innen an den Bildschirmen. Die beiden Stills oberhalb des Zeitstrahls zeigen diese Perspektive aus sogenannter Normalsicht in halbnaher Einstellung. Nur einmal unterbricht ein Schnitt die Perspektive und die Kamera wechselt in eine weite Einstellung und zeigt das Parlament aus einer Aufsicht. Zimmermanns Stimme klingt währenddessen aus dem Off weiter. Die Rede ist von zwei weiteren Einstellungswechseln gerahmt, die mit dem Auf- und Abtreten Zimmermanns zum Pult einhergehen. Zu Beginn wird Zimmermann in seinem Gang nach vorne von der Kamera begleitet, um darauf in einem Schwenk und Zoom-Out das gesamte Parlament von links oben sichtbar werden zu lassen. Die letzte totale Einstellung ist erneut statisch. Vergleichbare Rahmungen zeigen sich im gesamten Datenkorpus der Bundestagsreden.

Das Gesprächsinventar ermöglicht einen ersten deskriptiven und kommentierenden Zugriff auf die gesamte Rede. Hier werden die Metadaten der Rede erfasst: Datum, Redner*in, und es erfolgt eine erste deskriptive Charakterisierung des/der Redner*in. In einer Sichtung wird die Rede weiter in interaktive Einheiten (*interactive units*) unterteilt. Die Einheiten sind entweder durch Beifall oder durch hörbare und/oder sichtbare Zwischenrufe begrenzt, die wiederum Reaktionen auf Seiten der sprechenden Person hervorrufen können. Die *Timecodes* dieser Einheiten dienen später der besseren Orientierung im Transkript und der Elan-Datei.

Im Gesprächsinventar wird die Rede zunächst als orthographisches Transkript festgehalten. Für Reden aus dem Bundestag kann dazu das online verfügbare Plenarprotokoll der Sitzung herangezogen werden. Die Transkription umfasst neben der Rede auch interaktive Aspekte wie Zwischenrufe und Applaus. In der Kommentarspalte werden parallel erste Beobachtungen von Auffälligkeiten der sprachlichen und audiovisuellen Modalitäten notiert. Auf diese Weise erschließt sich das vorliegende Material den Analysierenden in seiner Spezifik.⁸ Dies ist wichtig, denn damit kommt gewissermaßen das Material, in seiner audiovisuellen Orchestrierung multimodalen Sprechens, selbst zu Wort. Hypothesen werden so aus dem Material heraus entwickelt und nicht in Form präexistenter Fragen auf das Material projiziert.

Abb. 3: Gesprächsinventar ›Zimmermann-Rede‹

Time of Interactive Units	Speaker	Orthographic Transcription	General Comments
00:00:00-00:00:04	WK, JK	der SPD-Fraktion. (Johannes Kahrs [SPD]: Bester Mann! – Beifall bei der SPD)	from WK's close-up to tracking shot of JZ from zooming out high right angle shot
00:00:05-00:00:09	(WK)	- (Vizepräsident Wolfgang Kubicki: Das sind Sie doch! – Heiterkeit bei Abgeordneten der SPD)	high right angle shot further zooms out, podium adjustment sound
00:00:09-00:00:59	JZ	Sehr geehrter Herr Präsident. Meine Damen und Herren. Der Bundeshaushalt trägt die Überschrift „Investitionen und Chancen im Wandel“. Und wenn ich mir anschau, was in der Digitalpolitik in diesem Herbst alles passiert: Alleine in dieser Woche ist der Bericht der Wettbewerbskommission 4.0 vorgelegt worden. Das Digitalkabinett tagt. Wir werden den Digitalgipfel haben. Wir haben eine Konferenz der Vereinten Nationen, das Internet Governance Forum, hier in Berlin. Die Blockchain-Strategie der Bundesregierung wird vorgestellt und auch die Datenethikkommission präsentiert ihre Ergebnisse.	cam shot is still zooming out until furthest point, showing JZ stepping back and forth behind the podium, until cam changes to close-up of JZ; he engages in some sudden accentuated discursive beat gestures, employing also his head and upper body and occasional

Abbildung 3 zeigt die ersten drei Zeilen des Gesprächsinventars der Zimmermann-Rede. Die Kommentarspalte dokumentiert die oben beschriebene Kamerafahrt (vgl. Abb. 2) in zeitlicher Relation mit der sicht- und hörbaren Handlung. Hier rahmt sie die Zeitspanne, vom Aufruf des Redners bis zum Beginn seiner Rede am Pult. Sie zeigt den von Applaus und Zwischenrufen beglei-

8 Zur Reliabilität der Analyseschritte vgl. Abschnitt 3.2.1.

teten Weg des Redners zum Pult und den Beginn seiner Rede. Die Bewegung der Kamera und der Zoom auf den Sprecher am Ende der Kamerafahrt setzen den Auftakt der Rede in Szene. Im Gesprächsprotokoll werden zudem erste Auffälligkeiten der sprachlichen Modalitäten notiert. Die Abbildung verweist beispielsweise auf rhythmische diskursstrukturierende Gesten, die nicht nur mit den Händen ausgeführt werden, sondern »employing also his head and upper body [...]« (Abb. 3: General Comments). Diese Dokumentation erlaubt einen ersten materialbasierten, deskriptiven Zugriff auf die später ausführlich analysierte Orchestrierung potenzieller multimodaler Ausdrucksbewegungen.⁹

3.2 Die Analyse von Ausdrucksbewegungen als Ansatzpunkt für das Zusammenspiel sprachlicher und audiovisueller Multimodalität

Wir haben eingangs erwähnt, dass unsere kulturwissenschaftliche Perspektive auf die Multimodalität des Sprechens in audiovisuellen Daten die Position der Zuschauer*in einnimmt und dass audiovisuelle Bilder und die dort gezeigten Formen multimodalen Sprechens sich im Akt des Zuschauens als multidimensionale Erfahrungsgestalten entfalten. Diese Gestalten sind unsere methodischen Ansatzpunkte. Wir fassen sie als audiovisuelle Bewegungsbilder, deren Bewegungsentfaltung als audiovisuelle Ausdrucksbewegungen in ihren affektiven Qualitäten beschrieben werden (vgl. hierzu ausführlich Müller/Kappelhoff 2018: Kap. 9, sowie der Überblick bei Scherer et al. 2014).

Das heißt, in diesem Schritt der Analyse nehmen wir die Ausdrucksbewegungen als filmisch orchestrierte Bewegungsfigurationen in den Blick (eMAEX; siehe Müller/Kappelhoff 2018, Appendix). Wir gehen dabei in drei Schritten vor: Zunächst fragen wir danach, wie man Ausdrucksbewegungen als audiovisuelle und sprachlich-körperliche Gestalten »identifiziert« (3.2.1). Im nächsten Schritt beschreiben wir die affektive Qualität der audiovisuell orchestrierten Ausdrucksbewegung (3.2.2) und analysieren sodann im Detail, wie gestische, sprachliche und gesprächsorganisatorische Aspekte in der multimodalen Orchestrierung der Ausdrucksbewegung zusammen wirken (3.2.3).

9 Die detaillierte Annotation in Elan kann an dieser Stelle aus Platzgründen nicht ausgeführt werden. Sie wird jedoch im Rahmen der multimodalen Analyse immer wieder aufgegriffen.

3.2.1 Affektive Höhepunkte und ihre Einbettung in Momente hoher Affektdynamik

Wir nähern uns den audiovisuellen Daten in einer phänomenologischen Grundhaltung, d.h. die analytische Perspektive ist die einer betrachtenden – oder besser – zuschauenden Person am Fernsehgerät oder Laptop. Die Bewegungsbilder – bestehend aus den audiovisuellen Bildern, die Körperbewegungen zeigen und sich als zusammengehöriges, dynamisches Ganzes entfalten – weisen Bewegungsqualitäten auf, die von den Zuschauer*innen leiblich erfahren werden. Die Rede wird im Ganzen in Elan angesehen und währenddessen Bereiche markiert, die sich durch besonders hohe Affektivität vom Rest der Rede unterscheiden. Diese werden als affektive Höhepunkte in der Annotation festgehalten. Dabei folgt die »Identifikation« der Höhepunkte nicht dem Prinzip des Addierens einzelner Ausdrucksmodalitäten, sondern vielmehr dem der zeitlichen Entfaltung multidimensionaler Ausdrucksgestalten. Diese Höhepunkte treten als deutlich wahrnehmbare Veränderungen der Intensität von der sogenannten *baseline*, d.h. dem Grundrhythmus der Rede, als z.B. explosive, stetig steigende oder auch plötzlich abfallende Qualitäten hervor. *Baseline* bezeichnet dabei den spezifisch, individuell-charakteristischen Sprachduktus – oder die multimodale Dynamik des Sprechens, von der sich der Höhepunkt absetzt. Affektive Höhepunkte und *baseline* sind nur als aufeinander bezogene Qualitäten des multimodal orchestrierten Sprechens zu fassen. Ausgehend von diesen affektiven Höhepunkten wird die audiovisuelle Bewegungsgestalt beschrieben, in die diese eingebettet sind. Diese Gestalten werden in Elan festgehalten und als Momente hoher Affektdynamik bezeichnet. Sowohl die Identifikation der affektiven Höhepunkte als auch die Beschreibung der Affektdynamiken wurde intersubjektiv und interkulturell in Tandems abgesichert (*two independent coder procedure*)¹⁰. Die aus diesen Tandems hervorgehenden Dynamiken ähneln in ihrer Grundidee der Kendon'schen Segmentierung der zeitlichen Entfaltung gestischer Bewegungen in *preparation*, *stroke* und *recovery* (2004: 111f.): Die einzelnen Phasen der Bewegungsgestalten können, genau wie im Falle von Handgesten, spezifische

10 Selbst in bilingualen Tandems konnten im Rahmen des *Mumo Stance* Projektes deutliche Überschneidungen in den Markierungen von Höhepunkten und den Momenten hoher affektiver Intensität erreicht werden. Dies spricht für den Ansatz des Zuschauer*innengefühls und gegen die Perspektive individueller Bewertung. Siehe hierzu auch die Anmerkungen bezüglich der Annotation audiovisueller Daten und »*pattern-seeking*« (Bakels et al. 2020b: 115; Herv. i. Orig.) im Rahmen der AdA Filmontology.

Bewegungskonfigurationen und Dynamiken entfalten. Die *preparation* etwa kann sich als langsamer und kontinuierlicher Aufbau oder aber als kurze, abrupte starke Rhythmisierung hin zum Höhepunkt darstellen. Das gleiche gilt für den affektiven Höhepunkt und das Abklingen in der Phase der *retraction*. Im Falle der politischen Reden beobachten wir oft ein Einhergehen von *retraction* und Ende einer interaktiven Einheit, in der der Moment hoher affektiver Intensität im Applaus verklingt, so auch bei Zimmermann. Die Dynamiken werden abschließend mit einem Vektorzeichenprogramm (hier Inkscape) in ihren Qualitäten, ihrem zeitlichen Verlauf und ihrer Position innerhalb der gesamten Rede visualisiert.

Abb. 4: Ein affektiver Höhepunkt (rot) eingebettet in einen Moment hoher Affektdynamik (orange) und weitere abgeschwächte affektive Höhepunkte (grau) in der Zimmermann Rede

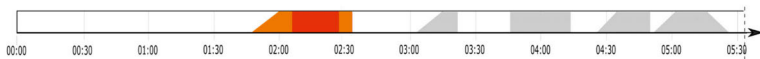


Abbildung 4 verdeutlicht, dass in Zimmermanns Rede nur ein außerordentlich affektiver Höhepunkt identifiziert wurde, welcher in einen Moment hoher affektiver Intensität eingebettet ist. Dieser befindet sich etwa in der Mitte der Rede und hat eine Länge von 46 Sekunden. Der Höhepunkt (markiert in rot) beginnt mit einer stetig ansteigenden, affektiven Intensität in der Vorbereitungsphase und endet mit einem abrupten Abfall (markiert in orange). In grau sind vier Bereiche in der Rede abgebildet, die ebenfalls eine höhere affektive Intensität aufweisen und ihrerseits durch unterschiedliche Dynamiken gekennzeichnet sind. So wechseln sich langsame und stetige Anstiege mit plötzlichen Anstiegen und abrupten Abbrüchen ab. Die Abbildung dokumentiert, dass in der etwa fünfminütigen Rede von Zimmermann die affektive Dynamik etwa nach zwei Minuten Redezeit deutlich zunimmt und dann bis zum Ende der Redezeit in abgeschwächter, fast rhythmischer Form immer wieder präsent ist. Auf diese Weise lassen sich spezifische affektive Profile für die Reden der verschiedenen Sprecher*innen sichtbar machen. Hier fokussieren wir uns auf den Moment der Zimmermann Rede, der die höchste affektive Intensität aufweist.

3.2.2 Ausdrucksbewegungen und ihre Bewegungsqualitäten

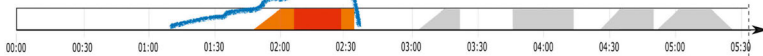
Ausgehend von den Momenten hoher Affektdynamik und den darin eingebetteten Höhepunkten lassen sich nun die audiovisuellen Ausdrucksbewegungen bestimmen (Müller/Kappelhoff 2018: Kap. 9; Scherer et al. 2014).¹¹ Auch bei der Ausdrucksbewegung haben wir es mit einer körperlich gefühlten und erlebten Einheit zu tun, die im Moment des Betrachtens der Rede entsteht. Sie entwickelt sich aus dem Zusammenspiel sprachlicher und audiovisueller Multimodalität zu einer Bewegungsgestalt. Diese Bewegungsgestalt ist ein sich zeitlich entfaltendes Bewegungsbild. Die Ausdrucksbewegung ist eine integrierte Ganzheit, deren affektive Qualität die Bewegungsqualität des sich entfaltenden Bewegungsbildes ist. Ihre Orchestrierung entsteht im Zusammenwirken der Multimodalität des Sprechens (semantische, pragmatische, syntaktische, gestische, interaktionale, prosodische Aspekte) auf der einen und der audiovisuellen Multimodalität (Einstellungsperspektiven, Montage, Bildraumgestaltung, Kamerabewegungen) auf der anderen Seite. Affektivität entfaltet sich hier als Gestalt, als integriertes Ganzes. Als Ausdrucksbewegung bildet sie einen zusammengehörigen Intensitätsbogen mit einem dynamischen Rhythmus. Ihre Grenzen ergeben sich aus dieser Bewegungsgestalt. Die Analysen in Tandems haben nicht nur innerhalb der Analyseteams, sondern auch im deutsch-polnischen Vergleich eine hohe intersubjektive Übereinstimmung ergeben, was sowohl die Grenzen als auch die Qualitäten der Ausdrucksbewegungen als ganzheitliche multidimensionale Gestalten angeht.

Unsere vergleichenden Analysen im Projekt haben gezeigt, dass die Grenzen der Ausdrucksbewegung mit den im vorherigen Analyseschritt beobachteten Grenzen der Momente hoher affektiver Intensität zusammenfallen können, aber nicht müssen. Tatsächlich beginnen die spezifischen Dynamiken der Ausdrucksbewegungen häufig deutlich, *bevor* die affektive Qualität in besonderem Maße ansteigt und als Moment besonderer affektiver Intensität erkennbar wird. So verhält es sich auch mit der Ausdrucksbewegung im Beispiel von Jens Zimmermanns Rede. Abbildung 5 zeigt, wann und wie sich

11 Ausdrucksbewegungen sind selbstverständlich nicht nur auf Momente hoher Affektdynamik beschränkt. Die Beschränkung ist dem analytischen Fokus des Projekts auf Momente hoher Affektdynamik geschuldet. Die Analyse der Qualität der Ausdrucksbewegung ermöglicht es dann, verschiedene Formen hoher Affektdynamik zu unterscheiden.

die Qualität des Ausdrucks um den Moment hoher affektiver Dynamik herum entfaltet (blaue Kurve).

Abb. 5: Affektiver Höhepunkt (rot) und Moment hoher affektiver Intensität (orange) eingebettet in die dynamische Entfaltung der Ausdrucksbewegung (blau)



Um die affektiven Dynamiken und Qualitäten der Ausdrucksbewegungen sprachlich zu erfassen, werden diese so beschrieben, dass der gefühlte und erlebte Eindruck der Dynamik möglichst bewahrt wird.¹² Die Forschenden nehmen damit ihre eigene körperliche Erfahrung der Rede zum methodischen Ansatzpunkt, denn auch sie sind grundsätzlich Zuschauende am Bildschirm. Die oben visualisierte Ausdrucksbewegung, in die der affektive Höhepunkt der Rede Zimmermanns eingebettet ist, entfaltet ihre affektive Qualität wie folgt:

Die Ausdrucksbewegung beginnt mit einem langsamen und stetigen Anstieg der Intensität, die sich in einem schneller und lauter werdenden Sprachgebrauch äußert und von akzentuierten gestischen Bewegungen begleitet wird. Gefolgt vom Gelächter des Publikums steigt die Intensität weiter. Die hohe Intensität wird gehalten und erstreckt sich in einer Art Plateau. Das Plateau ist von stark akzentuierten rhythmischen Bewegungen der Hände und des Oberkörpers gekennzeichnet und entfaltet so eine persistierende, repetitive, affektive Qualität. Eher abrupt – mit einem Lösen der Gerichtetheit des Körpers, einer tiefen Vorbeugung und einem Absenken der Stimme – lässt die Intensität der Ausdrucksbewegung nach.

Bei der sprachlichen Fassung der Bewegungsqualitäten stützen wir uns insbesondere auf ein Vokabular, das im Kontext der Beschreibung interaffektiver Prozesse in Mutter-Säugling Interaktion von Daniel Stern (2011) entwickelt und für die Beschreibung der affektiven Dynamiken und Qualitäten in der Filmanalyse von Sarah Greifenstein (2020) weiterentwickelt wurde.

12 Die Beschreibung der Ausdrucksbewegung folgt dabei dem für die Filmanalyse entwickelten Verfahren der Analyse filmischer Ausdrucksbewegungen in Form von kurzen Vignetten, die die affektiven Qualitäten beschreibend erfasst (eMAEX).

Methodisch-deskriptiv fokussiere ich entsprechend Bewegungsverben (oder ihre Substantivierungen) für die gesamte Verlaufsform (z.B. das Klopfen) und Adjektive zur Präzision der sinnlichen Eigenschaften dieser Bewegung (das sanfte Klopfen, das eilig hastige Klopfen, das harte-energisches Klopfen). (Greifenstein 2020: 69)

Die Qualitäten der Ausdrucksbewegungen werden als Qualitäten einer ganzheitlichen Gestalt greifbar. Sie zeichnen sich durch eine verschränkte Wahrnehmung von Bewegung, Kraft, Zeit, Raum und Gerichtetheit aus (Stern 2011: 13). Sterns Vokabular zielt dabei auf die Beschreibung der Formen, die die Bewegungsdynamiken ergeben. Es enthält Begriffe, die die Bewegung sowie die Intensität der Bewegung im zeitlichen Verlauf erfassen. Die sprachliche Beschreibung der sich in der Zeit entfaltenden Gestalt der Ausdrucksbewegung versucht – genuin phänomenologisch – die im Prozess des Zuschauens erlebten affektiven Qualitäten zu greifen. Dazu beschreibt sie die dynamischen Qualitäten und Intensitäten der situativen Bewegung. Sie konzentriert sich folglich auf die dynamische Bewegung und ihre sinnlichen Eigenschaften.

3.2.3 Multimodale Orchestrierung der Ausdrucksbewegung

An die Beschreibung der Bewegungsqualitäten der Ausdrucksbewegung schließt sich die detaillierte Analyse sprachlich-gestischer Multimodalität innerhalb der Ausdrucksbewegung an. Wie genau gestaltet sich das Zusammenspiel der sprachlichen und audiovisuellen Multimodalität in der aufgezeichneten Rede? Hier kommt die Spezifik der multimodalen Orchestrierung in den Blick, genauer das dynamische Zusammenspiel der Modalitäten, die den spezifischen Rhythmus einer Ausdrucksbewegung modulieren. Kappelhoff (vgl. 2008: 206) beschreibt diesen Rhythmus als die alles umspannende Ebene: Sie ist Struktur und Organisation des ästhetischen Wahrnehmungsvorgangs der kognitiven, affektiven und perzeptiven Prozesse des Zuschauers. Im Rhythmus, so Kappelhoff weiter, wird eine Bewegung zur Ausdrucksbewegung, zu einem kinematografischen Bild, in dem Gefühle nicht repräsentiert, sondern im Empfinden realisiert werden. Die Entfaltung dieses Rhythmus kann empirisch in der Analyse der multimodalen Orchestrierung der Ausdrucksbewegungen rekonstruiert werden. Neben dem Rhythmus von Handgesten und Intonation betrifft dies etwa auch rhythmische Veränderungen der Lautstärke, der Kamerabewegungen oder der Schnitttechniken, sowie interaktive Aspekte wie ein plötzliches Aufwallen von Applaus oder Zwischenrufen. Wie bereits angeführt, sind nicht immer alle Modalitäten

gleichermaßen an der Gestaltung dieses Rhythmus beteiligt. So zeigt der Blick auf die Ebene der audiovisuellen Multimodalität in der Zimmermann Rede eine klare Zurückhaltung in der Inszenierung. Die Kamera verbleibt durchgehend in einer stabilen Position und zeigt Zimmermann in halbnaher Einstellung von Kopf bis Hüfte am Rednerpult, welches zentral im Bildraum steht. Mit dieser Einstellung geht eine besondere Perspektivierung auf den Sprecher einher.

Die sprachlich-gestische, multimodale Orchestrierung der Ausdrucksbewegung und ihre Momente hoher Affektivität verweben sich mit den Inhalten der Rede. Wie diese Verschränkung sprachlicher und audiovisueller Multimodalität in solchen Momenten konkret funktioniert, wollen wir am Beispiel der Zimmermann Rede nun etwas genauer beschreiben. Worum geht es inhaltlich überhaupt? Und was ist der thematische Bezugspunkt des affektiv hochaufgeladenen Moments? Das zentrale Thema der gut fünf Minuten langen Rede ist die »Digitalpolitik der Bundesregierung«. Zimmermann beginnt mit einer Aufzählung der sogenannten »Tagesordnungspunkte« im Bereich der Digitalisierung des aktuellen politischen Geschehens und folgert, dass sie, die Sozialdemokrat*innen, diesen digitalen Wandel mitgestalten möchten. Nach abklingendem Applaus eröffnet er einen thematisch anschließenden Aspekt: klare Regeln auf dem Arbeitsmarkt für große digitale Konzerne. An dieser Stelle setzt die oben in ihren dynamischen Qualitäten beschriebene Ausdrucksbewegung an. Hier beginnt der kontinuierliche Anstieg der affektiven Intensitäten, den wir im Abschnitt 3.2.2 in seinen affektiven Qualitäten beschrieben und in Abbildung 5 als Bewegungsgestalt visualisiert haben. Abbildung 6 zeigt die Ausdrucksbewegung nun als gesprächsanalytisches Transkript.

Zimmermann lobt die Arbeit von Arbeitsminister Hubertus Heil, der den digitalen Strukturwandel aktiv gestalte (Zeile 001–010). Im weiteren Verlauf entfaltet sich interaktiv ein Missverständnis rund um Zimmermanns Empörung über die Behauptung, im Haushalt von Hubertus Heil würde »das Geld zum Fenster rausgeschmissen« (Zeile 011–014). Dieses beginnt mit Gelächter und Applaus aus dem Publikum (015), was zu einer Unterbrechung Zimmermanns führt.

An dieser Stelle beginnt die Entfaltung des affektiven Höhepunkts als Plateau hoher affektiver Intensität (visualisiert im gepunkteten Rahmen). Hier verdeutlicht Zimmermann erneut die Relevanz der Gestaltung des digitalen Strukturwandels (016–019) und endet mit dem Hinweis auf eine mögliche Konsequenz, sollte diese Relevanz ignoriert oder verharmlost wer-

den (019–020), wobei die affektive Qualität deutlich abnimmt. Das Ende der Ausdrucksbewegung ist dann deutlich markiert durch die Pause des Redners und Applaus im Publikum und zeichnet sich (im Gegensatz zum Beginn) durch einen abrupten Abfall der affektiven Intensität aus. Im Rest der Rede widmet Zimmermann sich weiteren Aspekten des Themenbereiches »Digitalpolitik der Bundesregierung«, es kommt aber zu keinem weiteren vergleichbar explosiven affektiven Ausbruch.

Abb. 6: Transkription der Ausdrucksbewegung (blaue Umrandung) nach GAT 2 in Verbindung mit Gestenannotation in der Sprachzeile: Nummerierte Gesture Units ($\{^i\}$) und Gesture Strokes (Fettsatz)

001	JZ	und deshalb (.) ist es GU:T, dass WIR-
002		mit hubertus HEIL einen sozialdemokratischen arbeitsminister haben,
003	JZ	¹ {der auch (.)} LIEFert an dieser stelle. wir {ham, ² h die NATIONALE weiterbildungsstrategie-
004	P	(((Beifall)))
005	JZ	wir ham das qualifizierungs]CHANgegesetzt- ³ h wir ² {ham-
006		die DENKfab rik in der digitalen arbeit- und wir haben auch das arbeit von MORGEN gesetzt.
007		und ³ {es ist WICHTig-} (.) ⁴ { DIESEN digiTALen strUKTURwandel-
008		⁵ h Aktiv zu ge]STALTEN; ⁶ h ² {denn die beSCHÄFtig}ten,
009		⁷ h die haben zu RECHT an einigen stellen auch beDENKen ⁸ {wie es} weitergeht;
010		und ⁷ {deswegen ist es} !WICHTig, ⁹ h ⁴ {dass wir dort}! ¹⁰ {AKtiv HANdeln.}
011		¹¹ {und wenn ich dann HÖre meine damen und herrn; ¹² h im HAUSHalt von hubertus hEIL-
012		da wÜrde das geld zum FENSTER rausgeschmissen werden-
013		¹³ h dann ist das nicht FALSCH- (.) <<stockend> ¹⁴ h geNUG- (.) zu- (.)> problemATISIE:ren ==
014		[= ¹⁵ weil das GELD- das geld= ¹⁶ ja da können sie LACHen, da können sie LA]CHen meine] kollegen von der ef: der; pe;
015	P	(((Gelächter, Applaus)))
016	JZ	¹⁷ h ¹¹ {aber an DIEser stelle den digITalen strUKTURwandel zu gestALTen; ¹⁸ h das ist eine MAMMutaufgabe;
017		¹⁹ h und die MENschen die angst haben IHren job durch digitalisier]ung zu verlieren,
018		²⁰ h um die müSsen wir uns geNAU]so kümmern, ²¹ h wie die menschen die ang ²² {st haben durch den !KOH!ausstieg ihren} job zu verlieren;
019		²³ h und wenn wir uns SO um die kümMern; ²⁴ h so verÄCHTlich wie SIE grade] geLAcht haben;
020		²⁵ h dann sind DAS [die nächsten,] ²⁶ h die die extre!MIS]ten wählen werden meine damen und herrn.]
021	P	(((Applaus))) ((das war nicht verächtlich) das war ein] netter verSPRECher,)

Um die multimodale Orchestrierung der Ausdrucksbewegung gesprächsanalytisch zu fassen, wurde sie nach GAT 2 (Selting et al. 2009) transkribiert. Dies ermöglicht nicht nur einen Einblick in die sequenzielle Dynamik der Rede, sondern ist auch eine Voraussetzung für eine mikroanalytische Rekonstruktion der sprachlichen Multimodalität, hier insbesondere des Zusammenspiels von Rede und Geste innerhalb der Ausdrucksbewegung. Die Mikroanalyse fokussiert im vorliegenden Beispiel auf den Höhepunkt der affektiven Intensität in der Ausdrucksbewegung, also den Anstieg, das Plateau und den markant abfallenden Ausklang. Neben der Kennzeichnung der Momente hoher affektiver Intensität (Anstieg und Ausklang; Rahmen in orange, Plateau: Rahmen rot-gepunktet) wird das Transkript durch eine einfache Notation der Handgesten erweitert. Diese Notation orientiert sich am multimodalen Transkriptionssystem von Müller (1998: Appendix).

Wie sich die perzeptive Erfahrungsgestalt der Ausdrucksbewegung in ihrer multimodalen Orchestrierung genauer beschreiben lässt, illustrieren die folgenden Unterkapitel am Beispiel des Zusammenspiels interaktiver und sprachlich-gestischer Dynamiken (3.2.3.1) sowie einer Analyse der gestischen Komplexität (3.2.3.2). Beides sind Aspekte, die die Spezifik der Ausdrucksbewegung mit ihrem charakteristischen Bewegungs- und Intensitätsprofil in der Zimmermann Rede in besonderer Weise auszeichnen.

3.2.3.1 Ausdrucksbewegung als sprachlich-gestisches und interaktives Geschehen

Wenn wir uns die identifizierten und auf ihre expressiven Qualitäten hin beschriebene Gestalt der Ausdrucksbewegung nun mit Blick auf das sprachlich-gestische und interaktive Geschehen betrachten, dann zeigt sich, dass ganz dem Diskurstyp »Rede« entsprechend, Zimmermann der primäre Sprecher ist. Dies wird nicht nur auf gesprächsstruktureller Ebene, sondern auch in der filmischen Inszenierung deutlich, die Zimmermann durchgehend zentral im Bildraum aus einer frontalen, halbnahen Perspektive zeigt. Aus der Perspektive der Zuschauer*innen ähnelt dies einer direkten Anrede. Dennoch ist die Rede kein Monolog, sondern kann dialogisch verstanden werden, da sie interaktiv gemeinsam mit dem Publikum geformt wird. Dies zeigt sich deutlich an den drei Stellen, an denen das Publikum sich mit Redebeiträgen in die Rede »einklinkt«. Während der Beifall in Zeile 004 keinen Einfluss auf Zimmermanns Redefluss nimmt, führen Gelächter und Applaus in Zeile 015 zu einer Unterbrechung, die auch den Umschwung vom Anstieg zum Plateau des affektiven Höhepunkts innerhalb der Ausdrucksbewegung markiert. Dieser Moment ist somit Resultat eines *interactive boosts*, d.h. einer hohen interaktiven Dynamik, rund um einen »Versprecher« Zimmermanns.

Was passiert hier nun genau? Den Anstieg affektiver Intensität markiert eine doppeldeutige Formulierungen und ein stockender Redefluss. In Zeile 013 stolpert Zimmermann zunächst über eine Formulierung und sein Redefluss gerät ins Stocken. Dadurch kommt es zu einer Bedeutungsverschiebung. Die eigentlich intendierte Problematisierung der Behauptung des Geldverschwendens (Zeile 012) kehrt sich durch die intonatorische Markierung von »FALSCH-(.)« (013) als Phrasenende in eine doppelte Verneinung und damit in die genteilige Bedeutung: Die Behauptung ist wahr. Zimmermann repariert das semantische Problem durch eine syntaktische Weiterführung der Äußerung (013), die sich jedoch durch wiederkehrendes Stocken ungewöhnlich dehnt. Trotz dieser eigentlich erfolgreichen syntaktischen und semantischen Repara-

tur bricht laut hörbares Gelächter und Applaus aus (015). Zimmermann stockt erneut in seiner Äußerung, wiederholt diese und bricht sie dann schließlich ab: »=weil das GELD- das geld-« (014). Dies markiert den Beginn des affektiven Höhepunkts, das Plateau erstreckt sich dann insgesamt über die Zeilen 014–19. Zimmermann wendet sich körperlich und pragmatisch direkt den »Störern« zu, indem er sich von seiner bisher mittigen Ausrichtung mit seinem gesamten Körper nach rechts orientiert und Richtung Publikum spricht »da können sie LACHen, da können sie LACHen meine kollegen von der ef de: pe:« (014). Deutlich lauter und schneller spricht er nun mit wiederkehrenden Überbetonungen direkt in Richtung der FDP (die jedoch nicht im Bild sichtbar wird). Das Plateau hoher affektiver Intensität geht einher mit seiner Anklage an die FDP, dem Vorwurf, sie würde die Dringlichkeit der Lage verkennen und die Bedürfnisse der Menschen lächerlich machen. Die politische Konsequenz aus diesem Fehlverhalten bildet den Abschluss der interaktiven Sequenz und der Ausdrucksbewegung: »°h dann sind DAS die nächsten, °h die die extre!MIS!ten wählen werden meine damen und herrn« (Zeile 020).

Zimmermann markiert das Ende seines Redebeitrages mit einer Wendung des Körpers zurück zur Mitte, senkt den Blick auf seine Notizen und schaut dann erneut in das gesamte Publikum, bzw. zu uns als Zuschauer*innen an den Bildschirmen. Währenddessen ertönt erneuter Applaus, ebenso ein Zwischenruf, vermutlich aus Reihen der FDP, der versucht, das Missverständnis noch aufzulösen (>das war nicht verächtlich das war ein netter verSPRECHer-«) und die Ausdrucksbewegung findet ein eher abruptes Ende.

Die Ausdrucksbewegung mit der langsam ansteigenden und dann gleichbleibend hohen affektiven Intensität und einem plötzlichen Abfall der Affektdynamik am Ende wird somit interaktiv getriggert. Die Unterbrechung durch das Publikum wirkt gleichsam wie ein interaktiver *boost*, ein interaktives Anheizen der affektiven Dynamik, und leitet den abrupten Wechsel nicht nur des Sprechaktes, sondern auch der Dynamik ein.

Die multimodale Orchestrierung der Ausdrucksbewegung bezogen auf die Multimodalität des Sprechens zeichnet sich dann durch eine deutliche Intensivierung der Lautstärke, der Sprechgeschwindigkeit und auch des Sprachrhythmus im affektiven Höhepunkt, sowie eine Zunahme komplexer gestischer Bewegungen aus. Die Ausdrucksbewegung ist somit durch einen pragmatischen *shift* vom Erklären zum Anklagen und eine Veränderung in der Rhythmizität im affektiven Höhepunkt charakterisiert. Wie sich dies in

handelt sich um zwei Gesten mit je verschiedenen kommunikativen Funktionen: eine rekurrente Geste (S2), genauer den gestischen Präzisionsgriff (*precision grip*; Müller 2014; Kendon 2004) und eine Zeigegeste (S3), hier realisiert mit ausgestrecktem Arm und Zeigefinger. Dabei wird deutlich, dass der Präzisionsgriff S2 in der Anfangsphase der Anstiegs der affektiven Dynamik deutlich überwiegt und S3, also die Zeigegeste, exakt im Moment des pragmatischen Umschwungs von Erläuterung zu Anklage auftritt (Zeile 014). Dieser Umschwung von der dynamisch ansteigenden affektiven Qualität zum Plateau hoher Affektdynamik innerhalb der Ausdrucksbewegung realisiert sich damit nicht nur sprachlich, sondern auch auf der Ebene der kommunikativen Funktionen, der gleichzeitig verwendeten Gesten. Wie genau lässt sich dies nun aus Sicht der Gestenforschung beschreiben und begründen?

Die Ringgeste ist Teil des Repertoires rekurrenter Gesten und ist regelmäßig in metakommunikativer modaler Funktion zu beobachten, wo sie die Präzision einer sprachlichen Aussage artikuliert (Bressem/Müller 2014: 1583). Ihre kommunikative Funktion leitet sich aus der präzisen Greifbewegung ab, die erforderlich ist, um mit Zeigefinger und Daumen kleine Objekte ergreifen zu können. Sie wird daher auch als Präzisionsgriff beschrieben (Bressem/Müller 2014: Kap. 1583; Kendon 2004; Müller 2014). Wenig überraschend wird sie deshalb häufig im Kontext politischer Diskussionen verwendet (vgl. auch Neumann 2004). Mit der Ringgeste heben Sprecher*innen spezifische Aspekte des Gesagten nicht nur hervor, sondern behaupten gestisch deren Präzision und hohe Relevanz. So auch in der im Transkript dargestellten Äußerung, in der Ringgesten rhythmisch in einzelnen Schlägen auf Fokusakzente (HAUSHALT, FENSTER, *geNUG*, *GELD*) gesetzt oder in rhythmischen Auf- und Abbewegungen über längere Äußerungsteile getragen werden (z. B. »*da würde das geld zum*«). Die Ringgeste zieht sich im Wechsel zwischen einzelnen gestischen Höhepunkten und repetitiven Bewegungen über die komplette gestische Einheit, bis sie zum Ende von der Zeigegeste mit ausgestrecktem Arm abgelöst wird. Die rhythmisierte Verwendung der Ringgeste orchestriert somit den Anstieg der affektiven Dynamik. Der Wechsel von Ring- zu Zeigegeste markiert die Anfangsphase des Höhepunkts affektiver Intensität, und er geht mit einem Wechsel der kommunikativen Funktion der Handgeste einher. Nun werden die Störer auf den Seiten der FDP gestisch aufgezeigt (»*ja da können sie LACHen*«). Allerdings verschiebt sich die Funktion der Zeigegeste im Verlauf des affektiven Höhepunktes vom Hinweisen auf die Quelle der Störer im Publikum (d. h. einer referentiellen Zeigegeste) zum anklagenden Zeigen (d. h. eine pragmatische Zeigegeste) (Kendon 2004: 159) (»*h und wenn wir uns SO um die küm-*

mern; °h so verÄCHTlich wie SIE gerade gelAcht haben; °h dann sind DAS die nächsten, °h die die extre!MIS!ten wählen«). Mit dem Wechsel von der Ringzur Zeigegeste, der hohen Bewegungsdynamik der rhythmisierten Zeigegesten, der ostentativen Ausrichtung des Oberkörpers und Kopfes auf die Gruppe der Störer parallel zur hochaufgeladenen sprachlichen Anklage entfaltet sich der Höhepunkt der Affektdynamik.

Wir halten fest: Auch mit Blick auf die Entfaltung der Gesten zeigen sich die unterschiedlichen Rhythmisierungen, affektiven Dynamiken der Ausdrucksbewegung in den drei Phasen Anstieg, Plateau und Abfallen. In ihren wechselnden kommunikativen Funktionen (Präzision der Aussage, Identifikation der Störenden und Anklage) verstärken sie die Sprechhandlungen. Sie haben zudem durch die Verkörperung der dynamischen Bewegungsqualitäten einen erheblichen Anteil an der Formierung der Ausdrucksbewegung.

3.2.3.2 Ausdrucksbewegung und gestische Komplexität

Bereits bei einem ersten Blick auf die Sequenz wird deutlich, dass sich die hohe affektive Intensität der Ausdrucksbewegung sowie ihr dynamisches Profil mit einer Ausdehnung gestischer Einheiten und einer Häufung gestischer Höhepunkte (*Strokes*) verbindet. Der folgende Abschnitt betrachtet daher etwas genauer, wie sich Formen gestischer Komplexität über den Verlauf der gesamten Ausdrucksbewegung entfalten.

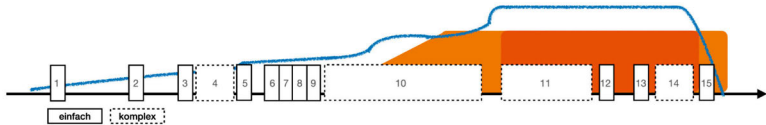
Müller (2024 a/b) betrachtet Gesten, aufbauend auf Kendons (2004) Ansatz, als lineare Zeitformen, die aus einer analytischen Perspektive in gestische Einheiten (*gesture units*) unterteilt werden können. Eine gestische Einheit ist dabei durch Ruhepositionen (rp) begrenzt (Kendon 2004: 111f.). Gestische Einheiten können weiter in einfache oder komplexe Formen unterschieden werden. Die Unterscheidung ist abhängig von der Anzahl der gestischen Phrasen (*gesture phrases*), die sich innerhalb einer gestischen Einheit entfalten. Gestische Phrasen umfassen, in Abweichung zu Kendon¹⁴, Vorbereitung, Höhepunkt und Rückbewegung der Geste. Eine einfache Einheit beinhaltet nur eine Phrase, während eine komplexe Einheit zwei oder mehr Phrasen umfasst (z. B. die in Abb. 7 visualisierte gestische Einheit 10).

Abbildung 8 zeigt eine schematische Darstellung der gestischen Einheiten und ihrer Komplexität, die die Ausdrucksbewegung in der zunehmend affektiv aufgeladenen Sequenz auszeichnet. Dargestellt auf dem Zeitstrahl als kleine

14 Kendon schließt in die gestische Phrase nur Vorbereitung und Höhepunkt der Bewegung ein (vgl. Kendon 2004: 111ff.).

Boxen, reihen sich insgesamt fünfzehn gestische Einheiten. Eine durchgezogene Linie verweist auf eine einfache, eine gestrichelte Linie auf eine komplexe gestische Einheit. Bereits auf dieser Ebene zeigt die Ausdrucksbewegung von Zimmermann eine deutliche Steigerung der Komplexität, insbesondere mit dem Beginn der ansteigenden affektiven Dynamik (orange). Während die gestischen Einheiten 1 bis 9 (mit Ausnahme von Einheit 4) kurze, simple Formen darstellen, die außerdem zu Beginn durch anhaltende Ruhephasen getrennt sind, sind die gestischen Einheiten 10 bis 15 deutlich komplexer und länger.

Abb. 8: Steigende gestische Komplexität im Verlauf der Ausdrucksbewegung



Diese mit der Entfaltung der Ausdrucksbewegung zunehmende Komplexität der gestischen Sequenzen wird auch mit Blick auf die gestische Phrasenstruktur deutlich. Zum einen steigt ihre Anzahl auf bis zu elf Phrasen in einer gestischen Einheit an und bildet damit eine rhythmisierte Bewegungssequenz (vgl. Einheit 10). Zum anderen werden die gestischen Phrasen an sich komplexer, da sie zunehmend aus Phrasen mit komplexen gestischen Höhepunkten (*›multi-strokes‹*) bestehen. Im Gegensatz zu einfachen Single-Stroke Phrasen (in Abb. 7 als S), umfassen diese mehrere Stroke Phrasen. Beispiele sind repetitive Gesten, wie die Kurbel-Geste (Ladewig 2014) oder das Nachzeichnen komplexer Formen. Alle komplexen gestischen Phrasen der Ausdrucksbewegung in der Rede Zimmermanns stellen Formen diskursiver Taktgebungen, *›discursive beatings‹* oder *›parsings‹* (Kendon 2004: 159) dar, also das rhythmische Auf- und Abbewegen der Hände (vgl. gestische Einheit 10).

Betrachten wir zudem die Entfaltung der gestischen Komplexität bezogen auf den fließenden Übergang verschiedener Handformen, dann zeigt sich, dass am Beginn der Ausdrucksbewegung einzelne gestische Betonungen und einfache gestische Einheiten (1–4, 6–9) überwiegen. Die Handformen sind hier vergleichsweise unspezifisch. Mit offener, lateral ausgerichteter Handfläche (PLOH; *palm lateral open hand*; vgl. auch Annotationssystem LASG von Bressemer et al. 2013) werden rhythmische Akzente als Auf- und Abbewegungen in Relation zur Äußerung gesetzt. Häufig fallen diese auf die Fokusakzente (*die*

DENKfabrik (006), *ist WICHTig-* (007)) oder diskursstrukturierende Elemente wie »deswegen« (010). Sie stellen klassische Formen diskursiven *beatings* dar und haben akzentuierende und redestrukturierende Funktion (Kendon 2004: Kap. 9). Wie lässt sich diese Zunahme der gestischen Komplexität über den Verlauf der Ausdrucksbewegung genauer beschreiben?

Am Beginn der Ausdrucksbewegung sehen wir viele kleine Gesten die im seitlichen Gestenraum, teilweise halb verborgen durch die Mikrofone, ausgeführt und schnell in Ruhephasen zum Rednerpult zurückgeführt werden. Mit der Zunahme der komplexen gestischen Phrasen verändert sich der Rhythmus von diesen kleinen Akzenten zu einer betonenden und stetig ratternden Dynamik. Dies geht auch mit dem Wechsel der Handform einher. So beobachten wir die pragmatische Ringgeste ausschließlich innerhalb komplexer Einheiten (4 und 10) und überwiegend in Multi-Stroke Phrasen mit einer Variabilität von zwei bis sechs Schlägen pro Phrase. Zusätzlich rückt die gestikulierende Hand weiter nach oben in das Zentrum des Gesten- und Bildraums und gerät damit ins Zentrum des Blickfeldes der Zuschauer vor Ort und vor den Bildschirmen. Während die Hände zuvor in die Ruheposition zurückgekehrt sind, verbleiben sie nun zwischen den einzelnen gestischen Taktschlägen in gehaltener Handform im zentralen Gestenraum. Der abrupte Wechsel der Handform zur Zeigegeste führt auch zu einer neuen Dynamik der gestischen Taktschläge. Deutlich eindringlicher und vor allem raumeinnehmender sticht der ausgestreckte Finger weit in den peripheren Gestenraum. Durch einen kleinen Schritt zurück ist Zimmermanns ausgestreckter Arm nun plötzlich mittig im Bildraum. Das Vor- und Zurückstrecken des Armes ist so in seiner vollen Länge zu sehen und erlangt dadurch hohe Prominenz. Es ist unmöglich, das anklagende Zeigen nicht zu sehen. Durch die neue Position sind auch die Vorbereitungsphasen aus den Ruhepositionen deutlich präsenter, sie wirken nun wie ein Anlaufnehmen für die ausladende Bewegung. Das rhythmische Rattern ist nun zu einem ausladenden Stechen geworden, das sich in ausholenden Bewegungen rhythmisch etabliert. Die letzte Zeigegeste (gestische Einheit 15) markiert das Ende des affektiven Ausbruchs wie einen Abschlusspunkt. In einer kurzen, abrupten Vorwärtsbewegung streckt Zimmermann die Hand mit ausgestrecktem Zeigefinger nach vorne, als würde er der FDP gewissermaßen »vor den Kopf« stoßen. Diese letzte Zeigegeste bildet den rhythmischen Abschluss des affektiven und argumentativen Höhepunkts, gefolgt von einem akzentuierten Nicken des Kopfes und einer nachdrücklichen Vorwärtsbewegung des gesamten Oberkörpers, all dies zeitgleich mit dem deutlich hervortretenden Fokusakzent »extre!MIS!ten« (Abb. 6: Zeile 020).

Betrachtet man die Entfaltung gestischer Komplexität im Zusammenspiel mit der prosodischen Dynamik über den Verlauf der gesamten Ausdrucksbewegung, so zeigt sich in ihrer Rhythmisierung die zunehmende Verdichtung gestischer Bewegungsmuster. Wir sehen eine Verschiebung von eher kleinen, unscheinbaren Gesten, die seitlich vom Körper und am Rande des Bildausschnittes ausgeführt werden (am Beginn der Ausdrucksbewegung), zu sehr raumgreifenden gestischen Bewegungen, die vor dem Körper deutlich sichtbar im Zentrum des Bildraums ausgeführt werden und so den affektiven Höhepunkt als ein Plateau, das sich über einen längeren Zeitraum hinweg erstreckt, orchestrieren und die Ausdrucksbewegung schließlich mit einer letzten pointierten Zeigegeste zu einem abrupten Ende führen.

3.3 Zusammenfassung

Die hier vorgestellte Systematik der Analyse sprachlicher und audiovisueller Multimodalität zeigt am Beispiel von Parlamentsreden, wie sich die Frage nach der Generierung, Mobilisierung und Modellierung von Zuschaueraffekten systematisch rekonstruieren lässt. Ausgangspunkt ist dabei das Konzept der Ausdrucksbewegung, das einen phänomenologischen Zugang zum Datenmaterial eröffnet. Einen Zugang, der die leibliche Erfahrung der Forschenden zum (intersubjektiv überprüfbar)en Ansatzpunkt für die Analyse audiovisueller multimodaler Daten nimmt. Dass dies keine esoterische Behauptung ist, haben die Analysen im deutsch-polnischen Forschungsprojekt zur affektiven Positionierung von parlamentarischen Redner*innen zeigen können. Am Beispiel der Analysen von Momenten hoher Affektdynamik ließen sich zweifelsfrei multidimensionale Erfahrungsgestalten am Material herauspräparieren, die dann einer detaillierten Mikro-Analyse unterzogen wurden. Schon der Blick auf eine Rede zeigt, dass sich auf diese Weise Affektdynamiken verschiedener Intensität und unterschiedlicher dynamischer Verläufe ausfindig machen lassen. Die Beschreibung der Bewegungsqualitäten der Ausdrucksbewegung erfolgt sodann als Beschreibung eines audiovisuell gestalteten Bildes, d.h. die Analyse des Sprechens und Gestikulierens, der Stimm- und der Körperbewegungen ist immer audiovisuell gerahmt, orchestriert. Zuschauer*innen an den Bildschirmen, auch wenn sie Forschende sind, sehen den/die Sprecher*in so, wie sie ihr von der Kamera (von der Regie) gezeigt wird. In unserem Fall scheint das unspektakulär: Wir sehen Zimmermann immer in der halbnahen Einstellung – nie schneidet die Regie auf die Totale (und verkleinert damit die Körperbewegungen auf Miniaturformat) – aber auch in dieser Einstellung ist unsere Ana-

lyse letztlich die eines Bildraumes, in dem wir den Oberkörper des Redners in frontaler Ansicht zu sehen bekommen. Manche Gesten sind von Mikrofonen verdeckt, die Interaktion mit dem Publikum können wir nur erahnen, denn auch diese ›liefert‹ uns die audiovisuelle Dokumentation der Rede nicht. Das heißt mithin, wenn wir von Ausdrucksbewegungen und ihren Bewegungsqualitäten als affektiven Qualitäten sprechen, dann sind dies immer Qualitäten, die als ganzheitliche Erfahrung eines audiovisuellen Bildes beschrieben werden. In diesem Sinne formulieren wir also die Position der Zuschauenden. Wir beschreiben damit unmittelbar, wie diese Bewegungsbilder uns berühren, bewegen, wie wir sie empfinden.

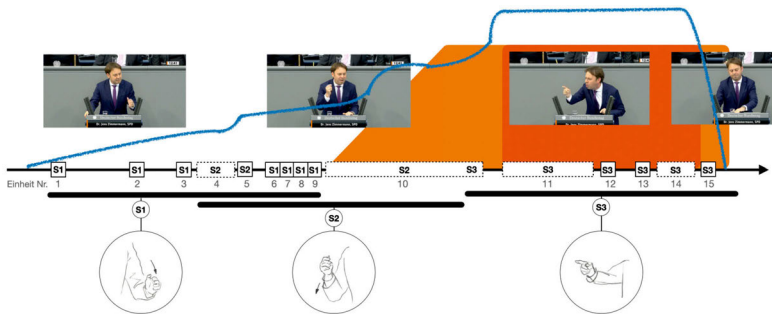
An diese phänomenologische Beschreibung schließt sich sodann eine Mikro-Analyse dessen an, was wir multimodale Orchestrierung der Ausdrucksbewegung nennen. Mit dieser Mikro-Analyse stützen wir die analytischen Einheiten, die in der phänomenologischen Analyse ausgemacht wurden, und beschreiben, wie sich die affektiven Dynamiken sprachlich-gestisch und in der audiovisuellen Inszenierung realisieren. Im vorliegenden Fallbeispiel stechen hier insbesondere die Rolle des interaktiven Geschehens sowie die sprachlich-gestischen Dynamiken als markant hervor. So zeigt sich, wie interaktive Trigger (hier die Einwürfe, Gelächter des Publikums), die sprachlich-gestische ›Doppelung‹ der kommunikativen Funktionen (von der Beschreibung zur Anklage), die Vergrößerung der entsprechenden Gesten, die damit ins Zentrum der visuellen Aufmerksamkeit rücken (*gestisches foregrounding*; Müller/Tag 2010) sowie die bemerkenswerte Zunahme an temporaler gestischer Komplexität (längere und komplexere gestische Einheiten) den Verlauf der Ausdrucksbewegung (gleichmäßiger Anstieg, Plateau hoher Intensität und abruptes Ende) ausmachen, ihn orchestrieren.

Abbildung 9 fasst die Analysen der multimodalen Orchestrierung noch einmal zusammen. Im Verlauf der Ausdrucksbewegung können drei verschiedene Handformen identifiziert werden: S1, die *palm lateral open hand* (PLOH); S2, der Ring (Daumen und Zeigefinger geschlossen); und S3, eine Zeigegeste mit ausgestrecktem Zeigefinger. Mit Blick auf die kommunikativen Funktionen zeigt sich, dass alle Gesten vordergründig pragmatische Funktionen haben.

Die multimodale Orchestrierung der Ausdrucksbewegung ist durch eine anwachsende Intensivierung charakterisiert, die sich als zunehmende Komplexität der Gesteneinheiten (von einfachen zu komplexen Einheiten) und Gesten-Phrasen (deutlich mehr Multi-Stroke Phrasen im Höhepunkt) realisiert. Gleichzeitig zeichnen sich die gestischen Bewegungen durch eine Zunahme der Rhythmisierung aus. Auch auf der Ebene der kommunikativen Funktion

lässt sich eine dynamische Veränderung beobachten, die mit der affektiven Qualität der Ausdrucksbewegung einhergeht. Während die Sequenz mit einer rein rhythmischen Betonung ohne spezifische Handform beginnt, wechselt dies im Verlauf der Sequenz über die Ringgeste, die die Präzision der zeitgleichen sprachlichen Äußerung behauptet, um plötzlich einen Sprechakt des Vorwurfs und der Anklage zu entfalten und schließlich abrupt in einem gestischen Abschlusspunkt zu enden.

Abb. 9: Gestische Formen mit pragmatischen Funktionen im zeitlichen Verlauf



Bezogen auf die audiovisuelle Multimodalität wird deutlich, dass trotz der »Zurückhaltung« der audiovisuellen Inszenierung eine klare Perspektivierung erfolgt. Die »Bühne« gehört eindeutig Zimmermann, der die gesamte Sequenz hindurch zentral im Bildraum steht. Seine Gesten sind gut zu sehen und können als rhythmische und pragmatische Komponenten der Ausdrucksbewegung zur Geltung kommen (anders etwa bei Redner*innen, die aus der Totalen gezeigt werden). Im Gegensatz dazu bleiben Zwischenrufe und Applaus aus dem Publikum rein akustisch und erscheinen damit eher zurückgesetzt. Sie sind zentraler Teil der Ausdrucksbewegung, auch wenn sie nicht dialogisch inszeniert sind, etwa durch Schuss-Gegenschuss-Montagen. Der affektive Höhepunkt der Ausdrucksbewegung ist deutlich geprägt durch die performative Sprechhandlung des Anklagens und Vorwerfens.

Am Beispiel der parlamentarischen Bundestagsrede von Jens Zimmermann (SPD) haben wir illustriert, wie ein medienästhetischer Ansatz für die Analyse audiovisueller und sprachlicher Multimodalität methodisch umgesetzt werden kann. Dabei wurde zunächst die Aufbereitung und materialbasierte Exploration multimodaler Daten jeweils bezogen auf audiovisuelle

und sprachliche Multimodalität vorgestellt, um dann die Ausdrucksbewegung als eine multidimensionale Erfahrungsgestalt als methodische Größe einzuführen. Die Analyse der multimodalen Orchestrierung der Ausdrucksbewegung hat gezeigt, wie die phänomenologisch-deskriptive Identifizierung von affektiven Höhepunkten, eingebettet in Momente hoher Affektdynamik, über die Beschreibung ihrer Bewegungsfigurationen empirisch rekonstruiert werden kann. Die Analyse der multimodalen Orchestrierung von Ausdrucksbewegung macht deutlich, dass sich die Ausdrucksbewegung als multidimensionale Gestalt dynamisch entfaltet, eine Gestalt, in der die Multimodalität des Sprechens immer mit der audiovisuellen Inszenierung verwoben ist. Es zeigt sich, dass Affektivität in Form von Bewegungsqualitäten und multimodalen Dynamiken nicht nur unmittelbar erfahrbar wird, sondern dass das gehörte, gesehene Sprechen immer in diese körperliche Erfahrung eingebettet ist.

Unser Ausgangspunkt ist damit die Affektivität der multidimensionalen Gestalt, nicht die sprachlich-gestische Gestaltung der Rede. Affektivität als Erfahrung gründiert den ›Inhalt‹ politischer Rede. Selbst eine zurückhaltende Inszenierung ist immer noch eine Rahmung und Positionierung des Redners als zentrale Figur. Dieser medienästhetische Ansatz, der die Wahrnehmung von Bewegungsbildern als Ausgangspunkt für die Analyse nimmt, hat die Verschränkung von sprachlicher und audiovisueller Multimodalität dokumentiert. Konsequenter ist der Ausgangspunkt also nicht die Sprachanalyse, sondern die Analyse der affektiven Modulation der Rede als leibliche Erfahrung eines sich in der Zeit entfaltenden audiovisuellen Bewegungsbildes.

Was die Zuschauer berührt ist weniger der Inhalt des Gesagten als das *Wie*. Diesem *Wie* sind wir hier auf verschiedenen Ebenen nachgegangen. Im Fall der Rede von Jens Zimmermann standen sprachlich-gestische und interaktive Aspekte im Vordergrund. Für andere Reden können dies andere Faktoren sein, die die Dynamik besonders triggern, auch kann die audiovisuelle Gestaltung stärker daran beteiligt sein, indem sie z. B. in Momenten hoher affektiver Expressivität eines Sprechers von der Halbnahen auf die Totale wechselt und damit die Sichtbarkeit des Körperausdrucks bis ins Mikroskopische verkleinert.

4 Abschließende Überlegungen und Fazit

Der vorliegende Beitrag hat einen medienästhetischen Ansatz zur Analyse der Multimodalität des Sprechens in audiovisuellen Medien vorgestellt, der einen

kulturwissenschaftlichen Zugang zu verschiedensten Phänomenen sprachlicher und audiovisueller Multimodalität eröffnet. Dabei haben wir argumentiert, dass die empirische Analyse multimodaler Sprachdaten, die in audiovisuellen Medien erscheinen, eine Reflexion und methodologische Berücksichtigung ihrer spezifischen medialen Ökologien notwendig macht.

Der transdisziplinäre Ansatz verbindet deshalb eine filmtheoretische Analyse der audiovisuell generierten Affektdynamiken mit einer linguistischen Analyse ihrer multimodalen Orchestrierung (Gesprächsanalyse, Gestenforschung, Interaktionsforschung). Dabei wurde deutlich, dass sich beide Perspektiven wechselseitig stützen. Ziel war es dabei, die Position der Zuschauer*innen zu rekonstruieren, die das Sprechen in audiovisuellen Medien als multidimensionale Bewegungs- und Zeitgestalten wahrnehmen – und nicht als isolierte Sätze, die aneinandergereiht, hier und da gestisch ausgeschmückt, oder von Wechseln der Kameraperspektive »abwechslungsreicher« gestaltet werden.

Indem wir von der körperlich erfahrbaren Entfaltung affektiver multidimensionaler Erfahrungsgestalten, sogenannten Ausdrucksbewegungen, ausgehen und nicht von einem spezifischen linguistischen Phänomen, nehmen wir einen wesentlichen Perspektivwechsel vor. Ein solch transdisziplinärer Ansatz deckt nicht nur die Komplexität und Vielschichtigkeit multimodalen Sprachgebrauchs auf, er ermöglicht auch eine Objektivierung der Wahrnehmungsdimension. Was wir in unserer Analyse beschrieben haben, sind keine subjektiven, individuellen Gefühle oder inneren Haltungen. Vielmehr handelt sich um ein geteiltes »Zuschauergefühl« (Kappelhoff 2018: 144–146; Kappelhoff/Bakels 2011), das sich in der Verschränkung mit den audiovisuellen Bewegungsbildern entfaltet. Die detaillierte Rekonstruktion der multimodalen Orchestrierung dieser Bilder kann gewissermaßen als »empirischer Nachweis« der »gefühlten« audiovisuellen Ausdrucksbewegung angesehen werden. In ihr fügen sich rhythmische, semantische und pragmatische Bedeutung in einem komplexen Zusammenspiel von Gesten, sprachlicher Äußerung und medialer Ökologie zu einer multidimensionalen Wahrnehmungsgestalt.

Wir schließen mit der Frage, was die Analyse vor einem kulturwissenschaftlichen Hintergrund zeigen kann. Dazu betrachten wir die Rede in ihrer Einbettung in mediale Ökologien als ein frei zugängliches Videodokument des offiziellen Archivs des Bundestages. Wir argumentieren, dass die Öffnung des Kontexts heraus aus der filmwissenschaftlich-linguistischen Betrachtung der medialen Ökologien hinein in gesellschaftliche Ökologien spezifische Formen einer Generierung, Modulation und Mobilisierung von Affekten aufdeckt,

die wie eine übergreifende Ebene der Bedeutung die Analysen rahmt und gesellschaftlich kontextualisiert. Wir beziehen uns hier auf Kappelhoff (2016; 2018). Er führt aus, dass sich im Moment des Sehens ein Wahrnehmungsraum entfaltet, in dem »die ästhetischen, semiotischen und pragmatischen Koordinaten tradierter Sinnkonstruktionen neu figuriert werden« (Kappelhoff 2018: 13f.). Dieser Wahrnehmungsraum ist immer eingebettet in ein Gefüge von sozial, kulturell und historisch situierter, gesellschaftlichen Ökologien (vgl. Kappelhoff 2018: 14, in Bezug auf Bourdieus Soziologie der legitimen Kultur). Die Inszenierung ist Teil eines etablierten »Medien-Konsums«, welcher wiederum das Sehen der audiovisuellen Daten formt. In diesem Konsum werden die Bewegtbilder zu einer »kulturellen Praxis« (Kappelhoff 2018: 10). Im Rahmen der im Archiv des Bundestages als audiovisuelles Dokument von Parlamentsdebatten der Öffentlichkeit zugänglich gemachten Rede von Zimmermann ist diese Wirklichkeit eine Form des geteilten Verständnisses von demokratischem Geschehen, in dem Streit und Auseinandersetzung akzeptierter und wichtiger Bestandteil politischen Handelns sind, welche jedoch in einem kontrollierten Rahmen verlaufen. Diese Kontrolliertheit zeigt sich in der Perspektivierung und der Affektmodulation. Wem wird wann die Bühne gegeben? Wie werden Streit und Auseinandersetzung inszeniert? Das Beispiel von Zimmermann verdeutlicht, dass selbst in Momenten direkten »Angriffs« aus dem Publikum, welche als Gefahr der primären Sprecherrolle angesehen werden können, kein Kontrollverlust dieser Rolle erfolgt. Die mediale Ökologie moduliert eine vermeintliche Neutralität, die den aktuellen Sprecher stützt. Es entfaltet sich ein Gefühl von Ordnung, Struktur und Kontrolliertheit. Die Kamera hält Zimmermann zentral, sie gibt ihm Raum. Die Auseinandersetzung wird nicht als ein gesichtsbedrohender Akt Zimmermanns erfahren. Es werden keine hämischen Gesichter in Großaufnahmen gezeigt, keine Schenkelklopfer oder Klamauk in Szene gesetzt. Die audiovisuelle Form der Rede reiht sich so in den Kanon des Archivs des Bundestages ein, in dem eine Form von demokratischem Handeln und Geschehen erfahren wird.

Mit einer medienästhetischen Perspektive auf die Multimodalität des Sprechens in audiovisuellen Medien verschiebt sich der Gegenstand sprachwissenschaftlicher Analyse von einer kontextfreien Sprachbetrachtung zu einer kontextreichen Sprachanalyse, die Sprache als sozio-kulturelles Phänomen in ihren medienkulturellen Ökologien und damit in einer kulturwissenschaftlichen Perspektive betrachtet.

Literaturangaben

- Bakels, Jan-Hendrik (2017): *Audiovisuelle Rhythmen. Filmmusik, Bewegungskomposition und die dynamische Affizierung des Zuschauers*, Berlin/Boston: De Gruyter.
- Bakels, Jan-Hendrik/Thomas Scherer/Jasper Stratil/Henning Agt-Rickauer (2020a): AdA Filmontology – a machine-readable Film Analysis Vocabulary for Video Annotation. In: Estill, Laura/Jennifer Guiliano (Hg.): *15th Annual International Conference of the Alliance of Digital Humanities Organizations*, DH 2020, Ottawa, Canada, July 20–25, 2020, Conference Abstracts.
- Bakels, Jan-Hendrik/Matthias Grotkopp/Thomas Scherer/Jasper Stratil (2020b): Computergestützte Filmanalyse im Spannungsfeld von Datenmodellen und Gestalttheorie. In: *montage AV* 29:1, 99–118.
- Bressemer, Jana/Cornelia Müller (2014): A repertoire of German recurrent gestures with pragmatic functions. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva Ladewig/David McNeill/Sedinha Tessendorf (Hg.): *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, Berlin/New York: De Gruyter Mouton, 1575–1591.
- Bressemer, Jana/Silva H. Ladewig/Cornelia Müller (2013): Linguistic annotation system for gestures (LASG). In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Sedinha Teßendorf (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, Berlin/New York: De Gruyter Mouton, 1098–1124.
- Bühler, Karl (1933): *Ausdruckstheorie. Das System an der Geschichte aufgezeigt*, Jena: Verlag von Gustav Fischer.
- Deppermann, Arnulf (2008): *Gespräche analysieren. Eine Einführung*, Wiesbaden: Verlag für Sozialwissenschaften.
- Goodwin, Marjorie/Asta Cekaite/Charles Goodwin (2012): Emotion as Stance. In: Sorjonen, Marja-Leena/Anssi Perakyla (Hg.), *Emotion in Interaction*, Oxford: Oxford University Press, 16–41.
- Greifenstein, Sarah (2020): *Tempi der Bewegung – Modi des Gefühls. Expressivität, heitere Affekte und die Screwball Comedy*, Berlin/Boston: De Gruyter.
- Greifenstein, Sarah/Dorothea Horst/Thomas Scherer/Christina Schmitt/Hermann Kappelhoff/Cornelia Müller (Hg.) (2018): *Cinematic Metaphor in Perspective. Reflections on a Transdisciplinary Framework*, Berlin/Boston: De Gruyter Mouton.
- Hickethier, Knut (2012): *Film- und Fernsehanalyse*, 5. aktualisierte und erweiterte Aufl., Stuttgart: J.B. Metzler.

- Horst, Dorothea (2018): *Meaning Making and Political Campaign Advertising. A Cognitive-Linguistic and Film-Analytical Perspective on Audiovisual Figurativity*, Berlin/Boston: De Gruyter.
- Horst, Dorothea/Franziska Boll/Christina Schmitt/Cornelia Müller (2014): Gesture as interactive expressive movement: Inter-Affectivity in face-to-face-communication. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressemer (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, Berlin/Boston: De Gruyter Mouton, 2112–2125.
- Kappelhoff, Hermann (2004a): Unerreichbar, unberührbar, zu spät. Das Gesicht als kinematografische Form der Erfahrung. In: *montage/av* 13:2, 29–53.
- Kappelhoff, Hermann (2004b): *Matrix der Gefühle. Das Kino, das Melodrama und das Theater der Empfindsamkeit*, Berlin: Vorwerk 8.
- Kappelhoff, Hermann (2008): Zuschauergefühl. Die Inszenierung der Empfindung im dunklen Raum des Kinos. In: Spiekermann, Geraldine/Beate Söntgen (Hg.), *Tränen*, München: Fink, 195–206.
- Kappelhoff, Hermann (2016): *Genre und Gemeinsinn. Hollywood zwischen Krieg und Demokratie*, Berlin/Boston: De Gruyter.
- Kappelhoff, Hermann (2018): *Kognition und Reflexion. Zur Theorie filmischen Denkens*, Berlin/Boston: De Gruyter.
- Kappelhoff, Hermann/Cornelia Müller (2011): Embodied meaning construction. Multimodal metaphor and expressive movement in speech, gesture, and feature film. In: *Metaphor and the Social World* 1:2, 121–153.
- Kappelhoff, Hermann/Jan-Hendrik Bakels (2011): Das Zuschauergefühl. Möglichkeiten qualitativer Medienanalyse. In: *Zeitschrift für Medienwissenschaft* 5:2, 78–96.
- Kendon, Adam (2004): *Gesture: visible action as utterance*, Cambridge/New York: Cambridge University Press.
- Ladewig, Silva H. (2014): The cyclic gesture. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressemer (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Interaction*, Berlin: De Gruyter Mouton, 1605–1618.
- Löffler, Petra/Florian Sprenger (2016): Medienökologien. Einleitung in den Schwerpunkt. In: *Zeitschrift für Medienwissenschaft* 14:1, 10–18.
- Merleau-Ponty, Maurice (1968): *The Visible and the Invisible*, Evanston, IL: Northwestern University Press.

- Michel, Sascha/Daniel Pfurtsc Keller (2021): »Ich bin seit Montag Zuhause in Quarantäne« – Zur Verbindung von Erzählen und Argumentieren in Social-Media-Kommentaren zu politischen Reden. In: *Zeitschrift für Literaturwissenschaft und Linguistik* 51:2, 255–279.
- Müller, Cornelia (1998): *Redebegleitende Gesten: Kulturgeschichte – Theorie – Sprachvergleich*, Berlin: Berlin Verlag Arno Spitz.
- Müller, Cornelia (2010): Wie Gesten bedeuten. Eine kognitiv-linguistische und sequenzanalytische Perspektive. In: *Sprache und Gestik* 1:41, 37–68.
- Müller, Cornelia (2014): Ring-gestures across cultures and times: Dimensions of variation. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bresse m (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Interaction* 2, Berlin: De Gruyter Mouton, 1511–1522.
- Müller, Cornelia (2019): Metaphorizing as embodied interactivity: What gesturing and film viewing can tell us about an ecological view on metaphor. In: *Metaphor and Symbol* 34, 61–79.
- Müller, Cornelia (2024a): Mimesis as as-if action. In: Zywiczynski, Przemyslaw/Slawomir Wacewicz/Monika Boruta-Żywiczyńska/Johan Blomberg (Hg.), *Perspectives on pantomime: evolution, development, interaction*, Amsterdam/Philadelphia: John Benjamins, 220–243.
- Müller, Cornelia (2024b): A toolbox for methods of gesture analysis. In: Cienki, Alan (Hg.), *The Cambridge Handbook of Gesture Studies*, Cambridge: Cambridge University Press, 182–216.
- Müller, Cornelia/Christina Schmitt (2015): Audio-visual metaphors of the financial crisis: meaning making and the flow of experience. In: *Revista Brasileira de Linguística Aplicada* 15:2, 311–342.
- Müller, Cornelia/Hermann Kappelhoff (2018): *Cinematic Metaphor. Experience – Affectivity – Temporality*, Berlin/Boston: De Gruyter Mouton.
- Müller, Cornelia/Susanne Tag (2010): The Dynamics of metaphor: Foregrounding and activating metaphoricity in conversational interaction. In: *Cognitive Semiotics* 6, 85–120.
- Neumann, Ranghild (2004): The semantics and pragmatics of everyday gestures. Proceedings of the Berlin conference, April 1998. In: Müller, Cornelia/Roland Posner (Hg.), *The conventionalization of the ring gesture in German discourse*, Berlin: Weidler, 217–223.
- Opiłowski, Roman (2017): Forschungsfelder der Medienlinguistik und neue Herausforderungen für Medienlinguistik 3.0. In: Iwona Bartoszewicz/

- Joanna Szczeł/Artur Tworek (Hg.), *Grenzen der Sprache – Grenzen der Sprachwissenschaft I*, Wrocław/Dresden: Neisse, 179–190.
- Scherer, Thomas J. J. (2024): *Inszenierungen zeitgenössischer Propaganda. Kampagnenfilme im Dienste des Gemeinwohls*, Berlin/Boston: De Gruyter.
- Scherer, Thomas/Sarah Greifenstein/Hermann Kappelhoff (2014): Expressive Movements in Audiovisual Media. Modulating Affective Experience. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressemer (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, Berlin/Boston: De Gruyter Mouton, 2081–2092.
- Schmitt, Christina (2020): *Wahrnehmen, fühlen, verstehen, Metaphorisieren und audiovisuelle Bilder*, Berlin/Boston: De Gruyter.
- Schmitt, Christina/Sarah Greifenstein/Hermann Kappelhoff (2014): Expressive Movement and Metaphoric Meaning Making in Audio-Visual Media. In: Müller, Cornelia/Alan Cienki/Ellen Fricke/Silva H. Ladewig/David McNeill/Jana Bressemer (Hg.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction*, Berlin/Boston: De Gruyter Mouton, 2092–2112.
- Selting, Margret/Peter Auer/Dagmar Barth-Weingarten/Jörg Bergmann/Karin Birkner/Elizabeth Couper-Kuhlen/Arnulf Deppermann/Peter Gilles/Susanne Günthner/Martin Hartung/Friederike Kern/Christine Mertzluft/Miriam Morek/Frank Oberzaucher/Jörg Peters/Uta Quasthoff/Wilfried Schütte/Anja Stukenbrock/Susanne Uhmann (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung* 10, 353–410.
- Sobchack, Vivian (1992): *The Address of the Eye. A Phenomenology of Film Experience*, Princeton: Princeton University Press.
- Stern, Daniel N. (2011): *Formen der Vitalität. Die Erforschung dynamischen Erlebens in Psychotherapie, Entwicklungspsychologie und den Künsten*, Frankfurt a.M.: Brandes & Apsel.
- Stöckl, Hartmut (2012): Medienlinguistik. Zu Status und Methodik eines (noch) emergenten Forschungsfeldes. In: Held, Gudrun/Christian Grösslinger/Hartmut Stöckl (Hg.), *Presstextsorten jenseits der »News«. Medienlinguistische Perspektiven auf journalistische Kreativität*, Frankfurt a.M.: Suhrkamp, 13–34.
- Xia, Qiang (2023): Turn-taking in video-mediated and co-present dialogues. A corpus-based study of German. In: *Register Aspects of Language in Situation (REALIS)* 2:4, 1–29.

Clara Kindler-Mathôt, M.A.
kindler@europa-uni.de

Cornelia Müller, Prof. Dr.
cmueller@europa-uni.de

Jana Junge, M.A.
junge@europa-uni.de

Katerina Papadopoulou, M.A.
Europa-Universität Viadrina
Große Scharnstrasse 59
15230 Frankfurt (Oder)
papadopoulou@europa-uni.de