

Gutachten und Kommentare unter: <http://dp.jfml.org/2021/opr-habscheid-hector-hrncal-waldecker-intelligente-personliche-assistenten/>

## **Intelligente Persönliche Assistenten (IPA) mit Voice User Interfaces (VUI) als ‚Beteiligte‘ in häuslicher Alltagsinteraktion. Welchen Aufschluss geben die Protokolldaten der Assistenzsysteme?**

*Stephan Habscheid & Tim Moritz Hector & Christine Hrncal & David Waldecker*

---

### **Abstract**

The paper presents research results emerging from the analysis of Intelligent Personal Assistants (IPA) log data. Based on the assumption that media and data, as part of practice, are produced and used cooperatively, the paper discusses how IPA log data can be used to analyze (1) how the IPA systems operate through their connection to platforms and infrastructures, (2) how the dialog systems are designed today and (3) how users integrate them into their everyday social interaction. It also asks in which everyday practical contexts the IPA are placed on the system side and on the user side, and how privacy issues in particular are negotiated. It is argued that, in order to be able to investigate these questions, the technical-institutional and the cultural-theoretical perspective on media, which is common in German media linguistics, has to be complemented by a more fundamental, i.e. social-theoretical and interactionist perspective.

---

**Keywords:** Smart Speaker, Praxistheorie, Plattform, Infrastruktur, Mensch-Maschine-Dialog, Datenschutz

### **Kontaktperson:**

Prof. Dr. Stephan Habscheid  
Universität Siegen  
Germanistisches Seminar  
Hölderlinstraße 3  
57068 Siegen  
[habscheid@germanistik.uni-siegen.de](mailto:habscheid@germanistik.uni-siegen.de)

## 1. Phänomenbereich, Zielsetzung, Gegenstände und Aufbau des Beitrags

Mit der zunehmenden Verbreitung von Smart Home-Geräten halten Programme im Alltag Einzug, die in der Forschung generisch als „Intelligente Persönliche Assistenten“ (IPA) mit „Voice User Interfaces“ (VUI) bezeichnet werden (vgl. Porcheron et al. 2018). Konzipiert als Infrastrukturen „soziotechnische[r] Zukünfte“ (vgl. Strüver 2020: 2) sollen, so das Werbeversprechen, solche Systeme gesprochensprachlichen Input von Nutzer\_innen erkennen, im Fall einer Adressierung des Systems internetbasiert verarbeiten und in Verbindung mit synthetischen akustischen Sprachausgaben adäquat beantworten bzw. nach Sprachbefehlen elementare automatische Aufgaben (z. B. im Haushalt) erfüllen (vgl. Strüver 2020: 1–10). Dazu werden in technischer Hinsicht Funktionen von Speech Processing, Natural Language Processing und Information Retrieval miteinander vereint (vgl. Natale 2020: 5).

Zu den bekanntesten Systemen dieser Art gehört „Alexa“ in Verbindung mit den vielfältigen „Echo“-Geräten des Unternehmens Amazon (vgl. Strüver 2020: 1–10). So werden beispielsweise Smart Speaker mit Hilfe der entsprechenden Smartphone-App über das Internet mit einem Nutzerkonto bei Amazon verknüpft, und ein Wake Word zur Adressierung des Systems wird zugewiesen. Anhand verschiedener Farben einer LED-Leuchte am Gerät ist erkennbar, ob die eingebauten Mikrofone in Betrieb sind und ob, sofern dies der Fall ist, das System lediglich die akustische Umgebung nach dem Wake Word absucht oder, nachdem das Wake Word technisch erkannt wurde bzw. das Gerät sich im „Aufmerksamkeitsmodus“ befindet, den akustischen Input in der Cloud oder gelegentlich geräteseitig als Befehl versteht und verarbeitet. In diesem Zusammenhang haben, nicht zuletzt auch durch Medienberichte über technische Fehler, auch Fragen des Datenschutzes ein größeres öffentliches Interesse gefunden (vgl. Wissenschaftliche Dienste des Deutschen Bundestages 2019; Strüver 2020: 1–10).

In diesem Zusammenhang ist bemerkenswert, dass die Anbieter der IPA den Nutzer\_innen in der App (Teile der) Logdaten in einer vom System aufbereiteten Form zur Verfügung stellen. Wir bezeichnen diese Daten als „Protokolldaten“, da sie nicht den primären Zweck der Gerätenutzung darstellen, sondern diesen Gebrauch dokumentieren und protokollieren. Mit diesen Daten, genauer: der Frage nach ihrer wissenschaftlichen Verwendbarkeit wollen wir uns in diesem Beitrag näher beschäftigen. Darüber hinaus gehen wir auf die Frage ein, welche *potenzielle* alltägliche Nutzung der Protokolldaten durch die Aufbereitung in der App nahelegt wird. Zu der Frage, wie die User\_innen die Protokolldaten *tatsächlich* nutzen, lässt

sich auf der Basis der Protokolldaten selbst allenfalls spekulieren. Dies gilt auch und noch mehr für die Frage, wie solche automatisch erfassten Logdaten durch die Anbieter der Systeme bzw. durch Dritte zur Verbesserung der Systeme und Skills<sup>1</sup> und/oder im Kontext der Datenverwertung und u. U. des Datenhandels genutzt werden. In dem vorliegenden Artikel beschränken wir uns auf die Protokolldaten selbst und versuchen – über die nahegelegten alltäglichen Nutzungsweisen hinaus – zu explorieren, welchen Aufschluss ein methodisch reflektierter Umgang mit solchen Daten im Kontext einer sozialtheoretisch und techniksoziologisch informierten, praxeologischen Medienlinguistik und einer gesprächsanalytischen Untersuchung der sprachbasierten Nutzung von IPA und ihrer Einbindung in alltägliche Mehrparteieninteraktion geben kann.<sup>2</sup>

Hierzu wird im Anschluss an eine erste Beschreibung des Phänomenbereichs, der Darstellung des Forschungskontextes und der Einführung basaler Begriffe (Abschnitt 1) zunächst knapp der Forschungsstand rekapituliert (Abschnitt 2), dann – durchaus auch als Nebenziel mit eigenem Gewicht – ein geeigneter sozialtheoretischer Hintergrund für die Medienlinguistik erörtert (Abschnitt 3) und im Anschluss die Methodik der Datenerhebung und -aufbereitung dargelegt (Abschnitt 4). Es schließen sich Analysen an, die zunächst, in der gebotenen Kürze, techniksoziologische Untersuchungen als Grundlagen für die Medienlinguistik und dann die linguistischen Gegenstandsbereiche betreffen: In techniksoziologischer Perspektive geht es um den Status der Phänomene als Daten und die Verflechtung dieser Daten mit Infrastrukturen und Plattformen (Abschnitt 5.1), linguistisch um die Mensch-Maschine-Dialoge mit den IPA, die soziale Interaktion ‚um die IPA herum‘ (vgl. Porcheron et al. 2018) (Abschnitt 5.2) und schließlich die Dialektik von sprachlich-kommunikativen Nutzungspraktiken (Abschnitt 5.3) zwischen systemseitiger „Kuratierung“ (Dolata 2019: 195; vgl. Abschnitt 5.1) und nutzungsseitiger „Domestizierung“ (Brause/Blank 2020; vgl. Abschnitt 2). Der Beitrag schließt mit Fazit und Ausblick (Abschnitt 6).

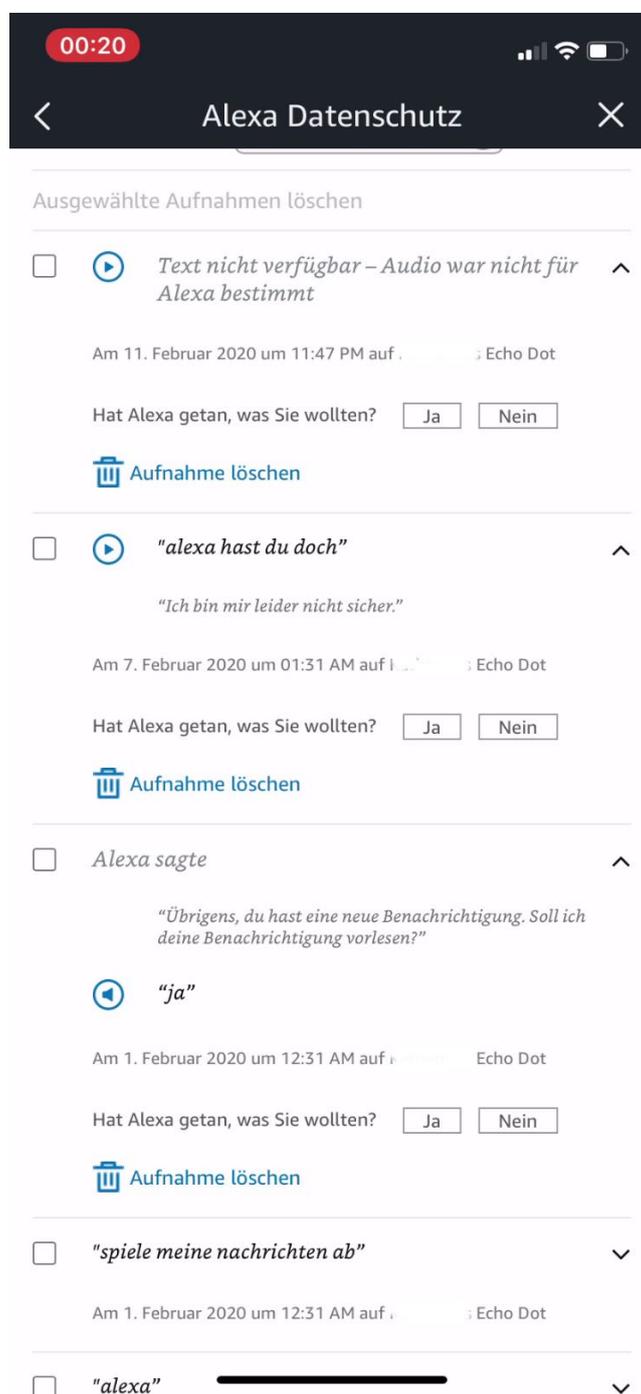
Betrachten wir zunächst das Phänomen an seiner Oberfläche. Im Fall der „Alexa“-App sind die Protokolldaten nicht vom Startbildschirm aus abzurufen, sondern an zwei hierarchisch tieferen Stellen in der App-Architektur einzusehen: Zum einen findet sich eine Darstellung der Protokolldaten in den „Einstellungen“ unter „Alexa-Datenschutz“ → „Sprachaufnahmenverlauf überprüfen“. Eine ähnliche

---

1 Durch die Installation von „Skills“ – Programm-Applikationen, die von Drittanbietern für das IPA-System von Amazon bereitgestellt werden – kann der vorinstallierte Funktionsumfang von „Alexa“ vielfältig erweitert werden.

2 Wir danken den Gutachter\_innen des Journals für Medienlinguistik für diverse hilfreiche Verbesserungshinweise zu einer früheren Fassung (Diskussionspapier).

Darstellung findet sich auch über den Eintrag „Aktivität“ im Menü der App. Die ausführlichere Darstellung unter dem Datenschutz-Menü ermöglicht explizit die Löschung einiger oder aller Einträge. Auf diese Variante wollen wir uns in diesem Beitrag fokussieren. Screenshot 1 zeigt die entsprechende Darstellung der Daten.



**Abbildung 1:** Screenshot: Protokolldatendarstellung unter dem Datenschutz-Menü

Die Daten werden schriftlich in Listenform dargestellt, wobei jeder Eintrag durch ein Antippen detaillierter angezeigt werden kann. In der eingeklappten Ansicht, in welcher hier der vierte Eintrag – „spiele meine nachrichten ab“ – zu sehen ist, können Datum und Uhrzeit der Aufzeichnung und das Gerät, das die Aufzeichnung erstellt hat, sowie der von „Alexa“ verstandene Text eingesehen werden. Zugleich kann jeder Eintrag mittels eines Kästchens am linken Rand ausgewählt und die Auswahl anschließend gelöscht werden (wie tief die Löschung systemseitig greift, kann anhand der Benutzeroberfläche nicht überprüft werden). Beim Aufklappen des Eintrags werden weitere Aktionen zu jedem Eintrag angeboten. So lässt sich hier jede von „Alexa“ aufgezeichnete Aufnahme auch mittels des blauen Pfeilsymbols akustisch abspielen. Außerdem wird die von dem IPA gesprochene Antwort in Textform wiedergegeben, sie selbst kann jedoch nicht angehört werden. Bei jedem Eintrag können die Nutzenden eine Rückmeldung geben, ob „Alexa“ den Anweisungen Folge geleistet hat. Zudem lässt sich jede Aufnahme mittels des Tippens auf ein blaues Mülleimer-Icon auch ohne Vorauswahl löschen. An dem Screenshot zeigen sich außerdem einige besondere Ausprägungen von „IPA-Dialogen“, wie wir die „neue Form von Dialogizität“ (Lotze 2020: 363) bzw. den „hybriden“ und „ambigen Austausch“ (Krummheuer 2010: 323–324) zwischen den Nutzenden und den Interaktion simulierenden technischen Systemen<sup>3</sup> hier bezeichnen wollen (Näheres in Abschnitt 5.2). Der erste Eintrag in dieser Liste zeigt eine Aufnahme, die als „nicht für Alexa bestimmt“ erkannt wurde; sie liegt nicht als transkribierter Text vor. An dieser Stelle wird für die Nutzenden unter anderem kontrollierbar, welche Interaktionen im Haushalt vom IPA aufgezeichnet wurden, ohne dass der IPA nutzerseitig dazu aufgefordert wurde. Der zweite Eintrag zeigt eine Interaktion, die zwar durch das *Wake Word* „Alexa“ ausgelöst wurde, jedoch vom IPA nicht verarbeitet werden konnte. Der dritte Eintrag wiederum zeigt eine jener – offenbar (derzeit noch) selteneren – Interaktionen, die nicht durch Nutzende, sondern durch den IPA initiiert wurden.

Unser Interesse an solchen Protokolldaten ist eingebettet in das Teilprojekt „Un/erbetene Beobachtung in Interaktion: Intelligente Persönliche Assistenten“ unter der Leitung von Stephan Habscheid

---

3 Zu bedenken ist bei der Anlehnung an solche Begriffe allerdings, dass sich die einzelnen Studien mit sehr unterschiedlichen Arten von Conversational Agents befassen, die sich wiederum von Smart Speakers mit IPA/VUI, wie sie im Mittelpunkt des vorliegenden Beitrags stehen, unterscheiden: Chatbots in den Studien von Lotze (2016; 2018; 2020), einem virtuellen Embodied Conversational Agent, der über eine Tastatur bedient wird, in der Untersuchung von Krummheuer (2010). Wieder anders im Blick auf die Medialität und den Kontext liegt der Fall bei einem Museumsroboter, wie ihn Pitsch et al. (2017) untersuchen (vgl. dazu Abschnitt 5.3 unten).

und Dagmar Hoffmann (seit 2020)<sup>4</sup> im Sonderforschungsbereich (SFB) 1187 „Medien der Kooperation“.<sup>5</sup> Im Rahmen dieses interdisziplinären, im Grenzbereich von Medienlinguistik und Mediensoziologie angesiedelten Projekts gehen wir dem eingangs skizzierten Innovationsversprechen bzw. den kritischen Fragen im Zusammenhang mit IPA nach und versuchen besser zu verstehen, wie derartige Systeme in Verbindung mit Plattformen und Infrastrukturen operieren, wie die Dialogsysteme heute gestaltet sind, wie Nutzer\_innen die VUI bzw. IPA in ihre soziale Interaktion, in Alltagspraxis und -diskurs einbinden und wie sie dabei mit Belangen des Datenschutzes umgehen. Ein Teilziel besteht darin, zu untersuchen und mithin zu explorieren, welchen Aufschluss die Protokolldaten in verschiedenen Analysefeldern geben können.

Am oberen Rand des Screenshots 1 ist eine Sekundenanzahl auf rotem Grund zu sehen; diese ist den Umständen der Datenerhebung und der Einbettung der Daten in die Plattform von Amazon geschuldet: Die Audiodaten sind zwar über die App abspielbar und löschar, aber nicht ohne Weiteres zu exportieren. Eine Möglichkeit, die wir als Forschende gesehen haben, um sie in ihrem Nutzungskontext zu erheben, bestand in der Bitte an Datenspendende, ein Bildschirmvideo vom sukzessiven Abspielen der Audio-Daten zu erzeugen. Wie wir mit diesen Videos, die uns zur Verfügung gestellt wurden, weiterarbeiten, wird in Abschnitt 4 zur Methodik näher erläutert.

Auf den Ebenen von Alltag und Wissenschaft basal für unseren Gegenstand, wie er im Titel des Beitrags bezeichnet wird, ist eine Grundannahme des SFBs: Demnach sind *Daten* nicht als Objekte sui generis für die medienwissenschaftliche Forschung interessant, sondern als Teil von Praxis, in deren Kontext Daten jeweils kooperativ hervorgebracht bzw. genutzt werden. In diesem Sinn verweist der Begriff der Datenpraktiken im Kontext des SFBs darauf, dass mit Daten umgegangen werden muss, damit sie *als* Daten je situativ relevant werden. *Praxis* wiederum muss sich in ihrem Vollzug nicht nur an anderen Beteiligten – dem Alter des Ego in der klassischen Sozialtheorie – ausrichten, sondern auch an den materiellen Gegebenheiten einer Situation. An den Protokolldaten wird dies insofern deutlich, als diese nicht nur die Aufzeichnung aus einem Haushalt auf einer App abrufbar machen, sondern dabei in das für die Nutzenden weitestgehend opake, weltumspannende soziotechnische System der Kommunikationsinfrastruktur und in die Rechenkapazitäten sowie betrieblichen Plattformen der Betreiberfirmen

---

4 Das Vorläuferprojekt in der ersten Förderphase des SFBs (2016–2019) wurde von Wolfgang Ludwig-Mayerhofer geleitet.

5 Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – SFB-Geschäftszeichen Projektnummer 262513311 (SFB 1187 „Medien der Kooperation“).

und ihrer Kooperationspartner\_innen eingebunden sind, einschließlich der hierfür erforderlichen materiellen und menschlichen Ressourcen im Kontext einer globalen politischen Ökonomie (vgl. Crawford/Joler 2018). Solche Gegebenheiten sind in der Literatur unterschiedlich theoretisch konzeptualisiert worden: Während Latours (2007) Programm einer Akteur-Netzwerk-Theorie mit dem Begriff „Aktanten“ darauf abzielt, Menschen und Nicht-Menschen in Bezug auf ihre *Agency* auf die gleiche Stufe zu stellen, zählen ontologisch weniger radikale bzw. agnostische Theorien aus dem praxistheoretischen Umfeld technische und andere Geräte zu den „materiellen Partizipanden des Tuns“ (Hirschauer 2004), da sie in der Praxis und für die Praxis „rekrutiert“ werden. In diesem Sinn ist auch unsere Bezeichnung der IPA als „Beteiligte“ zu verstehen.

Vor diesem Hintergrund fragen wir im vorliegenden Beitrag danach, welchen Aufschluss die Protokolldaten der IPA potenziell den Nutzer\_innen, besonders aber den im Rahmen eines rekonstruktiven Ansatzes forschenden Wissenschaftler\_innen darüber geben können,

- wie die IPA-Systeme durch ihre Verbindung mit Plattformen und Infrastrukturen operieren (Abschnitt 5.1),
- wie die Dialogsysteme heute gestaltet sind und wie Nutzer\_innen in Zwei- und Mehrparteienkonstellationen diese in ihre alltägliche soziale Interaktion einbinden (Abschnitt 5.2),
- in welche alltagspraktischen bzw. diskursiven Kontexte die IPA systemseitig gestellt sind bzw. in der nutzungsseitigen sozialen Interaktion gestellt werden und wie auf beiden Ebenen insbesondere Belange des Datenschutzes verhandelt werden (Abschnitt 5.3).

Um diese Fragen untersuchen zu können, sind die technisch-institutionelle und die kulturtheoretische Betrachtung von Medien, wie sie in der germanistischen Medienlinguistik weithin üblich sind, um eine elementare, sozialtheoretische und interaktionistische Perspektive zu ergänzen (Abschnitt 3). Werfen wir aber zunächst einen kurzen Blick auf den Forschungsstand.

## 2. Zum Forschungsstand

Obwohl sie noch nicht lange verfügbar sind, sind IPA mit VUI bereits mit unterschiedlichen disziplinären und methodischen Zugängen untersucht worden. Die folgende Darstellung konzentriert sich auf IPA-Systeme, die stationär in der häuslichen Umgebung eingerichtet sind und auch als Smart Speaker bezeichnet werden. Es treten allerdings Überlappungen mit anderen Typen von VUI auf, die etwa in andere Geräte integriert sind (etwa „Siri“ auf Apples iPhones oder Sprachsteuerungsanwendungen im Auto).

Zu den Schwerpunkten der bisherigen Forschung gehört die kritische Untersuchung von Gender-Aspekten durch stimmliche und mediale Inszenierung und „Vermenschlichung“ der Geräte (vgl. z. B. Both 2014; Phan 2017; Strengers/Nicholls 2018; Hennig/Hauptmann 2019; Natale/Cooke 2020). Andere gesellschaftliche Aspekte wie sozialen Status und Bildungsniveau nehmen etwa Schiller/McMahon (2019) oder Phan (2019) aus einer kritischen Perspektive in den Blick. Herausgestellt wird in diesen Studien mit unterschiedlichen Vorgehensweisen die Manifestierung gesellschaftlicher Stereotype bzw. Ungleichheiten durch das (u. a. stimmliche) Design der Geräte und deren Darstellung in der Werbung.

Ein weiterer Schwerpunkt der Untersuchungen waren Datenschutz- und Privatsphäre-Fragen. Ford/Palmer (2019) zeigen etwa in einer Netzwerktraffic-Analyse, dass die durch IPA beim „Absuchen“ nach dem Wake Word aufgezeichneten akustischen Signale zwar nicht (wie mitunter befürchtet) sämtlich an Cloud-Anbieter übertragen werden, dass aber möglicherweise einzelne Datenübertragungen erfolgen, die weder durch ein Aktivierungswort ausgelöst noch in der Smartphone-Anwendung protokolliert werden (siehe auch Gray 2016; Apthorpe et al. 2017). Bezüglich der rechtlichen Zulässigkeit kommt ein Papier der Wissenschaftlichen Dienste des Deutschen Bundestages (2019) zu dem Schluss, dass die Vorgaben der Datenschutz-Grundverordnung zwar wahrscheinlich eingehalten werden, dies aber aufgrund der Intransparenz der Auswertung schwierig zu überprüfen ist und IPA im Hinblick auf den Schutz von Dritten, die von der akustischen Aufzeichnung nicht wissen, und von Minderjährigen dennoch problematisch sein könnten (siehe auch Datenethikkommission der Bundesregierung 2019). Bei den Nutzer\_innen zeigt sich in ersten Studien ein resignativer Pragmatismus im Hinblick auf die Privatsphäre (vgl. Lau/Zimmermann/Schaub 2018: 18).

Ein dritter Schwerpunkt waren Nutzungsstudien zu IPA, in denen belegt werden konnte, dass oft einfache Alltagsroutinen vorherrschen, und zwar sowohl für den amerikanischen (vgl. Luger/Sellen 2016; Lopatovska et al. 2019) wie auch für den deutschen Kontext

(vgl. Pins et al. 2020). Vielfältigere Nutzungsweisen werden besonders erkennbar in Studien, die nach der „Domestizierung“ der technischen Systeme durch die Nutzenden fragen und methodisch weniger standardisiert vorgehen (vgl. Brause/Blank 2020). Dazu gehört auch die Studie von Porcheron et al. (2018), die mit einem ethnografisch-konversationsanalytischen Vorgehen die Einbindung von IPA in häusliche Alltagsaktivitäten und deren sequenziellen Ablauf untersucht und dabei auf Audio-Aufzeichnungen zurückgreift, die das sprachliche Material *vor* und *nach* der Nennung des Aktivierungsworts miteinschließen.

Methodisch wurde ansonsten bei der Betrachtung der Nutzung von IPA bisher überwiegend mit Befragungen, Umfragen und Dokumentationsinstrumenten (z. B. mit Tagebüchern) gearbeitet. Nur wenige Studien werten bisher die Protokolldaten der Nutzer\_innen aus. Diese konzentrieren sich zumeist auf eine quantitative Betrachtung der Nutzung. So können Bentley et al. (2018) bestätigen, dass bei einer Betrachtung über (durchschnittlich) drei Monate einfache Nutzungsszenarien dominant sind und wenige neue Dienste ausprobiert werden (vgl. auch Ammari et al. 2019; Sciuto et al. 2018). Die Protokolldaten werden dabei meist in Kombination mit Nutzungsdokumentation und Befragungen eingesetzt. Die Aufzeichnungen erwiesen sich auch bei einer Untersuchung zur Personifizierung und ontologischen Kategorisierung von IPA-Dialogen als nützlich, wurden aber auch hier zusätzlich zu Interviews herangezogen (vgl. Pradhan et al. 2019). Es zeigte sich dabei, dass die Anwender\_innen von IPA zwischen einer Kategorisierung der Geräte als „humanlike“ und „object-like“ schwanken und häufig fließende Übergänge entstehen (zu diesen Übergängen siehe auch Krummheuer 2010 sowie Abschnitt 5.2). Die qualitative Auswertung von automatisiert generierten Protokolldaten konnte auch in anderen Zusammenhängen genutzt werden. So kommt eine solche Analyse (hier von Chatbot-Logs) z. B. bei Candello/Pinhanez (2018) auch zum Einsatz, um Fehler in der Gestaltung der Mensch-Maschine-Dialoge seitens der Maschine zu erkennen und zu analysieren.

### 3. Theoretischer und konzeptueller Hintergrund

In der Debatte über den Medienbegriff im Kontext der Germanistischen Linguistik stehen traditionell ‚Kommunikationsformen‘ im Mittelpunkt: Hierunter versteht man Strukturbedingungen von Kommunikation und Sprachgebrauch, die durch die Verwendung technischer Artefakte (Medien i. e. S.), z. T. auch durch Medieninstitutionen, geprägt sind (z. B. Infrastrukturen und Plattformen der kommerziellen IPA-Systeme). Neben den an Genre-Konventionen

(z. B. Wetterabfragen) gebundenen empirischen Ausprägungen können so auch (noch) nicht oder nur im Ausnahmefall, mehr oder weniger kreativ genutzte kommunikative *Potentiale* (z. B. Erproben und Austesten, Herausforderung und/oder Verspottung der IPA unter Anwesenden<sup>6</sup>) ins Blickfeld kommen (vgl. Brock/Schildhauer 2017; Habscheid 2020). Zum anderen versteht man unter Medien (i. w. S.) kulturell verfestigte Techniken oder Verfahren, von denen technische Medien ein Bestandteil sein können: Solche „medialen Verfahren“, wie z. B. durch Sprachassistenzsysteme vermittelte Kommunikation mit digitalen Plattformen, bilden demnach die materiale und prozedurale Seite des Gebrauchs von Zeichen. Dieser ist zudem eingebettet in konventionelle (kommunikative) Praktiken (z. B. Wissensrecherche; Internet-Shopping) und hängt außerdem von der individuellen Kompetenz der (Zeichen-)Verwender\_innen ab (vgl. Schneider 2017: 45).

Nicht im Blickfeld liegt in beiden Fällen, wie Medien – jenseits ihrer Potentiale und der kreativen Beteiligung von Individuen – als *soziale* Instanzen zustande kommen und auf der Grundlage ihrer Materialität geprägt werden (vgl. Meiler 2019), dynamisch variieren und sich in der Zeit wandeln können. Um diese Fragen untersuchen zu können, sind die technische und die kulturtheoretische Betrachtung von Medien, wie sie in der germanistischen Medienlinguistik weithin üblich sind, um eine elementare, sozialtheoretische Perspektive zu ergänzen. In einer praxeologischen Perspektive, wie sie im SFB 1187 „Medien der Kooperation“ an der Universität Siegen verfolgt wird, werden Medien verstanden als „kooperativ erarbeitete Kooperationsbedingungen“ oder, kurz gesagt, als „Medien der Kooperation“ (Schüttpelz 2016: 5). Diese Kooperation wird als den Zeichen, deren materialen und prozeduralen Strukturbedingungen und ggf. konventionellen Nutzungsweisen, einschließlich der Herausbildung und Veränderung von Symbolsystemen, *logisch* vorgelegt aufgefasst (vgl. Meyer/Ayaß 2012: 14–15). Auch *Daten*, wie sie im vorliegenden Beitrag fokussiert werden, sind nur in Zusammenhängen kooperativer sozialer Praxis zu verstehen (vgl. Abschnitt 3).

Ein solcher sozialtheoretischer Ansatz, der den Begriff der Praxis zum Dreh- und Angelpunkt der Theoriebildung macht, fordert kulturalistische Medientheorien und das Konzept der Kommunikationsformen gleichermaßen heraus. Diese Position, wie sie in der konversationsanalytischen Tradition von Charles Goodwin (2018) bzw. im Grenzbereich von Konversationsanalyse und Medientheorie durch Erhard Schüttpelz und Christian Meyer (2017) entwickelt wurde, blendet kulturelle Verfestigungen und Übereinkünfte (wie Techniken, kommunikative Gattungen oder Symbole) keineswegs

---

6 Vgl. Krummheuer (2010: Kap. 9) sowie Abschnitt 5 unten.

aus, geht jedoch auf einer elementareren Ebene der Ontologie nicht von diesen aus, sondern stellt stattdessen das Konzept der wechselseitigen ‚Praxis‘ (Schüttpelz/Meyer) bzw. der ‚Co-operative Action‘<sup>7</sup> (Goodwin) in den Mittelpunkt: Beteiligte an der Herstellung von Sinn verfertigen demnach wechselseitig Abläufe, indem sie jeweils die von ihren Vorgänger\_innen ins Spiel gebrachten, zeichenförmig geprägten materiellen Ressourcen (nicht zuletzt indexikalische Verweise)<sup>8</sup> partiell aufgreifen und in der Verfolgung ihrer Ziele transformierend wiederverwenden. Auf dieser Grundlage können sich mit dem Effekt einer Vereinfachung der Verständigung Konventionen herausbilden und verändern, Medien selbst – ohne die Kommunikation nicht denkbar ist – werden jedoch elementarer gefasst (wir kommen darauf zurück). Theoriebildungen, die Medien konstitutiv an Konventionen binden, greifen insoweit zu kurz (vgl. Habscheid 2020: 2–3).

Auch Brock und Schildhauer (2017) gehen von der Überlegung aus, dass Kommunikationsformen aufgrund ihres auch potentiellen Charakters stets offen sind für unterschiedlichste kommunikativ-funktionale Nutzungen, die sich auf längere Sicht zu Genres verfestigen *können*.<sup>9</sup> Dieser Aspekt mag kulturwissenschaftlich unbedeutend erscheinen, insofern das *Potential* einer Kommunikationsform nicht sinnvoll zu ergründen ist: Der Kreativität der Nutzer\_innen und der Eigendynamik interaktionaler Ordnungsbildungen in situierter Kommunikationsprozessen sind keine Grenzen gesetzt. Gleichwohl lassen sich aus dem Aspekt der Potentialität von Kommunikationsformen zwei wichtige Einsichten ableiten: Zum einen betonen Brock und Schildhauer, dass im Blick auf die konkrete Nutzung von Kommunikationsformen im Rahmen von Genres zwischen *notwendigen* und nur *typischen* Eigenschaften systematisch unterschieden werden sollte. Zum anderen beugen Brock und Schildhauer mit ihrer Betonung des Potentials und der Offenheit von Kommunikationsformen einem konventionalistisch verengten Blick auf kommunikative und sprachliche Praxis vor, sei es in Bezug auf Genres, sei es hinsichtlich der Medien/Kommunikationsformen. Hier wiederum kann eine praxeologische Perspektive unmittelbar anschließen.

---

7 Wobei der Bindestrich den operativen, inkrementellen Aspekt der durch situierter Zeichengebrauch vermittelten kooperativen Verständigung markiert.

8 Für eine praxeologische Beschreibung verschiedener sprachlicher Zeichentypen, die den Status von Symbolen und Repräsentation relativiert, vgl. Meiler (2019: 68–72).

9 Diese sind selbst vielfach (wenn nicht immer) polyfunktional (vgl. Brock/Schildhauer 2017: 20–21; Hausendorf et al. 2017: 229–271).

Interaktionale, soziale und kognitive Prozesse liegen nach Goodwin (2018: 1) vielfältigsten Dimensionen und Aspekten des (in vielerlei Hinsicht einzigartigen) menschlichen (Zusammen-)Lebens zugrunde (vgl. Schüttpelz/Meyer 2018): Dazu zählen gleichsam kleinteilige Aspekte wie die multimodale und sequenzielle Interaktion, aber auch Sequenzen und Schichten von Handlungen und letztlich die funktionale und historische Ausdifferenzierung von Gesellschaften. Goodwin zeigt auf, wie neue Handlungen systematisch ko-operativ hervorgebracht werden, also jeweils zustande kommen und verständlich werden durch sinnhaft strukturierende und geschichtete Operationen auf dem, was andere an Ressourcen und Lösungen zuvor geschaffen haben (Goodwin 2018: 431; vgl. Schüttpelz/Meyer 2018: 179–180). Auf diese Weise können nach Goodwin Erträge menschlichen Handelns transformiert, akkumuliert und tradiert werden, und Akteur\_innen können nur vor diesem Hintergrund ihre situative bzw. situationsübergreifende Handlungsmacht erlangen (Goodwin 2018: 440).<sup>10</sup>

Dem Konzept „co-operative action“ bei Goodwin entspricht – bei aller Kritik an der Theoriearchitektur und Begriffsbildung im Einzelnen (vgl. Schüttpelz/Meyer 2018) – bei Schüttpelz und Meyer in etwa der Begriff der „Praxis“, hier verstanden als „das in einer wechselseitigen Verfertigung befindliche Geschehen“ (Schüttpelz/Meyer 2017: 158). Im Rahmen dieser Theorie ist die Praxis „allen anderen sozialen Größen vorzuordnen“: „Kooperation, Interaktion, Praktiken, Handlungen, Routinen, Techniken, technische Medien werden ‚in der Praxis‘ hervorgebracht“ – und das heißt eben: „in einem sich in wechselseitiger Verfertigung befindenden Geschehen“ (Schüttpelz/Meyer 2017: 159). Demnach können durch Praktiken und Handlungen Routinen und Techniken ausgebildet werden, die jedoch nicht mit den Praktiken zur Deckung kommen, vielmehr beruhen Praktiken auch auf einer „wechselseitigen Improvisation“ (Schüttpelz/Meyer 2017: 156).

Wichtig in unserem Kontext erscheint, dass das grundlegende Prinzip der wechselseitigen Verfertigung – wie übrigens auch bei Goodwin – nicht nur für die Inhalte, sondern auch für die Mittel der Verständigung gilt, also auch für an Medien gebundene Zeichen (um an Schneider 2017 anzuknüpfen). Auch diese Mittel liegen nicht einfach bereits vor, sondern sie müssen stets aufs Neue situativ und ko-operativ hervorgebracht werden. Vor diesem Hintergrund werden, wie bereits erwähnt, bei Schüttpelz Medien allgemein gefasst als

---

<sup>10</sup> Auf einen ähnlich umfassenden Gegenstandsbereich richtet sich die *Nexus Analysis* von Scollon/Scollon (2004) (Hinweis J. Androutsopoulos). Eine vergleichende Befassung mit diesem Ansatz ist im Rahmen des vorliegenden Beitrags nicht möglich.

„kooperativ erarbeitete Kooperationsbedingungen“ (Schüttpelz 2016: 5). Nach Goodwin (2018: 445) gehört es zu den Grundprinzipien von Praxis, dass sie semiotisch „opportunistisch“ und „gefräßig“ ist: Alle Arten von wahrnehmbaren Materialien können zur lokalen Konstruktion von Handlungen einbezogen und damit zu sinnhaft strukturierenden und strukturierten Medien werden (Goodwin 2018: 445), einschließlich aller Elemente der Systemebenen von Sprache in der Interaktion (vgl. Schüttpelz/Meyer 2018: 179–180; 182).

An dieser Stelle muss betont werden, dass eine derartige Kooperation *Wechselseitigkeit* voraussetzt, aber nicht unbedingt *Gemeinsamkeit* im Sinne bereits geteilter Ziele, Werte, Bräuche, Zeichensysteme usw. Schüttpelz und Meyer (2017: 6) verweisen an anderer Stelle auf die in der englischsprachigen Forschungsliteratur weithin übliche Unterscheidung von *mutual* („wechselseitig“), wie in *mutual constitution, assistance, repair* usw., und *common* oder auch *joint* oder auch *shared* („gemeinsam“), wie in *common goals, means, actions* usw. Derartige Gemeinsamkeiten stellen das Ergebnis und eine Erleichterung von Kooperation dar, nicht jedoch deren notwendige Voraussetzung.

Damit positioniert sich der Ansatz nicht zuletzt auch in einer sozialtheoretischen Tradition. Wenn Menschen im Alltag miteinander handeln, stellt die bewusste, rational reflektierte Verständigung über Ziele, Mittel usw. stets nur einen kleinen Ausschnitt – sozusagen die Spitze des Eisbergs (Garfinkel 2012: 56) – dessen dar, was für die Verständigung tatsächlich relevant ist. Weite Teile dessen, was in der Situation relevant ist, müssen als fraglos gegeben erachtet werden (Garfinkel 2012: 56–57). Mit anderen Worten: Wann immer in der Kommunikation ein Konsens erreicht wird, muss, wie Meyer und Ayaß es formulieren, eine „außer-konsensuale“ — man könnte auch sagen: vor-vertragliche — „Grundlage“ bereits gegeben sein, „aus welcher der Konsens überhaupt erst entstehen und begründet werden kann“ (Meyer/Ayaß 2012: 14; vgl. Durkheim 1992). Umgekehrt ist in der Perspektive der Praxistheorie aber auch eine Begründung von Kooperation durch gesellschaftlich-kulturelle Konventionen für sich genommen unbefriedigend. So ging etwa Harold Garfinkel in kritischer Distanz zur Theorie seines Lehrers Talcott Parsons davon aus,

„dass Normen keine externen, abstrakten Leitlinien sein können, sondern von den Akteuren selbst auf der Grundlage einer verkörperten und reflexiven Sozialität lokal hervorgebracht, verwaltet und situationsspezifisch im Hier und Jetzt angewendet werden müssen.“ (Meyer/Ayaß 2012: 14).

An die zentrale Stelle der Sozialtheorie tritt damit ein Begriff von ‚Praxis‘ im Sinn von Schützpelz und Meyer bzw. von ‚Ko-operativität‘ im Sinn Goodwins.

#### 4. Methodische Überlegungen und Korpusbeschreibung

Im Sinne einer so konturierten linguistischen Praxeologie können auch IPA-Systeme über ihre Medien- und Datenpraktiken verstanden werden: auf der Seite der Nutzer\_innen im „front end“, auf der Seite der Anbieter\_innen als Verarbeitung und Verwertung im „back end“ (siehe auch Zuboff 2018). Schwerpunktmäßig rücken wir die potentiellen Daten- und Medienpraktiken der Nutzer\_innen auf der „Vorderseite“ der IPA-Systeme in den Blick (vgl. Abschnitt 1).<sup>11</sup> Zur Verfertigung dieser Praktiken kann auf verschiedene Ressourcen zurückgegriffen werden: Sprachliche Zeichen sind ein wesentlicher Bestandteil dieser, doch auch nicht-sprachliche Ressourcen gehören dazu, wie allgemein das Wiederaufgreifen und Transformieren von bereits eingebrachten Ressourcen (vgl. Abschnitt 3).

Um einen Zugriff auf diese Praktiken zu bekommen, liegt es zunächst nahe, die in der Smartphone-App hinterlegten Aufzeichnungen der IPA selbst zu betrachten. Zu diesem Zweck wurde ein Protokolldatenkorpus erstellt, das mit Stand Januar 2021 aus 244 IPA-Dialogen aus drei verschiedenen Haushalten besteht. Die Haushalte wurden über das private Umfeld der Autor\_innen akquiriert. Diese setzen sich aus zwei bzw. drei Personen zusammen (zwei studentische Wohngemeinschaften und eine Familie bestehend aus einer Mutter mit zwei erwachsenen Söhnen).

Die in der „Alexa“-App gespeicherten Sprachverlauf-Aufnahmen (vgl. Abschnitt 1) wurden von den IPA-Nutzer\_innen durch Bildschirmvideos aufgezeichnet. Die dazu genutzte Funktion *Bildschirmaufnahme* war auf den Smartphones der Beteiligten vorinstalliert. Aus den Videos wurden daraufhin die Audiospuren extrahiert.

Die extrahierten Audiodateien wurden anschließend geschnitten. Die so gewonnenen Aufzeichnungen sind „natürliche Daten“ im Sinne der ethnomethodologischen Konversationsanalyse nach Harold Garfinkel (vgl. Bergmann 2010), die als ein methodologischer Stützpfeiler unserer Analysen dient. Sie sind nicht durch die Erhebungssi-

---

11 Praxeologische Arbeiten wie etwa von Wiedemann (2019), Dang-Anh (2019), Meiler (2018) oder Schubert (2019) zeigen, dass mit solchen (unterschiedlich ausgeformten) Ansätzen eine „Analyse digitaler Kultur“ (Schäfer 2021) unter verschiedenen Gesichtspunkten möglich ist.

tuation beeinflusst, da sie erst mit zeitlichem Abstand als Forschungsdaten erhoben wurden (vgl. Gerwinski/Linz 2018; Salheiser 2019).

Die Dateien wurden anschließend inventarisiert. Das tabellarische Inventar umfasst die vom IPA für die weitere Verarbeitung verschriftete Aufzeichnung sowie die von uns erstellte Umschrift des Audios nach GAT2 (vgl. Selting et al. 2009) als Basistranskript. Ein „Protokolleintrag“ in der App wird dabei immer als ein Tabelleneintrag, d.h. als eine Zeile, erfasst. Es wurden in diese Einträge außerdem die in der App hinterlegten Zusatzinformationen aufgenommen, die für die Auswertung der Aufzeichnungen relevant sein können, darunter Uhrzeit und Dauer der Aufnahmen, das verwendete Gerät sowie die Anzahl der auf der Aufnahme zu hörenden Sprecher\_innen. Dokumentiert wurde auch die produzierte „Antwort“ des IPA, die in der App festgehalten ist, in der Situation aber auditiv wiedergegeben wurde. Dies ermöglicht, das gesprochensprachliche Material auch einer basalen prosodischen Analyse zuzuführen und die Einbettung in soziale Situationen besser verstehbar zu machen. Außerdem war so auch ein Abgleich zwischen der Umschrift durch das IPA-System und dem von den Autor\_innen gehörten Sprachmaterial möglich. Diese doppelte Dokumentation und Transkription ermöglicht darüber hinaus eine Sequenzierung auf Basis der verbalen und schriftlichen Sprachdaten sowie anhand der Zusatzinformationen (z. B. zur Dauer zwischen zwei Eingaben). Ein Tabelleneintrag entspricht dabei idealtypisch einer Sequenz (Nutzer\_innen-Eingabe – IPA-Antwort, siehe Beispiel 1 in Abschnitt 5.2). Sequenzen und darüber hinaus gehende Bestandteile von sequenziellen Abläufen können aber auch über Einträge hinweg verlaufen, umgekehrt können innerhalb eines Eintrags beispielsweise zwei Sequenzen dokumentiert sein (siehe Beispiel 2 in Abschnitt 5.2). Häufig ist in einem Eintrag nur das Aktivierungswort dokumentiert, und in einigen Fällen umfasst die Aufzeichnung Anderes als nur Sprachbefehle, z. B. Anschlusskommunikation. Da das soziale Geschehen zwischen zwei Einträgen weder visuell noch akustisch dokumentiert ist, bleiben diese Sequenzierungen aber unsichere Interpretationen.

Gemäß GAT2 wurden die Sequenzen weiter in Intonationsphrasen gegliedert, sofern diese im akustischen Material erkennbar waren. Die Darstellung der Beispiele im Analyseteil konzentriert sich aus Gründen der Übersichtlichkeit auf diese GAT2-Notation, bezieht aber die schriftlichen Ausgaben des IPA mit ein und wird jeweils um eine Erläuterung weiterer, für die Analyse relevanter Informationen ergänzt.

Potenzielle Auswertungen dieser Sequenzen explorieren wir, vor dem Hintergrund einer linguistisch orientierten Praxeologie, in dreierlei Stoßrichtungen: Erstens als *Daten*praktiken in Infrastrukturen

und Plattformen (Abschnitt 5.1), zweitens mit einem Fokus auf die Einbettung der *sprachlichen* Praktiken in soziale Interaktionen (Abschnitt 5.2) und drittens mit Fokus auf die alltäglichen und diskursiven Kontexte, in die sie eingebunden sind und die sie zugleich mit konstituieren (Abschnitt 5.3).

## 5. Analysen

### 5.1 Protokolldaten als Datenpraktiken in Infrastrukturen und Plattformen

Wenn nun in der Theorie der Praxis von Schüttpelz und Meyer Medien nicht als feste Rahmen für kommunikative Formen gelten, sondern selbst als Teil dieser Formen interaktiv hervorgebracht werden müssen, kann man sich fragen, was dies für die Betrachtung von Protokolldaten und überhaupt für ein entsprechendes sozialtheoretisches und medienlinguistisches Verständnis von *Daten* heißt. Daten sind aus dieser Perspektive nichts rein Gegebenes, worauf die Etymologie des Begriffs schließen ließe (lat.: *dare* = geben), sondern zum einen etwas *Hergestelltes*, zum anderen etwas, das situativ als Datum *relevant* gemacht werden muss, um als solches Verwendung finden zu können.

Die Produktion von Daten wurde sozialwissenschaftlich vor allem in der Wissenschaftssoziologie untersucht (vgl. Pickering 1993) – wegweisend waren Studien Bruno Latours, beispielsweise zur Hormonforschung (Latour/Woolgar 1986). Daten, so Latours These, liegen nicht einfach vor, sondern sind als Teil eines Netzwerks aus Personen, Geräten, Institutionen, Orten und Gelegenheiten aus diesem heraus entstanden und verständlich.<sup>12</sup> In einem anderen Kontext wären sie entweder unverständlich oder würden ihre Bedeutung verändern. In diesem Sinn kann man im Sinne Latours Daten als „immutable mobiles“ bezeichnen, als „unveränderliche und kombinierbare mobile Elemente“ (Latour 2009: 129). Mit Latour ist dabei auch hervorzuheben, dass die Daten selbst in einer konkreten materiellen Form vorliegen und für eine Darstellung entsprechend aufbereitet werden müssen (vgl. Latour/Woolgar 1986: 50).

Die These von der wechselseitigen Verfasstheit von Daten wird anhand des Labors besonders anschaulich, da diese Verfassung hier auf händischer Arbeit beruht. Im Fall der *Protokolldaten* ist sie weit

---

<sup>12</sup> In diesem Text beziehen sich Latour und Woolgar zwar auf die *construction of scientific facts* durch die Erzeugung von „literary inscription[s]“ (1986: 87); das Argument lässt sich aber ebenso auf die diesen Fakten zugrundeliegenden Daten ausweiten.

weniger sichtbar, da die Herstellung weitgehend automatisch und opak erfolgt (vgl. Crawford/Joler 2018). Zugleich zeigt jedoch die Präsentation der Protokolldaten durch eine App (vgl. Abb. 1), dass die Daten auch in diesem Fall in einer gewissen Form aufbereitet und für die Nutzenden als Daten präsentiert werden. In diesem Sinn ist der Begriff der „Datenpraktiken“ zu verstehen, wie er im Forschungsprogramm des SFB 1187 entwickelt wird.<sup>13</sup>

Die Präsentation der Protokolldaten in der App ist für die Rekonstruktion der Nutzungspraktiken grundsätzlich aufschlussreich, da sie sowohl den Nutzenden wie auch den Forschenden einen chronologischen Einblick in die konkreten Interaktionen erlaubt und darüber hinaus Fehlschläge sichtbar macht. Aber auch die Beschränkung in der Nutzung der Daten durch die Verhinderung eines direkten Exports dieser aus der App (vgl. Abschnitt 1) ist insofern relevant, als sie auf die Einbettung der Protokolldaten in die Plattform von „Alexa“ und damit der Firma Amazon verweist. Quellenangaben in von der App dokumentierten Folgezügen zu Befehlen (z. B. „spiele i want it that way von backstreet boys“ – „I Want It That Way von Apple Music“) bzw. in den schriftlich dokumentierten Antworten (z. B. „Laut dem Londoner Natural History Museum ...“) machen ersichtlich, dass der IPA z. T. auch auf Dienste anderer zugreift.

Diese Vermittlung der Angebote Dritter sowie die Rahmung der Protokolldaten für die Nutzenden lassen den IPA sowohl als Teil einer *Infrastruktur* wie auch als Teil einer *Plattform* erscheinen. Da beide Begriffe in der sozial- und kulturwissenschaftlichen Diskussion digitaler Medien relevant sind, sollen sie auch hier kurz in ihrer Relevanz für die IPA geklärt werden. Im Kontext der Germanistischen Linguistik wird das Konzept der Infrastruktur von Matthias Meiler herangezogen und mit dem Begriff der Kommunikationsform (vgl. Abschnitt 2) vermittelt (vgl. Meiler 2019: 73–76, am Beispiel von Weblogs).

Beide Begriffe, Plattform wie Infrastruktur, verweisen etymologisch auf Substrata, welche die Grundlage für andere Aktivitäten bilden und selbst meist unbemerkt bleiben. Insbesondere im Fall der Infrastruktur wird diese Grundlage sehr häufig dann im Alltag thematisiert, wenn sie ihren für selbstverständlich genommenen Dienst versagt – der Strom fällt aus, der Abfluss ist verstopft etc. Aus einer solchen Perspektive können IPA als Infrastruktur gelten, da sie die Grundlage für Informationsabfragen über die Dienste Dritter – Wetterdatenanbieter, Wikipedia, Nachrichtenportale – und für die Steuerung von Smart Home-Geräten darstellen. Aber auch durch die Integration in die Steuerung der häuslichen Grundeigenschaften

---

<sup>13</sup> Siehe <https://www.mediacoop.uni-siegen.de/de/forschungsprogramm/>

wie Wärme und Licht sind IPA mit den klassischen Infrastrukturen verknüpft.

In dieser Perspektive sind IPA als ein weiterer Schritt in der technischen Anbindung des Haushalts an ein Versorgungssystem zu verstehen. IPA basieren auf Infrastrukturen der Strom-, Telefon- und Internetverbindung von Haushalten. Des Weiteren wird für die Einrichtung der Geräte ein Smartphone vorausgesetzt. Der Smart Speaker im Wohnzimmer ist somit Ausläufer eines Netzes an Verbindungen zu auf der Welt verteilten Rechenzentren und den dort ablaufenden Rechenprozessen (vgl. Eggert/Kerpen 2018). IPA sind zum einen auf externe Dienstleister angewiesen und zum anderen auf eine Fülle an vernetzten Geräten, die in Form „smarter“ Glühbirnen und Thermostate von den IPA steuerbar sind.

In der Forschung zu digital-vernetzten Medien ist neben dem Begriff der Infrastruktur vor allem jener der Plattform zentral. Dieser Begriff der Plattform weist nun spezifisch auf die Eigentumsverhältnisse und organisatorische Einbettung der Technologien hin. Die „Plattform-Ökonomie“ (Srnicsek 2016) ist zu einem populären Begriff geworden, um nicht nur ökonomische Strategien von Social Media-Firmen wie Facebook, sondern einen Trend des Outsourcing und der betrieblichen Flexibilisierung auf Basis der Analyse großer Datenmengen zu beschreiben. Wesentlich für Plattformen ist mithin die Erhebung und Verwertung vielfältiger Nutzungs- und Nutzer\_innen-Daten (vgl. Srnicsek 2016: 39–43; Strüver 2020: 3).

Die ersten Medien, die als Plattform beschrieben wurden, und deren Thematisierung daher die *Platform Studies* begründete, waren Spielkonsolen wie jene der Firmen Atari, Sega oder Nintendo (vgl. Bogost/Montfort 2009). Sie standardisieren als Plattform technische Parameter der Computerspiele und können somit als „two-sided markets“ (Rochet/Tirole 2003) beschrieben werden, die Anbietende und Kaufinteressierte in Bezug zueinander setzen. Dabei behalten die Plattformen die Kontrolle über die Regeln des Markt-Zugangs und der Transaktionen. Kirchner und Beyer (2016: 329) bezeichnen eben dies als „Plattformlogik“ (vgl. Staab 2019). Diese zeigt sich insbesondere bei den beiden großen Smartphone-Betriebssystemen „iOS“ und „Android“, die Software nur über die jeweiligen App-Stores auf den Endgeräten installierbar machen. Langlois und Elmer (2019) weisen darauf hin, dass sich vormalig auf *einen* Plattformmarkt beschränkte Unternehmen zunehmend in anderen Bereichen engagieren und sich daher eine Bewegung von einer geschäftsmäßigen Organisation funktionspezifischer Plattformen hin zu Unternehmen beobachten lässt, die umfassende Infrastrukturen bereitstellen oder sich in bestehende einschreiben. Dies lässt sich anhand der drei Unternehmen illustrieren, welche die auf dem west-

lichen Markt erfolgreichen IPA anbieten. Diese haben jeweils ursprünglich eine bestimmte Plattform betrieben: Amazon im Versandhandel, Google die Suchmaschine und Apple die Kombination aus Computer-Hardware und Software. Inzwischen haben sie in verschiedene Bereiche expandiert, bei denen sie teilweise in Konkurrenz zueinander treten. Die jeweiligen IPA sind in die Plattformen der Anbieter integriert, denn die Firmen stellen jeweils intelligente Lautsprecher mit herstellereigenen Betriebssystemen her, die als Vermittlungsinstanz auch bei der Kommunikation mit *Artificial-Intelligence*-Produkten anderer Anbieter zum Tragen kommen. Diese vertikale Integration und die „Plattformlogik“ der durch die IPA konstituierten Märkte wird zugleich von einer Einbettung in eine Infrastruktur von Diensten der Anbieter sowie Dritter komplementiert. In diesem Sinn kann man die Einbettung der IPA in den häuslichen Alltag wohl mit jener Figur beschreiben, die Plantin et al. (2018: 306) auf digitale Medien insgesamt beziehen: als eine Gleichzeitigkeit der „platformization of infrastructures“ und der „infrastructuralization of platforms“.

Wenn man nun die Protokolldaten und ihre Einbettung in die „Alexa“-App im Hinblick auf Amazon als Plattform betrachtet, fallen vor allem die Beschränkungen im Umgang mit den Daten auf. Zugleich fällt es aus dieser Perspektive leicht, die Nutzenden lediglich in Abhängigkeit von diesen Geräten und ihren Einbettungen in Wertschöpfungsketten, Interfaces und Voreinstellungen zu begreifen. So betreiben Plattformen laut Ulrich Dolata (2019: 195) über die konkrete technische Ausgestaltung eine „Kuratierung sozialer Verhältnisse und sozialen Verhaltens“ – beispielsweise durch die mehr oder weniger große Möglichkeit, Beiträge oder Inhalte privat zu schalten. In der Wissenschafts- und Techniksoziologie finden sich jedoch primär Perspektiven, welche den scheinbar starren Begriff der Infrastruktur in vielerlei Hinsicht aufweichen. Zum einen wird auf die historische Genese, Pfadabhängigkeit und Wandelbarkeit von Infrastrukturen verwiesen, zum anderen wird die konstante Arbeit an der Aufrechterhaltung von Infrastrukturen betont. In Susan Leigh Stars Forschung (Star/Bowker 2006) wird aus der Infrastruktur daher auch ein Verb – *to infrastructure* –, das den Prozess der Infrastrukturerung beschreibbar macht. Infrastrukturen erscheinen aus dieser Perspektive als etwas, das der Pflege und Reparatur, Integration und Adaption bedarf. Des Weiteren weist die Forschung darauf hin, dass auch auf der Seite der Nutzenden, die nicht in die professionelle Planung, Konstruktion und Pflege der Infrastruktur eingebunden sind, sich gleichwohl Umgangsweisen mit Infrastrukturen herausbilden, die nicht immer den Intentionen der Herstellenden und Betreibenden der Infrastruktur entsprechen müssen. Beim IPA,

wie schon beim internetfähigen Computer um die Jahrtausendwende (vgl. Röser et al. 2019), müssen die Mitglieder von Privathaushalten z. B. entscheiden, in welchem Raum sie das jeweilige Gerät aufstellen und wer dieses wann und wie nutzen darf.

Die von Schüttpelz (2016: 5) auch auf „infrastrukturelle Medien“ gemünzte Formulierung von den „kooperativ erarbeitete[n] Kooperationsbedingungen“ lässt so nicht nur die Fabriziertheit der Infrastruktur deutlich hervortreten, sondern auch die daran anschließende Kooperation auf Basis der Infrastruktur als Bedingung und Ermöglichung von Kooperation und damit Interaktion. Des Weiteren kann man mit dem Begriff der Infrastruktur technische Gerätschaften und Kommunikationsmedien einerseits, die Körper der Akteur\_innen (Star/Bowker 2006: 231) und soziale Aspekte wie die Gepflogenheiten der Interaktion (Schegloff 2012) andererseits in ein Kontinuum einreihen, um so ihre Relevanz für die Interaktion zu betonen. Die Frage wäre nun, wie sich der praktische Umgang mit den IPA und dessen Erforschung im Hinblick auf die Einbettung in Plattformen und Infrastrukturen darstellen.

## 5.2 IPA-Dialoge und sprachliche Interaktion

Die Protokolldaten geben Einblicke sowohl in Charakteristika soziotechnischer IPA-Dialoge als auch – in deutlich eingeschränktem Maß – in soziale Interaktion unter anwesenden Nutzer\_innen in Situationen, in denen diese gemeinsam mit Sprachassistenzsystemen umgehen. IPA-Dialoge können mit Krummheuer (2010: 323–324) grundsätzlich als ein „hybrider“ bzw. „ambiger“ Austausch beschrieben werden (s.u.). In Verbindung mit diesem soziotechnischen Austausch stoßen wir in den Protokolldaten auch auf bruchstückhafte Dokumentationen sozialer Interaktion, die auf eine IPA-Nutzung in Zwei- oder Mehrparteienkonstellationen zurückzuführen ist (vgl. Porcheron et al. 2018; vgl. auch – im Blick auf die Interaktion mit einem Museumsroboter – Pitsch et al. 2017).

Bei der Konzeptualisierung der sozialen Interaktion im Verhältnis zum hybriden Austausch mit dem Gerät stützen wir uns – wie auch Krummheuer – auf die Arbeiten Goffmans, der soziale Interaktion als „wechselseitige Wahrnehmung und Kommunikation von zwei körperlich anwesenden Personen“ versteht, „die wahrnehmen, dass sie wahrgenommen werden, und einen gemeinsamen Aufmerksamkeitsfokus teilen“ (Krummheuer 2010: 13). Auch ohne die Ressource des Körpers, über die der Embodied Conversational Agent bei Krummheuer verfügt, weist der „hybride“ bzw. „ambige“ Austausch auch im Fall von IPA mit VUI aufgrund einer eben diese Illusion nähernden Gestaltung streckenweise Ähnlichkeiten zur sozialen Interaktion auf, bringt aber immer wieder auch Irritationen mit sich, die

dann den sozialen Charakter des Gegenübers in Frage stellen und den Simulationscharakter der (vorherigen) Dialoge deutlich werden lassen (Krummheuer 2010: 323–324). In Begriffen der Philosophie des Geistes (vgl. Lotze 2016: 68–75; Drösser 2020: 20–28) könnte man auch sagen: Phasenweise können die beteiligten Menschen aufgrund der quasi-geistigen Leistungen ihres technischen Gegenübers kognitiv gar nicht anders, als diesem Gegenüber ein geistiges Innenleben (Intentionalität) zu unterstellen. Zu einer entsprechenden Perception und einem hierauf beruhenden Verhalten trägt zudem ein anthropomorphes Design, etwa der Stimmqualität im Fall von „Alexa“, bei (vgl. Lotze 2016: 62–64). Dann wieder wird anhand inkohärenter, holpriger und unflexibler Dialog-Abläufe deutlich, dass eine angemessene semantische Verarbeitung der situierten sprachlichen Äußerungen nicht gelingt, die Systeme also selbst kein Bewusstsein ihrer eigenen Intentionalität und der Intentionalität ihres Gegenübers haben – oder anders gesagt: dass sie dieses Bewusstsein durch dialogische Strukturen und anthropomorphes Design bloß mehr oder weniger geschickt *simulieren* (vgl. Lotze 2016: 233–253). Echte soziale Interaktion ist in weit höherem Maße indexikalisch (vgl. Garfinkel 1967; vgl. Abschnitt 3) und zeichnet sich durch ein komplexes Geflecht von kontextgebundenen Erwartungen und Folgeerwartungen aus, die im hybriden Austausch mit Assistenzsystemen nicht in gleichem Umfang produziert und eingelöst werden können.<sup>14</sup> Hier zeigen sich, trotz einer verbreiteten Rhetorik des „noch nicht“, „heute schon“, „immer öfter“, „mehr und mehr“ etc. (vgl. Drösser 2020: 6–14) und im Ganzen nicht einsehbarer, im Einzelnen jedoch tiefgreifender technischer Neuerungen (so genannte „neuronale Netze“, *Machine Learning*, *Big Data* etc., vgl. Crawford/Joler 2018), prinzipielle Grenzen (vgl. Drösser 2020: 11; 55–59). Während mithin in den IPA-Dialogen trotz technischer Fortschritte die Anpassung der Nutzenden an die Dialogfähigkeit der Systeme in den Vordergrund tritt, gerade wenn man deren spezifische Funktionalitäten nutzen möchte (vgl. Drösser 2020: 72), zeigen sich bei der Einbettung der IPA in die soziale Interaktion deutlicher die kreativen Gestaltungsmöglichkeiten der Nutzenden im Umgang mit der Technik.

Das Verhältnis zwischen den IPA-Dialogen und sozialer Interaktion soll nachfolgend anhand zweier Beispiele aus dem oben beschriebenen Protokollkorporus erläutert werden. In Beispiel (1) („Pantheon“) zeigt sich die für unser Korpus typisch erscheinende Form eines zweizügigen IPA-Dialogs: Auf den Sprachbefehl des

---

<sup>14</sup> Dies gilt umso mehr, wenn Assistenzsysteme in arbeitsteilige soziotechnische Netzwerke in institutionellen Kontexten eingebettet sind (vgl. Amrhein/Cyra/Pitsch 2016).

Nutzers M1, der explizit und direkt durch die verbale Imperativform „definiere“ in Verberststellung und das Akkusativobjekt „PANtheon“ (Z. 001) realisiert wird, folgt eine entsprechende Reaktion des Sprachassistenten:

**Beispiel (1): Pantheon**

```
001 M1 definiere PANtheon.
```

Die ursprünglich auditiv produzierte Reaktion des Sprachassistenten ist in der App ausschließlich schriftlich<sup>15</sup> dokumentiert (vgl. Abschnitt 1):

```
Das Wort „Pantheon“ hat folgende Bedeutungen: 1) ein antiker, allen Göttern geweihter Tempel und 2) die Gesamtheit der Götter einer (polytheistischen) Religion.
```

In dem in der „Alexa“-App einsehbaren Sprachaufnahmen-Verlauf wird im Anschluss an die Reaktion des Sprachassistenten in Form einer Erläuterung der Bedeutung des erfragten Wortes „Pantheon“ kein weiterer Eintrag (etwa eine Ratifizierung durch den Nutzer M1) angezeigt. Lediglich wenige Ausnahmen in unserem Korpus weisen mehr als diese beiden Dialogzüge auf, was insofern – jedenfalls im Blick auf die Form der Dokumentation in den Protokolldaten – die Annahmen von Krummheuer (2010: 320) zu bestätigen scheint, dass die „Verknüpfung der einzelnen Redebeiträge [...] im hybriden Austausch“ oft (noch) „nicht über drei, sondern über zwei Redezüge organisiert [ist]“ und „eine richtige Interpretation des ersten Redezugs unterstellt [wird]“. Allerdings scheint es inzwischen auch Abweichungen von diesem Muster zu geben: Im Zuge einer gewissen Annäherung an soziale Interaktion sollen auch dritte Dialog-Züge seitens der Nutzer\_innen (z. B. Reformulierungen von Fragen oder Befehlen) durch die Systeme ausgewertet werden, um die sequenzielle Angemessenheit der technischen Reaktion zu kontrollieren, und es sollen im Interesse einer Erhöhung des Nutzungskomforts – Drösser (2020: 78) spricht, durchaus ironisch, von einer „Rundum-sorglos-Betreuung“<sup>16</sup> – künftig auch komplexere interaktive Abläufe proj-

---

<sup>15</sup> Die Antwort wird für die Endnutzer\_innen nur auditiv ‚vorgetragen‘; die Produktion erfolgt jedoch als Teil des Natural Language Processings schriftlich und wird anschließend durch Speech Processing auditiv verbalisiert (siehe dazu Natale 2020).

<sup>16</sup> Möglicherweise nähern sich im Zuge dieser Entwicklung die IPA mit VUI im Blick auf den Grad ihrer Selbstaktivierung dem Museumsroboter an, wie er von Pitsch et al. (2017) untersucht wurde.

ziert und technisch unterstützt werden (Strüver 2020: 7–9). Während Beispiel (1) eine einfache Form des IPA-Dialogs bzw. von Mensch-Maschine-Interaktion illustriert, zeigen sich an dem nachfolgend präsentierten Beispiel (2) („Elefantentrompete“) zwei Phänomene, die eine genauere Untersuchung wert sind. In diesem Beispiel versuchen die beteiligten Sprecherinnen W1 und W2 den *Smart Speaker* zur Imitation von Tierstimmen zu bewegen.

**Beispiel (2): „Elefantentrompete“**

001 W1 wie macht die KATze,

002 W1 wie macht die KATze,

003 W2 wie macht der eleFA:NT,

Schriftlich ist als Antwort auf die von W1 und W2 an den Smart Speaker gerichteten und in der Smartphone-App auditiv abrufbaren Fragen folgende (ursprünglich vom Gerät auditiv wiedergegebene, in der App aber nur schriftlich erfasste) Ausgabe des IPA dokumentiert:

Der Elefant macht - Sie haben Elefantentrompete verwendet. Amazon empfiehlt Ihnen daher einen weiteren beliebten Bildung und Nachschlagewerke-Skill, Animal Sounds. Würden Sie ihn gern ausprobieren?

Die darauf folgenden mündlichen Äußerungen von W1 und W2 sind in der App wiederum auditiv abrufbar:

004 W1 nein.

005 W2 NE::IN,=

006 W1 =sehr UNgern würd ich dat AUsprob-

Beispiel (2) illustriert mehrere Phänomene: Erstens gibt es Abschluss über mögliche Störungen und Reparaturen in IPA-Dialogen. Zweitens zeigt sich, dass auch der IPA bzw. das integrierte Natural Language Processing zweite Züge derart erweitern können, dass Folgezüge relevant gesetzt werden: So produziert der IPA nicht nur den Austausch abschließende Turns, sondern auch sequenz-eröffnende Turn-Expansionen, die eine Fortsetzung erwartbar machen (vgl. zur Selbstaktivierung technischer Artefakte auch Pitsch et al. 2017 im Bezug auf Museumsroboter sowie allgemein zum Dialogdesign als Mittel zur Simulation eines intelligenten sozialen Gegenübers bei Chatbots Lotze 2016).

Wenden wir uns zunächst den Reparaturen zu: In den Zeilen 001 und 002 fällt auf, dass Sprecherin W1 zweimal hintereinander die gleiche Frage an das Gerät richtet („wie macht die KATze,“, Z. 001). Ob W1 zuvor das Wake Word „Alexa“ äußert und wie lange die Pausen zwischen den Äußerungen von W1 und W2 (Z. 001–003) sind, ist durch die Darstellung in der App nicht nachvollziehbar. Das wiederholte Einfordern einer Antwort durch W1 und W2 ist ein Indiz dafür, dass die von W1 relevant gesetzte Antwort des IPA nicht geliefert wurde. Auch die über die App abrufbaren Protokolldaten zeigen an, dass keine Aktivität des Geräts erfasst wurde. Nachdem auch der zweite Versuch von W1 scheitert, den Sprachassistenten zu einer Reaktion zu bringen, wiederholt Sprecherin W2 die Frage, ersetzt bei gleichbleibender syntaktischer Struktur das Lexem „KATze“ jedoch durch das Lexem „eleFA:NT“ (Z. 003). An dieser Stelle wird deutlich, wie Sprecherin W2 im Rahmen des Testens der sogenannten Skills des Smart Speakers mit der von Sprecherin W1 relevant gesetzten, aber vom Gerät nicht realisierten Reaktion umgeht und die von W1 an das Gerät gestellte Frage entsprechend anpasst (vgl. Schegloff 2012: 252–254). Solche von den beteiligten menschlichen Gesprächspartnern realisierten Lösungen können zudem Aufschluss darüber geben, was erstere als mögliche Ursache(n) für das Ausbleiben des relevant gesetzten zweiten Turns – hier des Adjazenzpaars Frage-Antwort – deuten: in diesem Fall, dass der IPA das zum Lexem „KATze“ erwartete Tiergeräusch nicht produzieren kann, aber andere Tiergeräusche gegebenenfalls verfügbar sind (vgl. Porcheron et al. 2018).

Betrachten wir mit Fokus auf die inkrementelle Erweiterung des IPA-Dialogs nun den zweiten Teil der Äußerung des Sprachassistenten. Nach dem Abspielen des Elefantengeräuschs (in dem in der App hinterlegten Transkript durch einen Gedankenstrich visualisiert) wird der – hier im einfachen Fall hybrider Dialoge abgeschlossene – Zug seitens des IPA um eine Empfehlung expandiert: An die beantwortete Frage anknüpfend wird eine Frage verbalisiert, die eine Antwort seitens des/der am Dialog beteiligten Menschen relevant setzt. Die vom Sprachassistenten in Bezug auf den Skill „Animal Sounds“ aufgeworfenen Frage „Würden Sie ihn gern ausprobieren?“ verneint W1 (Z. 004) und in unmittelbarem Anschluss auch ihre Gesprächspartnerin W2, letztere mit einem paraverbal verstärkten „NE::IN,“ (Z. 005). Die von W1 daraufhin geäußerte und die Sequenz abschließende Expansion „sehr UNgern würd ich dat AUsprob-“ (Z. 006) zeigt, wie W1 auf der Ausgabe des Sprachassistenten operiert, indem sie das in der Ausgabe enthaltene sprachliche Material in ihrer Äußerung aufgreift, durch syntaktische Variation transformiert (vgl. Goodwin 2018: 431) und dialektal durchfärbt. Ob W1 ihre Äußerung nicht vollständig zu Ende führt oder ob die Aufnahme des IPA

frühzeitig abbricht – z. B. nach dem erkannten Sprachbefehl „nein.“ (Z. 004) – lässt sich auf Basis des Ausschnitts nicht nachvollziehen; wahrscheinlich ist jedoch, dass der IPA nicht auf die Rezeption einer zweiten Sprecherin ausgerichtet ist und bereits nach dem Erkennen der ersten, turnabschließenden Äußerung („nein.“, Z. 004) nicht mehr im „Aufmerksamkeitsmodus“ ist, sodass die Aufnahme abbricht, bevor W1 ihre inkrementelle Erweiterung (Z. 006) vollendet hat (siehe dazu auch Pitsch et al. 2017: 396).

Die Äußerung von W1 (Z. 006) ist doppelt funktionalisiert: Einerseits bekräftigt sie den Abschluss des IPA-Dialogs, gleichzeitig nutzt die Sprecherin die Äußerung, um in der sozialen Interaktion mit der ko-präsenten Sprecherin W2 die Äußerungen des IPA als Ressource für die weitere Interaktion nutzbar zu machen: Ihre spöttische – ironisch höflich formulierte – Zurückweisung des Vorschlags verweist indexikalisch auf eine durch Geringschätzung des IPA-Skills geprägte Haltung und den Kontext des gemeinsamen unterhaltsamen Erprobens mehr oder weniger skurriler und überflüssiger technischer Funktionen (vgl. Abschnitt 5.3).<sup>17</sup>

Ein weiteres, bereits für Gruppeninteraktionen, in die ein technisches Artefakt eingebettet ist, beobachtetes Phänomen (vgl. Reeves/Porcheron/Fischer 2019; Habscheid et al. 2020) scheint sich in diesem Beispiel zu manifestieren: Die scheinbar von W1 an das Gerät gerichtete Äußerung adressiert (primär) ihre ko-präsente Gesprächspartnerin W2 – diesen Schluss lassen die dialektale Durchfärbung und die für den IPA-Dialog irrelevante Ergänzung um die Bewertung „sehr UNgern“ (Z. 006) zu. Die Äußerung erfüllt somit auch und sogar primär Zwecke für die soziale Situation, denn „what is said *to* the device is necessarily often said *around* others“ (Reeves/Porcheron/Fischer 2019: 49). Dabei nutzt W1 Bestandteile des IPA-Dialogs und bindet diesen so in die soziale Interaktion ein; es entsteht eine Verflechtung von hybridem Austausch *mit* dem Gerät und sozialer Interaktion mit W2 *vor* dem Gerät.

Da die von Amazon erhobenen Mitschnitte von Sprachein- und Sprachausgaben als verschriftete isolierte Einzelelemente ohne die Situation, in der sie realisiert werden, gespeichert werden, lässt sich über die genauere Einbettung sowie über die prosodische Realisierung der Äußerungen des Sprachassistenten im vorliegenden Beispiel (2) nur spekulieren. Da diese Informationen aber erforderlich für eine verlässliche Rekonstruktion der sprachlichen Interaktion sind, lassen sich über die mit dem IPA-Dialog verbundene soziale

---

<sup>17</sup> Wie das Austesten technischer Funktionen des IPA eingeordnet werden kann, – etwa als neue Form von „Computer Talk“ nach Zoeppritz (1985) – bleibt im Rahmen weiterer Untersuchungen zu prüfen (vgl. zu einer neuen Online-Variante von „Computer Talk“ Lotze 2016; 2018).

Interaktion *vor* und *mit* dem Gerät anhand der Protokolldaten nur in recht geringem Umfang Rückschlüsse ziehen. Multimodale Interaktionsanalysen, die den Kontext, die sozialräumliche Umgebung und insofern auch die materiale Beschaffenheit der Interaktionssituation mit dokumentieren, sind für eine umfassende gesprächsanalytische Betrachtung zwischenmenschlicher Interaktion im Umgang mit dem IPA unerlässlich.

Dagegen kann für die Beschäftigung mit IPA-Dialogen die Analyse der Protokolldaten aufschlussreicher sein. IPA-Dialoge sind deutlich von sozialer, zwischenmenschlicher Interaktion zu unterscheiden, in der die sequenzielle Organisation, das Turn-Taking und die Gleichzeitigkeit von Produktion und Rezeption sowie die gegenseitige Wahrnehmungswahrnehmung (vgl. Goffman 1983: 2) generische Merkmale der Interaktion und entscheidend für die Lösung von kommunikativen Problemen sind (vgl. Schegloff 2012: 246; Auer 2000). IPA-Dialoge sind in diesen Hinsichten grundlegend anders strukturiert. So lässt sich etwa im zweizügigen Fall nur mit großen Einschränkungen überhaupt von einer sequenziellen Organisation sprechen. Allerdings scheinen sich die Merkmale – in Grenzen – anzunähern, wenn Antworten des IPA durch Expansionen Folgeäußerungen erwartbar machen, die in die Position einer Verständnissicherung eintreten.

### 5.3 Produkt- und nutzungsseitige sprachliche Herstellung von Alltagskontexten

Nicht zuletzt geben die Protokolldaten auch einen Aufschluss darüber, in welche alltagspraktischen – kommunikativen und über Kommunikation hinausreichenden – bzw. diskursiven Kontexte die IPA systemseitig gestellt sind bzw. in der nutzungsseitigen sozialen Interaktion gestellt werden (vgl. Habscheid 2016). So zeigt etwa Beispiel 1 (vgl. Abschnitt 5.2) eine Wissensrecherche, die an der Bedeutung (alltagssprachlich: „Definition“) eines sprachlichen Ausdrucks (*Pantheon*) festgemacht wird.

In Beispiel 2, in dem neben dem IPA-Dialog mehr oder weniger zufällig auch ein Auszug aus der sozialen Interaktion „vor“ dem Gerät in den Protokolldaten dokumentiert ist, wird die Praktik erkennbar, in einer geselligen Aktivität das z. T. skurrile Funktions- und Leistungsspektrum des IPA zu erkunden. Zugleich zeigt sich bruchstückhaft, wie sich im Fall von Störungen in den IPA-Dialogen in Mehrparteienkonstellationen ein „Meta-Interaktionsraum“ vor dem Gerät öffnet (vgl. Pitsch et al. 2017; Hector/Hrncal 2020), in dem ein Diskurs von Nutzer\_innen *über* das Gerät, hier eine mokante Bewertung, seinen Platz findet, mit dem die Nutzenden ihre Erfahrungen verarbeiten (vgl. Krummheuer 2010, 263–311).

Protokolldaten wie Beispiel 1 dokumentieren Nutzungssituationen, die systemseitig angelegt sind. Sie sind damit sowohl Ausdruck der Nutzung als auch ihrer „Kuratierung“ durch den IPA (vgl. Dolata 2019: 95; vgl. Abschnitt 5.1). So lässt sich anhand der Protokolldaten erschließen, dass routinierte Nutzungsweisen des *Smart Speakers* über weite Strecken das erwartbare Repertoire betreffen, u. a. das Abrufen von Musiktiteln, Playlists und Filmen, Erzählwitzen, Tierlauten und Naturgeräuschen; Wissensrecherchen und Wetterabfragen; Rechenaufgaben; Befehle zu Kommunikationsfunktionen („spIEle meine ↑↑NACHrichten ab;“; „was steht in meinem kaLENDER?“) und *Smart Home*-Geräten („te VAU an;“). Dass in der Nutzung von IPA mit VUI oft einfache Alltagsroutinen vorherrschen, wurde vor allem durch quantitative empirische Studien belegt (vgl. Abschnitt 2).

Derzeit noch seltener werden nach unseren Protokolldaten Dialogsequenzen in routinierten Alltagssituationen auch von Seiten des IPA initiiert:

**Beispiel (3): „Benachrichtigung“**

```
Übrigens, du hast eine neue Benachrichtigung. Soll  
ich deine Benachrichtigung vorlesen?
```

```
001 M1 äh JA:,
```

Wie auch der Diskontinuitätsmarker (*übrigens*) zu Beginn der Äußerung markiert, handelt es sich hier um eine Selbstaktivierung des Systems in Form eines thematisch neuen Vorschlags (vgl. Pitsch et al. 2017; vgl. Abschnitt 5.2). Dadurch kann – etwa auch in längeren Abläufen – die Servicequalität der IPA erhöht und anhand der Reaktionen überprüft werden (vgl. Drösser 2010: 72). Zugleich sind durch ein strategisches Dialogdesign im Rückgriff auf konventionalisierte Adjazenzpaare (vgl. Lotze 2016: 100–103) Abläufe in höherem Maße systemseitig zu lenken und zu kontrollieren.

Im Gegensatz zu Standardsituationen zeigt Beispiel 2 die kreativere Herstellung einer Nutzungspraxis durch zwei Nutzer\_innen, wobei die in der App angelegten Möglichkeiten aufgegriffen und in den alltäglichen Diskurs eingebettet werden. Vielfältigere Nutzungsweisen werden besonders erkennbar in qualitativen Nutzungsstudien, die nach der „Domestizierung“ der technischen Systeme durch die Nutzenden fragen (vgl. Brause/Blank 2020; Abschnitt 2).

Dass spielerische Nutzungsweisen ihrerseits produktseitig antizipiert sind, zeigt die Kategorisierung eines ähnlichen Skills (in „Neuheiten und Humor“) in Beispiel 4:

**Beispiel (4): „Katzensimulator“**

Sie haben Katzensimulator verwendet. Amazon empfiehlt Ihnen daher einen weiteren beliebten Neuheiten und Humor-Skill, Ein Kompliment. Test gefällig?

Auch hier wird eine vorherige Sequenz durch den IPA um eine sequenzinitiiierende Expansion in Form einer standardisierten Empfehlung erweitert. Neben der Erhöhung der Service-Qualität durch zusätzliche Angebote ermöglicht auch eine derartige Interaktionsstruktur (vgl. Abschnitt 4.1) prinzipiell die Auswertung der initiierten Folgezüge unter dem Aspekt der Verständnissicherung und damit auch eine Verbesserung der Dialog-Qualität.

Auf spielerische Nutzungskontexte verweisen schließlich beziehungsensitive Äußerungen (Beleidigungen, Aufforderungen zu Komplimenten etc.), mit denen der humanoide Charakter des Geräts testend herausgefordert wird (vgl. Krummheuer 2010: 263), wie in Beispiel 5:

**Beispiel (5): „Liebe“**

001 M1 LIEBST du mich?

Hier trägt die paradoxe Antwort des IPA, die durchaus zum Weiterdenken anregen kann, zur Simulation eines ‚geistreichen‘ Gegenübers bei (vgl. Abschnitt 5.2):

Ich bin gerne bei dir, aber zu echter Liebe bin ich leider gar nicht in der Lage.

Zu den systemseitig angelegten, subsidiären Nutzungsmöglichkeiten gehört, wie – neben entsprechenden Nutzer\_innen-Äußerungen („schalte das Mikrofon aus.“) – auch die Protokolldaten in der App als solche zeigen, nicht zuletzt auch eine gewisse Regulierung des Datenschutzes durch die Funktion „Aufnahme löschen“, die damit in den Verantwortungsbereich der Nutzer\_innen gestellt wird. In einem von Niklas Strüver (2020: 1) beschriebenen Amazon-Werbevideo, in dem humorvoll-selbstironisch ein Bezug des IPA zu historischen Vorläufern hergestellt wird, tritt dementsprechend neben Dienstmädchen, Hofnarr, Zeitungsjunge und anderen auch eine Sekretärin auf, die Tonbandaufzeichnungen löschen soll.

## 6. Fazit und Ausblick

Als Teil der App geben die Protokolldaten und ihre Inszenierung durch die App einen Eindruck davon, wie sie in die Formierung eines bestimmten Gebrauchs durch die Plattform und Infrastruktur eingebunden sind. Unsere Erhebung der Protokolldaten über Bildschirmvideos zeigt darüber hinaus die im Programm angelegte Interaktion der Nutzenden mit den Protokolldaten. Während der Zweck der Darstellung unter „Aktivität“ erst einmal unklar bleibt, wird die Darstellung der Aufzeichnungen im anderen Fall deutlich als Möglichkeit gerahmt, als Nutzer\_in Einfluss auf die von Amazon gespeicherten Mitschnitte zu nehmen. Die angebotene Möglichkeit einer Löschung kann dabei als Reaktion von Amazon auf die öffentliche Debatte bezüglich der Verwendung von personenbezogenen Daten durch Internetfirmen verstanden werden. Aktuell ist die Löschung aller aufgezeichneten Befehle möglich, nicht nur in der App, sondern auch per Sprachbefehl (vgl. Herbig 2020). Wie bereits dargestellt, ist in der App die Löschung der Daten, nicht aber ihr Export möglich. Die Daten können hier also lediglich der Kontrolle durch Amazon entzogen werden, sollen jedoch entsprechend dem Interface-Design nicht ohne weitere Umstände für andere Zwecke verwendet werden.

Die Daten geben einen Einblick in den tatsächlichen Vollzug der Beobachtung der Nutzenden durch die Infrastruktur. Da hier nicht nur die erfolgreichen Sprachbefehle, sondern auch die Missverständnisse protokolliert werden, werden auch jene gescheiterten Anläufe des IPA studierbar, die vielleicht von den Nutzenden nicht bemerkt worden wären. Zugleich lässt sich nachvollziehen, wann und wie der IPA genutzt wurde. Es entsteht somit ein Protokoll der Nutzung, das um einiges genauer ist als indirekte Erhebungsmethoden (vgl. Tietze/Roßbach 1991): Wie erwartet, erlauben die Protokolldaten gewisse Aufschlüsse darüber, wie derartige Systeme in Verbindung mit Plattformen und Infrastrukturen operieren, wie die Dialogsysteme heute gestaltet sind und welche Nutzungskontexte sie nahelegen.

Da im Regelfall nur der unmittelbare Sprachbefehl nach dem Wake Word erfasst wird, werden die Einbettung in soziale Interaktion und deren praktische bzw. diskursive Kontexte dagegen nur im Ausnahmefall und unvollständig erfasst. Diese Unvollständigkeit betrifft sowohl die simultan ablaufende Interaktion in anderen als den dokumentierten Modalitäten als auch die Einbettung in größere sequenzielle Abläufe. Die Protokolldaten können u. U. begrenzt Aufschluss über soziale Interaktion und Alltagspraxis unter Einbindung

der IPA geben, bedürfen aber in dieser Hinsicht unbedingt der Ergänzung durch Beobachtungs- bzw. audiovisuelle Aufzeichnungsdaten.

Dagegen zeigt unsere Erhebung der Protokolldaten über Bildschirmvideos die im Programm angelegte Interaktion der Nutzenden mit den Protokolldaten selbst. Die App und damit Amazon scheinen jedenfalls daran interessiert zu sein, durch die Möglichkeit der Datenlöschung einen gewissen Datenschutz, und zwar als Aufgabe der Nutzer\_innen, zu offerieren.

Dabei bleibt es jedoch eine empirische Frage, die auf der Basis der Protokolldaten allein nicht zu klären ist, inwiefern diese Möglichkeit der Datenlöschung von den Nutzenden tatsächlich verwendet wird und wie diese eingeschätzt wird. In einer Interviewstudie im Rahmen unseres Projekts verfolgen wir diese Fragestellungen weiter. Bereits vorliegende Studien zum Gebrauch von IPA aus den Sozialwissenschaften und der Forschung zur Human-Computer-Interaction (vgl. Malkin et al. 2019) deuten darauf hin, dass dieses Wissen insgesamt nicht sonderlich verbreitet ist. Selbst dann, wenn die Möglichkeit der Dateneinsicht und -löschung bekannt ist, wird sie anscheinend selten genutzt. Vor diesem Hintergrund kann unsere Erhebung auch als Krisenexperiment (Garfinkel 1967: 37) verstanden werden, da sie die Nutzenden in vielen Fällen zum ersten Mal mit der Möglichkeit konfrontiert, ihren Datengebrauch zu kontrollieren.

## Literatur

- Ammari, Tawfiq/Kaye, Jofish/Tsai, Janice Y./Bentley, Frank (2019): Music, Search, and IoT: How People (Really) Use Voice Assistants. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* (26), 17:1–17:28.
- Amrhein, Antje/Cyra, Katharina/Pitsch, Karola (2016): Processes of Reminding and Requesting in Supporting People with Special Needs: Human Practices as Basis for Modeling a Virtual Assistant? In: *EDIA 2016. Proceedings of the 1st Workshop on Ethics in the Design of Intelligent Agents. In conjunction with the 22th European Conference on Artificial Intelligence – ECAI 2016. The Hague, Holland, August 30, 2016*, 14–19. URL: <http://ceur-ws.org/Vol-1668/paper3.pdf>.
- Apthorpe, Noah/Reisman, Dillon/Sundaresan, Srikanth/Narayanan, Arvind/Feamster, Nick (2017): *Spying on the Smart Home. Privacy Attacks and Defenses on Encrypted IoT Traffic*. URL: <http://arxiv.org/pdf/1708.05044v1>.

- Auer, Peter (2000): On-line-Syntax – Oder: was es bedeuten könnte, die Zeitlichkeit der mündlichen Sprache ernst zu nehmen. In: *Sprache und Literatur* (31), 43–56.
- Ayaß, Ruth/Meyer, Christian (Hg.) (2012): *Sozialität in Slow Motion. Theoretische und empirische Perspektiven. Festschrift für Jörg Bergmann*. Wiesbaden: Springer VS.
- Bentley, Frank/Luvogt, Chris/Silverman, Max/Wirasinghe, Rushani/White, Brooke/Lottridge, Danielle (2018): Understanding the Long-Term Use of Smart Speaker Assistants. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4 (3), 1–24.
- Bergmann, Jörg (2010): Harold Garfinkel und Harvey Sacks. In: Flick, Uwe/Kardorff, Ernst/Steinke, Ines (Hg.): *Qualitative Forschung. Ein Handbuch*. Reinbek: Rowohlt (Rowohlts Enzyklopädie 3990), 51–62.
- Bogost, Ian/Montfort, Nick (2009): Platform Studies: Frequently Questioned Answers. In: *UC Irvine: Proceedings of the Digital Arts and Culture Conference 2009, After Media: Embodiment and Context*. URL: <https://escholarship.org/uc/item/01r0k9br>.
- Both, Göde (2014): Multidimensional Gendering Processes at the Human-Computer-Interface: The Case of Siri. In: Marsden, Nicola/Kempf, Ute (Hg.): *Gender-UseIT. HCI, Usability und UX unter Gendergesichtspunkten*. Berlin/München/Boston: de Gruyter, 107–112.
- Brause, Saba/Blank, Grant (2020): Externalized domestication. Smart speaker assistants, networks and domestication theory. In: *Information, Communication & Society* 23 (5), 751–763.
- Brock, Alexander/Schildhauer, Peter (Hg.) (2017): *Communication Forms and Communicative Practices. New Perspectives on Communication Forms, Affordances and What Users Make of Them*. Frankfurt a. M.: Peter Lang (Language and Text Studies, 15).
- Candello, Heloisa/Pinhanez, Claudio (2018): Recovering from Dialogue Failures Using Multiple Agents in Wealth Management Advice. In: Moore, Robert/Szymanski, Margaret/Arar, Raphael/Ren, Guang-Jie (Hg.): *Studies in Conversational UX Design*. Cham: Springer VS, 139–157.
- Crawford, Kate/Joler, Vladan (2018): Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources (research report). In: *AI Now Institute and Share Lab*. URL: <https://anatomyof.ai>.
- Dang-Anh, Mark (2019): *Protest twittern. Eine medienlinguistische Untersuchung von Straßenprotesten*. Bielefeld: Transcript.
- Datenethikkommission der Bundesregierung (2019): *Gutachten der Datenethikkommission*. Berlin. URL: <https://www.bmi.bund.de/>

[SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf](#).

- Dolata, Ulrich (2019): Plattform-Regulierung. Koordination von Märkten und Kuratierung von Sozialität im Internet. In: *Berliner Journal für Soziologie* (29), 179–206.
- Drösser, Christoph (2020): *Wenn die Dinge mit uns reden. Von Sprachassistenten, dichtenden Computern und Social Bots*. Berlin: Dudenverlag.
- Durkheim, Emile (1893/1992): *Über soziale Arbeitsteilung. Studie über die Organisation höherer Gesellschaften*. Frankfurt a. M.: Suhrkamp (suhrkamp taschenbuch wissenschaft, 1005).
- Eggert, Michael/Kerpen, Daniel (2018): Wer Datengesellschaft sagt, muss auch Cloud-Computing sagen. Die Cloud als zentrale Infrastruktur der datafizierte Gesellschaft. In: Houben, Daniel/Prietl, Bianca (Hg.): *Datengesellschaft. Einsichten in die Datafizierung des Sozialen*. Bielefeld: Transcript, 155–177.
- Ford, Marcia/Palmer, William (2019): Alexa, are you listening to me? An analysis of Alexa voice service network traffic. In: *Personal and Ubiquitous Computing* 23 (1), 67–79.
- Garfinkel, Harold (1960/2012): Die rationalen Eigenschaften von wissenschaftlichen und Alltagsaktivitäten. In: Ayaß, Ruth/Meyer, Christian (Hg.): *Sozialität in Slow Motion. Theoretische und empirische Perspektiven. Festschrift für Jörg Bergmann*. Wiesbaden: Springer VS, 41–57.
- Garfinkel, Harold (1967): *Studies in Ethnomethodology*. Cambridge: Polity.
- Gerwinski, Jan/Linz, Erika (2018): Methodik II: Beobachterparadoxon – die Aufnahmesituation im Gespräch. In: Gerwinski, Jan/Linz, Erika/Habscheid, Stephan (Hg.): *Theater im Gespräch. Sprachliche Publikumspraktiken in der Theaterpause*. Berlin/Boston: de Gruyter, 105–163.
- Goffman, Erving (1983): The Interaction Order. In: *American Sociological Review* 48, 1–17.
- Goodwin, Charles (2018): *Co-Operative Action*. New York: Cambridge University Press.
- Gray, Stacey (2016): *Always On: Privacy Implications of Microphone-Enabled-Devices*. In: Future of Privacy Forum. URL: [https://fpf.org/wp-content/uploads/2016/04/FPF\\_Always\\_On\\_WP.pdf](https://fpf.org/wp-content/uploads/2016/04/FPF_Always_On_WP.pdf).
- Habscheid, Stephan (2016): Handeln in Praxis. Hinter- und Untergründe situierter sprachlicher Bedeutungskonstitution. In: Depermann, Arnulf/Feilke, Helmuth/Linke, Angelika (Hg.): *Sprachliche und kommunikative Praktiken*. Berlin/New York: de Gruyter (IDS Jahrbuch 2015), 127–151.

- Habscheid, Stephan (2020): Rezension zu: Brock, Alexander & Peter Schildhauer (Hg.) (2017): *Communication Forms and Communicative Practices. New Perspectives on Communication Forms, Affordances and What Users Make of Them*. Frankfurt a. M.: Peter Lang (Language and Text Studies, Volume 15). In: *Zeitschrift für Rezensionen zur germanistischen Sprachwissenschaft* 12 (1-2), 150–155.
- Habscheid, Stephan/Hrnal, Christine/Carros, Felix/Lüssem, Jens (2020): Professionelle Emotionalität und humanoide Robotik in der institutionellen Kommunikation. In: Gruber, Helmut/Spitzmüller, Jürgen/de Cillia, Rudolf (Hg.): *Institutionelle und organisationale Kommunikation. Theorie, Methodologie, Empirie und Kritik*. Wien: V&R Unipress, 169–188.
- Hausendorf, Heiko/Kesselheim, Wolfgang/Kato, Hiloko/Breitholz, Martina (2017): *Textkommunikation: ein textlinguistischer Neuansatz zur Theorie und Empirie der Kommunikation mit und durch Schrift*. Berlin: de Gruyter (Reihe Germanistische Linguistik, 308).
- Hector, Tim Moritz/Hrnal Christine (2020): *Intelligente Persönliche Assistenten im häuslichen Umfeld. Erkenntnisse aus einer linguistischen Pilotstudie zur Erhebung audiovisueller Interaktionsdaten*. SFB 1187 Working Paper Series 14.  
URL: <https://www001.zimt.uni-siegen.de/ojs/index.php/wps1187/article/view/57/61>.
- Hennig, Martin/Hauptmann, Kilian (2019): Alexa, optimier mich! KI-Fiktionen digitaler Assistenzsysteme in der Werbung. In: *Zeitschrift für Medienwissenschaft* 11 (21), 86–94.
- Herbig, Daniel (2020): Amazon Alexa: Nutzer können Speichern von Sprachaufnahmen abschalten. In: *heise.de*, 24.09.2020.  
URL: <https://www.heise.de/news/Amazon-Alexa-Nutzer-koennen-Speichern-von-Sprachaufnahmen-abschalten-4912086.html?view=print>.
- Hirschauer, Stefan (2004): Praktiken und ihre Körper. Über die materiellen Partizipanden des Tuns. In: Hörning, Karl H./Reuter, Julia (Hg.): *Doing Culture. Neue Positionen zum Verhältnis von Kultur und sozialer Praxis*. Bielefeld: Transcript, 73–91.
- Kirchner, Stefan/Beyer, Jürgen (2016): Die Plattformlogik als digitale Marktordnung. In: *Zeitschrift für Soziologie* 45, 324–339.
- Krummheuer, Antonia (2010): *Interaktion mit virtuellen Agenten? Zur Aneignung eines ungewohnten Artefakts*. Stuttgart: Lucius & Lucius (Qualitative Soziologie, 11).
- Langlois, Ganaele/Elmer, Greg (2019): Impersonal subjectivation from platforms to infrastructures. In: *Media, Culture & Society* 41, 236–251.

- Latour, Bruno (2007): *Eine neue Soziologie für eine neue Gesellschaft. Einführung in die Akteur-Netzwerk-Theorie*. Berlin: Suhrkamp.
- Latour, Bruno (2009): Die Logistik der immutable mobiles. In: Döring, Jörg/Thielmann, Tristan (Hg.): *Mediengeographie. Theorie – Analyse – Diskussion*. Bielefeld: Transcript, 111–144.
- Latour, Bruno/Woolgar, Steve (1986): *Laboratory Life. The Construction of Scientific Facts*. 2. Auflage. Princeton, NJ: Princeton University Press.
- Lau, Josephine/Zimmerman, Benjamin/Schaub, Florian (2018): Alexa, Are You Listening? In: *Proceedings of the ACM on Human-Computer Interaction 2*, 1–31.
- Lopatovska, Irene/Rink, Katrina/Knight, Ian/Raines, Kieran/Cosenza, Kevin/Williams, Harriet/Sorsche, Perachya/Hirsch, David/Li, Qi/Martinez, Adrianna (2019): Talk to me: Exploring user interactions with the Amazon Alexa. In: *Journal of Librarianship and Information Science 51* (4), 984–997.
- Lotze, Netaya (2016): *Chatbots. Eine linguistische Analyse*. Berlin: Peter Lang.
- Lotze, Netaya (2018): Zur sprachlichen Interaktion mit Chatbots – eine linguistische Perspektive. In: Hug, Theo/Pallaver, Günther (Hg.): *Talk with the bots – Gesprächsroboter und Social Bots im Diskurs*. Innsbruck: Innsbruck University Press, 29–50.
- Lotze, Netaya (2020): Künstliche Intelligenz im Dialog – Ein methodologisches Konzept zur Analyse von Mensch-Maschine-Interaktion. In: Marx, Konstanze/Lobin, Henning/Schmidt, Axel (Hg.): *Deutsch in Sozialen Medien. Interaktiv – multimodal – vielfältig*. Berlin/Boston: de Gruyter (Jahrbuch des Instituts für Deutsche Sprache 2019), 363–368.
- Luger, Ewa/Sellen, Abigail (2016): "Like Having a Really Bad PA". The Gulf between User Expectation and Experience of Conversational Agents. In: Kaye, Jofish/Druin, Allison/Lampe, Cliff/Morris, Dan/Hourcade, Juan (Hg.): *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. New York: ACM, 5286–5297.
- Malkin, Nathan/Deatrick, Joe/Tong, Allen/Wijesekera, Primal/Egelman, Serge/Wagner, David (2019): Privacy Attitudes of Smart Speaker Users. In: *Proceedings on Privacy Enhancing Technologies 2019*, 250–271.
- Meiler, Matthias (2018): *Eristisches Handeln in wissenschaftlichen Weblogs. Medienlinguistische Grundlagen und Analysen*. Heidelberg: Synchron (Wissenschaftskommunikation, 12).
- Meiler, Matthias (2019): Zur praxeologischen Verhältnisbestimmung von Materialität, Medialität und Mentalität oder: Medien als Praxis. In: *Zeitschrift für Semiotik 41* (1–2), 63–88.

- Meyer, Christian/Ayaß, Ruth (2012): Einleitung. In: Ayaß, Ruth/Meyer, Christian (Hg.): *Sozialität in Slow Motion. Theoretische und empirische Perspektiven. Festschrift für Jörg Bergmann*. Wiesbaden: Springer VS, 11–18.
- Natale, Simone (2020): To believe in Siri: A critical analysis of AI voice assistants. In: *Communicative Figurations Working Paper* 32.  
URL: <https://www.kommunikative-figurationen.de/en/publications/working-papers/>.
- Natale, Simone/Cooke, Henry (2020): Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces. In: *Media, Culture & Society*. URL: <https://journals.sagepub.com/doi/abs/10.1177/0163443720983295>.
- 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), 1–39.
- Pickering, Andrew (1993): The Mangle of Practice: Agency and Emergence in the Sociology of Science. In: *American Journal of Sociology* 99, 559–589.
- Pins, Dominik/Boden, Alexander/Stevens, Gunnar/Essing, Britta (2020): „Miss understandable“ – A study on how users appropriate voice assistants and deal with misunderstandings. In: *Proceedings of Mensch und Computer 2020 (MUC20)*. ACM, Magdeburg, 349–359.
- Pitsch, Karola/Gehle, Raphaela/Dankert, Timo/Wrede, Sebastian (2017): Interactional Dynamics in User Groups. In: Wrede, Britta (Hg.): *Proceedings of the 5th International Conference on Human Agent Interaction*. Bielefeld, 10/17/2017 - 10/20/2017. New York: ACM Press, 393–397.
- Plantin, Jean-Christophe/Lagoze, Carl/Edwards, Paul N./Sandvig, Christian (2018): Infrastructure studies meet platform studies in the age of Google and Facebook. In: *New Media & Society* 20, 293–310.
- Porcheron, Martin/Fischer, Joel E./Reeves, Stuart/Sharples, Sarah (2018): Voice Interfaces in Everyday Life. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal, QC, Canada.  
URL: <http://www.cs.nott.ac.uk/~pszsr/files/porcheron-2018-voice-interfaces-in-everyday-life.pdf>.
- Pradhan, Alisha/Findlater, Leah/Lazar, Amanda (2019): "Phantom Friend" or "Just a Box with Information". In: *Proceedings of the ACM on Human-Computer Interaction* 3 (CSCW), 1–21.

- Reeves, Stuart/Porcheron, Martin/Fischer, Joel (2019): “This is not what we wanted”: Designing for Conversation with Voice Interfaces. In: *Interactions* 26 (1), 46–51.  
DOI: <https://doi.org/10.1145/3296699>.
- Rochet, Jean-Charles/Tirole, Jean (2003): Platform Competition in Two-Sided Markets. In: *Journal of the European Economic Association* 1, 990–1029.
- Röser, Jutta/Müller, Kathrin Friederike/Niemand, Stephan/Roth, Ulrike (2019): *Das mediatisierte Zuhause im Wandel. Eine qualitative Panelstudie zur Verhäuslichung des Internets*. Wiesbaden: Springer VS.
- Salheiser, Axel (2019): Natürliche Daten: Dokumente. In: Baur, Nina/Blasius, Jörg (Hg.): *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer, 1119–1134.  
DOI: [https://doi.org/10.1007/978-3-658-21308-4\\_80](https://doi.org/10.1007/978-3-658-21308-4_80).
- Schäfer, Hilmar (2021): Der Gebrauch des Digitalen. Zur praxeologischen Analyse digitaler Kultur. In: *Mittelweg* 36 30 (1), 3–14.
- Schegloff, Emanuel A. (2006/2012): Interaktion: Infrastruktur für soziale Institutionen, natürliche ökologische Nische der Sprache und Arena, in der Kultur aufgeführt wird. In: Ayaß, Ruth /Meyer, Christian (Hg.): *Sozialität in Slow Motion. Theoretische und empirische Perspektiven. Festschrift für Jörg Bergmann*. Wiesbaden: Springer VS, 245–268.
- Schiller, Amy/McMahon, John (2019): Alexa, Alert Me When the Revolution Comes. Gender, Affect, and Labor in the Age of Home-Based Artificial Intelligence. In: *New Political Science* 41 (2), 173–191.
- Schneider, Jan Georg (2017): Medien als Verfahren der Zeichenprozessierung: Grundsätzliche Überlegungen zum Medienbegriff und ihre Relevanz für die Gesprächsforschung. In: *Gesprächsforschung* 18, 34–55.
- Schubert, Cornelius (2019): Repair Work as Inquiry and Improvisation: The Curious Case of Medical Practice. In: Strebel, Ignaz/Bovet, Alain/Sormani, Philippe (Hg.): *Repair Work Ethnographies*. Singapore: Springer.
- Schüttpelz, Erhard (2016): Infrastrukturelle Medien und öffentliche Medien. In: *Media in Action* 0, 1–21 (Pre-Publication).  
URL: [https://www.mediacoop.uni-siegen.de/wp-content/uploads/2016/06/schuettpelz\\_infrastrukturelle\\_medien.pdf](https://www.mediacoop.uni-siegen.de/wp-content/uploads/2016/06/schuettpelz_infrastrukturelle_medien.pdf).
- Schüttpelz, Erhard/Meyer, Christian (2017): Ein Glossar zur Praxistheorie. „Siegener Version“ (Frühjahr 2017). In: *Navigationen* 17 (1), 155–163.

- Schüttpelz, Erhard/Meyer, Christian (2018): Charles Goodwin's Co-Operative Action: The Idea and the Argument. In: *Media in Action* 1, 171–188.  
URL: <https://www001.zimt.uni-siegen.de/ojs/index.php/mia/article/view/37>.
- Sciuto, Alex/Saini, Arnita/Forlizzi, Jodi/Hong, Jason (2018): "Hey Alexa, What's Up?" A Mixed-Methods Studies of In-Home Conversational Agent Usage. In: Koskinen, Ilpo/Lim, Youn-kyung/Cerratto-Pargman, Teresa/Chow, Kenny/Odom, William (Hg.): *Proceedings of the 2018 Conference on Designing Interactive Systems June 9-13, Hong Kong*. New York: Association for Computing Machinery, 857–868.
- Scollon, Ron/Scollon, Suzie Wong (2004): *Nexus Analysis: Discourse and the Emerging Internet*. London: Routledge.
- Selting, Margret/Auer, Peter/Barth-Weingarten, Dagmar/Bergmann, Jörg/Bergmann, Pia/Birkner, Karin/Couper-Kuhlen, Elizabeth/Deppermann, Arnulf/Gilles, Peter/Günthner, Susanne/Hartung, Martin/Kern, Friederike/Mertzluff, Christine/Meyer, Christian/Morek, Miriam/Oberzaucher, Frank/Peters, Jörg/Quasthoff, Uta/Schütte, Wilfried/Stukenbrock, Anja/Uhmann, Susanne (2009): Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). In: *Gesprächsforschung* 10, 353–402.
- Srnicek, Nick (2016): *Platform Capitalism*. Cambridge: MIT Press.
- Staab, Philipp (2019): *Digitaler Kapitalismus. Macht und Herrschaft in der Ökonomie der Unknappheit*. Berlin: Suhrkamp.
- Star, Susan Leigh/Bowker, Geoffrey C. (2006): How to Infrastructure. In: Lievrouw, Leah A./Livingston, Sonia M. (Hg.): *Handbook of New Media*. London: Sage, 151–162.
- Strengers, Yolande/Nicholls, Larissa (2018): Aesthetic pleasures and gendered tech-work in the 21st-century smart home. In: *Media International Australia* 166 (1), 70–80.
- Strüver, Niklas (2020): *Amazons Alexa als Infrastruktur der Zukunft? Eine qualitative Betrachtung von Nutzungspraktiken mit Sprachassistenten im Kontext digitaler Plattformtechnologien*. Masterarbeit an der RWTH Aachen.
- Tietze, Wolfgang/Roßbach, Hans-Günther (Hg.) (1991): *Mediennutzung und Zeitbudget. Ansätze, Methoden, Probleme*. Wiesbaden: Deutscher Universitätsverlag.
- Wiedemann, Lisa (2019): *Self-Tracking. Vermessungspraktiken im Kontext von Quantified Self und Diabetes*. Wiesbaden: Springer VS.
- Wissenschaftliche Dienste des Deutschen Bundestages (2019): *Zulässigkeit der Transkribierung und Auswertung von Mitschnitten der Sprachsoftware „Alexa“ durch Amazon*. Berlin. URL:

<https://www.bundestag.de/resource/blob/650728/3f72e6abc1c524961e5809002fe20f21/WD-10-032-19-pdf-data.pdf>.

Zuboff, Shoshana (2018): *Das Zeitalter des Überwachungskapitalismus*. Frankfurt/New York: Campus Verlag.

Zoeppritz, Magdalena (1985): *Computer talk? Technical Report TN 85.05*. Heidelberg: IBM Heidelberg Scientific Center.