

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 Sprachassistentensystem 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; Sprachassistentensysteme; 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. Sprachassistentensysteme 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 ähnlichen 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 Sprachassistentensystem *Alexa*, die der eines menschlichen sprachlichen Austauschs ähnelt: »Communication repair refers to the work of restoring shared understanding after conversational partners misunderstand 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 Missverständen 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 ›versteht‹, ›halluziniert‹, oder eben ›missversteht‹, 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. Statt dessen 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 Sprachassistentensystem *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 personenzentrierte 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 Dialogicity* 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 im 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 Echtzeiteignisse 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 aufzubauen, 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 Sprachassiszenzsystems 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 Sprachassistentensystemen 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 konsensualen 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-Tourangeau 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 Sprachassistentensysteme bestimmen.

3 Sprachassistentensysteme: 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 Bilderkennungs-, 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 Sprachverständigen 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 gesciptet, 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 videobasierten Ethnographie wird dabei nicht nur untersucht, wie eine Nutzerin des Sprachassistentensystems *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 Zwischenpiel 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 Sprachassistentensystem 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 Sprachassistenzsysteme

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 Beobachtung 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 auftreten.

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 verharret 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 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 (Exzert 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 naahaaran 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 Sprachassistentensystem 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 Sprachassistentensystems 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 verwenden

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 Sprachassistentensystem 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 Sprachassistentensystem 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 Sprachassistentensystem als mögliche*n Kommunikationspartner*in, der*die 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 Sprachassistentensystem *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 Khaleiji (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 Sprachassistentensystems, 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 from 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 entfaltete, 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 aktiver 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 entkörpert werden, um sie maschinentauglich zu machen. Es sind die Nutzer*innen, die diesen Prozess umkehren und die entkörperten 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 Languaging. 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 languaging. 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 Sprachassistentensysteme. In: *Sprache und Literatur* 51:2, 197–229.
- Hepp, Andreas/Wiebke Loosen/Stephan, Dreyer/Juliane, Jarke/Sigrid Kannengießer/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, Kristina/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. Propen/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, Alaistar (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 Mathemagics: 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