Linguistic Practices as a Means of Domesticating Voice-Controlled Assistance Technologies

Stephan Habscheid, Tim Hector, and Christine Hrncal

Abstract *This chapter explores the linguistic practices involved in the domestication of voice-controlled smart speakers, drawing on findings from our research project "Un/desired observation in interaction: Intelligent Personal Assistants", conducted in Germany from 2020 to 2023. First, the characteristics of smart speakers and the methodological challenges of studying them empirically are outlined. This is followed by a theoretical discussion of how perspectives from interaction research and from linguistic media research can be integrated to complement one another. The empirical part of the chapter first elucidates the organizational characteristics of 'VUI dialogue' and then the organizational integration of VUIs in social interaction in multi-party constellations. Finally, we show how everyday usage practices in households are shaped by the linguistic conditions of their mediation on both organizational levels. The analyses also address the limitations of voice assistants as experienced by users, particularly when devices fail to meet expectations. Thus, this chapter sheds light on the complex dynamics of human–technology relationships and takes into account social and linguistic dimensions of technology domestication.*

1. Introduction: Smart Technologies between Public Discourse and Private Practice

Public discourse around AI-based 'smart' technologies has become increasingly controversial in the past few years – the more recent "hype" in the field of "communicative AI" (Hepp et al. 2023) has given this a new boost, but already before that, the increasing permeation of smart technologies into everyday life had already ignited an ongoing public debate. As Roe and Perkins (2023) point out in an analysis of AI discourse in the British press, public reporting on

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

AI sways dynamicly between two poles: the promise of great potential on the one hand and warnings about serious dangers on the other. Both poles can be potentially problematic – if the pendulum swings towards an overly positive and expectant attitude towards technological advancement, this not only creates exaggerated expectations of what applications can offer, but also plays into the hands of big tech companies. In academic studies, their products have been criticized for being non-transparent technologies (Liesenfeld, Lopez, and Dingemanse 2023) as well as a catalyst for racist (e.g. Phan 2017, 2019, Woods 2021; Leblebici, this volume) and sexist (e.g., Strengers and Kennedy 2020) biases in representing social order. Furthermore, the operations of the companies have been shown to be based on global inequality and exploitation (e.g., Crawford and Joler 2018, Couldry and Mejias 2019) as well as on excessive energy consumption and other environmental problems (Crawford 2021, Brevini 2023). On the other hand, AI-based, 'smart' technologies certainly have the potential to offer great benefits, e.g., in educational contexts (for an overview see Schiff and Rosenberg-Kima 2023), as assistance technologies (Albert, Hamann, and Stokoe 2023; Endter, Fischer, and Wörle 2023), in the context of smart cities, or in medical contexts (Levina et al. 2024).

However, the discourse on AI-based, networked, and data-driven technologies is not only debated in public, but also, significantly, in the private living environments of users – where they are actually used. In addition to pragmatic questions concerning device operation, users' and potential users' discussions and reflections revolve around comfort and assistance, surveillance, safety, data protection, and exploitation, as well as on human and non-human agency. Discourse in the 'public of the home' picks up on, continues, and evaluates public debates – and relates them to the everyday media practices of users themselves, as has been shown in research on the "domestication" of communication technologies such as television (Silverstone, Hirsch, and Morley 1992). Significantly, the integration of the internet and mobile technologies as well as data-driven and networked technologies into domestic life has blurred distinctions between public and private spheres and thus challenged the domestication research paradigm (Waldecker and Hector 2023, 14). Nevertheless, the domestic sphere continues to play a crucial role in society and in the organization of everyday life, and is one of the most significant areas of application for smart technologies (see Habscheid et al., this volume).

As earlier research grounded in the domestication research paradigm (Silverstone, Hirsch, and Morley 1992), sociology of knowledge and ethnomethod-

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

ology (Ayaß 2012) as well as conversational linguistics (Baldauf 2002) has shown for different media phenomena, language is central to processes of media domestication and appropriation. Analyzing linguistic practices is thus an apposite approach to investigate the anatomy of social practices that are affected and sometimes reshaped by media technologies – and to research the reflections and stances of users, in order to reveal not only how media technologies are embedded within the domestic community, but also how they are discussed and become part of its discourse. To date, however, empirical research on the everyday practices of users who actually use 'intelligent' and networked technologies in their living environments has been rare. Few studies have focused on linguistic practices and patterns in the domestication of media technologies.

In this chapter, we summarize the results of the linguistic strand of the research project "Un/desired observation in interaction: Intelligent Personal Assistants" within the Collaborative Research Center "Media of Cooperation" at the University of Siegen[1], which investigated linguistic media practices with voice assistants. We focus on the domestication of voice assistants – specifically, on how this is achieved linguistically – to make the contribution of this strand of research more visible in the increasingly complex academic discourse relating to such technologies. We begin with a brief characterization of smart speakers as a device type, elaborating on their specific features and the corresponding methodological consequences for empirical investigation (section 2). We then set out some theoretical assumptions concerning the relationship between linguistics, media, and praxeology (section 3). Based on these premises, we then draw on examples from our data corpus to illustrate the following practices of domestication: (a) the linguistic organization of one-on-one dialogues with smart speakers; (b) the speakers' linguistic embedding in multi-party-interactions; and (c) the ways in which users linguistically accomplish social practices in a collage of human and machinic utterances (sections 4.1–4.3). We finish with a summary and an outlook on the role of smart speakers within broader smart home ecosystems and the relevance of linguistic practices within these developments (section 5).

1 For a more detailed description of the research project and its contextualization within the Collaborative Research Center, see the the introduction to this volume.

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

2. Characteristics of Smart Speakers: How to Investigate them from an Empirical Linguistic Perspective

Smart speakers integrate virtual intelligent personal assistants with voice user interfaces (VUIs) in the form of a stationary device placed in the living environment. What interests us in particular about these interfaces is their voice-based operation via the acoustic channel: both inputs and outputs are processed verbally. Some dialogues are supplemented by acoustic signal tones, which can be conceptualized as "earcons" (Blattner, Sumikawa, and Greenberg 1989, 11), as well as by visual signs on the surface of the device. Voice assistants can be seen as a prototype for smart technologies: the recording, transmission, and utilization of data; the invisible connection to network publics as well as the embedding of the technology usage in sequential and incremental social interaction; the interweaving with everyday practices (and their transformation); the humanoid character of the systems; and the gradual adaptation of users to the linguistic restrictions (Hector and Hrncal 2024) are typical characteristics of smart technologies that come together in intelligent personal assistants with VUIs. Furthermore, the devices are associated with controversy, with (potential) users weighing the benefits of comfort, security, and assistance against surveillance, privacy breaches, and observability.

Krummheuer (2010) characterizes sociotechnical dialogue with virtual agents as a "hybrid exchange". On the one hand, such dialogues exhibit characteristics of interpersonal interactions based on a simulated similarity between technical and human actors. On the other hand, when malfunctions arise or communication is unsuccessful, differences between the human and the device come to the fore, and it is the human user who has to adapt to the limited interactional capabilities of the machine. Especially when disruptions occur, the focus can quickly shift from similarity to difference. The synthetic voice not only vocalizes machinic answers, but is modelled as an artificial companion, a 'persona'. Natale and Cooke (2021, 1009) stress that "[f]rom a technical viewpoint there isn't anything like one monolithic 'Alexa' or 'Siri'". From this perspective, these 'personae' function rather like metaphors to integrate a range of technical processes, such as speech recognition, natural language processing (NLP), and information retrieval within an interaction partner perceived as unified and singular. Sociotechnical exchange with the systems is mediated by such metaphors. This type of interface design follows on from attempts to make human–technology dialogues seem as 'natural' as possible: the non-human dialogue partner is addressed as a human conver-

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

sational partner would be, as advertisements for these products emphasize (Lind and Dickel 2024). Hence, the overriding design aim of this interface type is to allow users to integrate computer operations as 'seamlessly' as possible into their everyday lives. The interfaces therefore serve as a tool to mediate computer operations in the form of a linguistic dialogue (Merkle and Hector, forthcoming).

Voice assistants are always connected to a smartphone app. These apps include various functions: they give users control over the device in terms of settings and preferences, show connections to other smart home devices as well as to the internet, and enable touch-based control (for a detailed description, see Habscheid et al. 2021). Some of these apps also show a protocol of the usages of the smart speaker, including the recordings of audio in which a voice input was recognized. Habscheid et al. (2021) examine the analytic potential of the log file data that are recorded by voice-controlled systems and the documented activities in related smartphone applications. The log file data include not only the audio recording of the input, but also further information concerning the 'activities', presented graphically (such as a transcript of the recorded input and response and the time it occurred). They also provide further options for interaction with the database entries, such as providing feedback on whether the voice assistant did what was expected. At the same time, log file data serve to document fragments of the social situation they recorded. With these characteristics, they offer data by means of which the machine's performance and its 'understanding' of the recorded situation can be assessed. On the basis of the datafied recordings one can also draw conclusions about their further utilization as training data for speech recognition and NLP systems (see also Hector, forthcoming).

However, the data recorded by the systems do not provide enough information for research that seeks to analyze the entanglement of smart speaker systems, everyday life, and ongoing social situations in relation to linguistic practices (Hector and Hrncal 2020, 9; Habscheid et al. 2021, 44–45) – such a perspective calls for recordings of not just the 'voice command' itself, but of the social situation in which it takes place; the preparation and initiation of a voice dialogue and its subsequent evaluation and follow-up comments are crucial. This creates a methodological challenge for research into situational smart speaker usage: How can researchers record audio data in the private living environments of users, not just during discrete interaction situations, but whenever the smart speaker is used in daily life – without recording the whole living environment constantly (which would not just be ethically problematic

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

but would also produce an overwhelmingly vast volume of data)? Porcheron et al. (2018) developed a solution for this methodological problem: A specifically designed device called a "conditional voice recorder" (CVR) that can be placed in the living environments of study participants. Its basic function is to continuously record the audio in a certain room via a far-field microphone, but also to delete the recorded audio after a set duration of time, e.g., three minutes. The resulting three-minute audio recording is held in the buffer memory, only to be saved for longer if – and this is the key operating principle of the raspberry-pi based device – an invocation word such as "Alexa" is recognized by the built-in speech recognition of the CVR. When an activation word is recognized, the three buffer minutes are kept, together with three follow-up minutes, and saved on a connected flash drive[2]. Equipped with this technology for data collection, our research project was able to include within its ambit the analysis of multi-party situations involving more than one user and/or smart speaker, as well as the conversational preparation, initiation, and subsequent evaluation of VUI dialogues – and hence their embedding in turn-by-turn talk in social interaction (Habscheid, Hector, and Hrncal 2023, 7–8). We also drew on another form of data: video recordings produced to document how users installed their smart speakers for the first time.

3. Theoretical Foundations: Linguistics of Practices, Interaction, and Media

Approaches that seek to better understand communicative and cultural practices by tracing and reconstructing their linguistic (and multimodal) mediation have long been anchored in the theoretical traditions of anthropology and linguistics (e.g., Wittgenstein 1984; Hanks 1996; Luckmann 1986). More recently, linguistics has also been engaging with newer sociological approaches from the field of praxeology (e.g., Schatzki 2002; Reckwitz 2003; Hirschauer 2004, 2016; Deppermann, Feilke and Linke 2016a; on Hirschauer see also Habscheid, Hector, and Hrncal 2023).

The concept of 'practices' builds upon the insight that the use of linguistic means and forms indexically invokes – and thereby situationally modifies –

2 For a detailed description of the functions of the device, its further technological development for the context of the research project in Siegen, and the data practices associated with it, see Hector et al. (2022).

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

highly complex, 'gestalt-like' cultural knowledge contexts, the meaning of which extends far beyond simple speech acts and semantic content (Habscheid 2016). Additionally, the concept of 'practice' refers to a level of sociality that is logically prior to the handling of cultural resources in communication (Schüttpelz and Meyer 2017): Social practice is fundamentally established through cooperative production and temporality, it transcends language and knowledge in that it is also based on the participation of bodies, spaces, and technology (e.g., humans and voice user interfaces in connection with built environments, digital platforms, and infrastructures), often under asymmetrical conditions for cooperation (Hirschauer 2004, 2016; on VUIs see Habscheid et al. 2021). We revisit the asymmetrical types of participation associated with VUIs in section 4.2.

Cooperative practice requires material and sensory mediation, which may involve technical resources. Certain approaches in linguistics, like the tradition of 'Gesprächsforschung', focus in particular on cases of 'interaction' (Hausendorf 2015) in which "co-presence" or at least "tele-co-presence" (Hausendorf 2022) is established on the basis of synchronicity (and, in face-to-face conversation, also co-presence in physical space). The tradition of conversation analysis, which has strongly influenced linguistics (including 'Gesprächsforschung') investigates how interaction in co-presence is structured by means of language ("linguistic practices", cf. Deppermann, Feilke and Linke 2016b, 13) and manifold other material resources that can be functionalized as situated signs in the process of interaction (Goodwin 2018, 445). Whether and to what extent the findings of Conversation Analysis are applicable to dialogues with VUIs is one of the questions we address below (section 4.1). A further question is how social interaction changes under conditions of co-presence when VUIs participate asymmetrically (section 4.2). Finally, interaction forms the organizational backbone of communicative and cultural practices (including practices that extend beyond communication, like cooking or shopping), which can be observed in connection with the domestication of smart speakers in households (section 4.3).

One approach to incorporating media into linguistic theory is to assume that media formats presuppose and enable particular linguistic forms. This can be studied, for example, by focusing on 'communication forms' (*Kommunikationsformen*). This concept refers to the structural conditions of communication and language use that are characterized by the use of technical artifacts ('media' in a narrow sense) in connection with certain media institutions (e.g., the platforms of commercial IPA systems). In addition to

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

practices tied to 'genre' conventions (e.g., weather queries), communicative potentials that have not (yet) been exploited culturally can also come into focus: Thus, in their reinterpretation of the term, Brock and Schildhauer (2017, 13) define communication forms as "human-made clusters of technical and communicative constellations with communicative potentials, which are commonly restricted by conventions". Alternatively, a different perspective can be adopted, whereby 'media' are understood in a broader sense: as culturally established techniques, of which technical media can be a component. Accordingly, "medial procedures", e.g., communication mediated by language assistance systems, form the material and procedural side of the use of signs; this use is also embedded in conventional (communicative) practices (e.g., weather queries), and it furthermore depends on the individual competence of the sign user (Schneider 2017, 45).

Neither of these approaches take into account how media – beyond their communicative potential and their creative use by individuals – come into being *as social instances* and are thereby shaped on the basis of their materiality (Meiler 2019), vary dynamically, and can change over time. In order to be able to investigate such questions, the technical and cultural-theoretical view of media must be supplemented by a foundational, social-theoretical perspective: In a praxeological view, as pursued in the CRC 1187 "Media of Cooperation" at the University of Siegen, media are understood as "cooperatively created conditions of cooperation" or, in short, as "media of cooperation" (Schüttpelz 2017, 24).

This position, developed in the conversation analytic tradition by Charles Goodwin (2018) and at the intersection of conversation analysis and media theory by Erhard Schüttpelz and Christian Meyer (2017), does not ignore cultural consolidations and agreements (such as techniques, communicative genres, or symbols), nor does it take them as given. Instead, at a more foundational level of social order, it focuses on the concept of reciprocal "practice" (Schüttpelz and Meyer) or "co-operative action" (Goodwin): Participants in the production of meaning mutually produce processes by partially taking up the sign-like material resources (including, not least, indexical references) brought into play by their predecessors, and transforming them to reuse them for their own purposes. This theoretical premise is able to account for how communicative conventions can emerge and change, while identifying media themselves – without which communication is inconceivable – as always situationally emergent, brought forth in practice. It is this framework that underlies our research.

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

4. Analysis

The research that is discussed in the following sections 4.1–4.3, including the examples cited, is based on a corpus of video and audio data recorded in eight households. The recordings of initial installations comprised a total of one hour and 53 minutes of video filmed in six different situations. The audio files were recorded by the CVR in two different phases: generally, the first phase took place directly after the initial installation of the smart speaker, with a second phase three to four months later. In two households, smart speakers had already been installed before our data collection started, so we only conducted one recording phase. All the recording phases lasted between 20 and 30 days. This resulted in a total of 30 hours and 58 minutes of audio recordings, which were subsequently inventoried and transcribed according to the GAT 2-standard (Selting et al. 2011). For the video recordings, the transcription was supplemented multimodally following Mondada (2016).

4.1 Linguistic Organization I: The 'VUI Dialogue'

In order to be successful, sociotechnical dialogue between user and smart speaker – referred to in the following as VUI dialogue – must follow a specific sequential pattern, as shown in the following example. The excerpt stems from a CVR recording from the household of Beate (BW), a retired official in her sixties.

Example 1: How will the weather be today?

```
051 BW:  HEY google?
052 p:   (1.1)
053 BW:  wie wird das WETter heute?
         how will the weather be today?
054 p:   (1.7)
055 GA:  HEUte wird es in aachen überwiegend
         bewÖlkt,
         today in Aachen it will be mostly
         cloudy
```

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

```
056        bei temperaturen zwischen elf und
           NEUNzehn grAd.
           with temperatures between eleven and
           nineteen degrees
057 BW:    HM_hm;
058 p:     (1.0)
059 GA:    momenTAN ist es überwiegend bewÖlkt
           bei elf grAd.
           at the moment it is mostly cloudy with
           eleven degrees.
```

In addition to voice input (l. 053) and voice output (l. 055–056), the VUI dialogue in the example above also includes the invocation by means of an activation word or phrase – "HEY google?" (l. 051). The device responds to this by activating the listening mode, which enables the ensuing audio input to be processed. Devices also indicate activation visually with light signals on the speaker itself; Apple's smart speaker also emits an acoustic reception signal after a waiting period to indicate that it is ready to listen (Hector, in preparation). The exchange that follows activation is typically designed as a pair sequence – the input (first pair part) makes an output (second pair part) expected (Habscheid 2023, 188). Differently from the example above, the latter could take the form of a response to the input without any voice output from the smart speaker (e.g., fulfilling a task such as playing a certain song or turning on the lights). This basic sequence structure was represented very consistently across our data set (Hector, in preparation; forthcoming). Furthermore, dialogues can be expanded with a third element either by the user(s) or by the system. Sequence-ending third moves that are common in social interaction, such as the confirmation by Beate ("HM_hm;", l. 057) in the example above, often do not take place in VUI dialogues, and when they do, it often remains unclear whether or to what extent they are processed by the systems as indicators of comprehension. Longer user-initiated expansions without a further invocation are typically co-addressed to both the device and other co-present persons (Hector, in preparation) – as we elaborate below. When a third move is emitted by the device, this might, for instance, be to provide additional information as in the example (l. 059), or new features may be recommended, or tips, such as how to manage notifications. These utterances may be semantically linked to the preceding turns, but they do not have to be; they also can, but do not have to, make follow-up turns relevant

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

(Habscheid 2023, 189–90; Hector in preparation). In general, these expansions can function to generate user feedback, to compose more coherent sequential projects, or may be part of a general effort by providers to improve the service quality (Habscheid 2023, 188–90).

Regarding the linguistic design of the input, research on human–machine interaction has intensively discussed the notion of "computer talk" (first mentioned by Zoeppritz in 1985) and deliberated whether a distinct register for conversations with machines can be defined (see also Hector, in preparation). As summarized by Lotze (2016, 160–61), it has not been possible to empirically establish evidence of such a "register", as the actual linguistic practices involved in sociotechnical exchange with machines are too diverse and not specific. Short linguistic utterances, such as stand-alone nouns, imperatives, or deontic infinitives, which might at first glance appear to be characteristic of such conservations, have also been observed in other empractical contexts (e.g., Mondada 2014; Deppermann 2018), and are not unique to talk with machines (Hector, in preparation; Merkle and Hector, forthcoming). Furthermore, questions such as asking for a weather forecast in the example above (l. 053) are not the exclusive preserve of VUI dialogue. What does, however, seem to be specific to human–machine exchanges, is a "new form of dialogicity" (Lotze 2020, 363; Habscheid 2023, 174), which is characterized by a "broken-up" form of sequentiality (Krummheuer 2010, 229; Hector, in preparation). For VUIs, for example, this means that follow-up requests by users, if not initiated by the VUIs, always require a whole invocation sequence. Furthermore, sequential coherence between a first and a follow-up utterance – which in human-to-human conversations is often realized by the use of pronouns, for example – cannot be accomplished, or only to a very limited extent. The exchange between users and the system is most stable when the basic sequential structure mentioned above is adhered to, using adjacency pairs and short inputs (see also Barthel, Helmer and Reineke 2023). That does not mean that VUI dialogues are entirely predetermined, however. Their constitution is still an ongoing linguistic accomplishment between a machinic and a human interlocutor with very different initial conditions and it takes place under specific socio-spatial, material conditions.

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

4.2 Linguistic Organization II: VUIs as Participants in Multi-Party Interactions

A comparative analysis of one-on-one situations and multi-party settings makes it clear that sociotechnical dialogue and social interaction between humans are not one and the same – rather, users accomplish "transitions between sociotechnical exchange and social interaction" (Habscheid 2023, 176) and establish a dedicated "meta-interaction space" (Habscheid 2023, 176) – "a specific type of interaction between humans, which is directly related to the sociotechnical human–machine exchange" (Habscheid, Hector, and Hrncal 2023, 15). In this type of interaction participants may, for example, negotiate and reflect on the sociotechnical exchange, discuss smart speaker features in relation to ways of addressing the VUI, evaluate and deal with failures, or more generally discuss "the embedding of the smart speaker in the sequential unfolding of everyday practices" (Habscheid, Hector, and Hrncal 2023, 15). The relevance of this type of interaction has already been pointed out by Porcheron et al. (2018, 9): Users conceptually distinguish an exchange with a VUI based on inputs and outputs from one of "turns-at-talk". While the first is by design pre-planned to a certain degree, the unfolding of the talk is based on coherent conversational context and reciprocally constitutes itself in the latter.

The following excerpt (Example 2) illustrates how VUIs can be embedded in multi-party interactions by human users and to what extent the negotiation of agency can be intertwined with this. The excerpt was recorded in the shared apartment of Lukas (LF) and Alex (AK). Kurt (KS), a guest present at the time of the recording, introduces Lukas and Alex to a feature of their smart speaker that was previously unknown to them: the "Super Alexa Mode".

Example 2: "Super Alexa Mode"

```
155 KS:  (sag ma) kennst du den SUper alexa
         modus?
         say do you know the super alexa mode
156 p:   (1.7)
157 LF:  (den) WAS?
         the what
158 p:   (0.7)
```

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

```
159 KS:  <<p> WARte.>
              wait
160      alexa lautstärke
         SE[CHS-                  ]
         alexa volume six
161 k:     [((Musik wird leiser))]
              volume turned down
162      (1.2)
163 k:   ((Musik wird lauter))
           volume turned up
164 KS:  aLExa?
         alexa
165 k:   [((Musik wird leiser)) ]
            volume turned down
166      [(0.5)                  ]
167 KS:  hoch (.) hoch (.) runter runter (.)
         links rechts (.) links rechts (.)
         be: a: (.) START;
         up up down down left right
         left right b a start
168      (1.4)
169 AL:  super alexa modus wird aktiVIERT.
         super alexa mode activated
170      (0.6)
171 AL:  hiHI;
         hehe
172      (0.4)
173 AL:  zehn extra leben adDIERT.
         ten extra lives added
174      (0.6)
175 AL:  speed level drei FREIgeschaltet;
         speed level three unlocked
176      (0.6)
177 AL:  alle powerUPS verfügbar;
         all power-ups available
```

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

```
178 AK:  (die kann ja doch)
         she can
         [((unverständlich)) ]
            unintelligible
179 AL:  [FEHLfunktion;      ]
          malfunction
180 AK:  die kann auch in (deu[tschland) sich
         selbst ((unverständlich))]
         in Germany she can also herself
         ((unintelligible))
181 AL:                        [powerups haben
         keine POwer;             ]
                                  power-ups do
         not have any power
182      (0.4)
183 KS:  ((lacht))
           laughs
184 AL:  ABbruch;
         aborting
```

In this excerpt, a hidden function of the smart speaker, a so-called 'Easter egg' that has no function beyond entertainment, is personally presented 'in a favorable light' by Kurt to his interaction partners Lukas and Alex, who at the time were relatively inexperienced in using the device. Kurt's performance showcases the smart technology as exceptional and innovative (Habscheid, Hector, and Hrncal 2023, 8). The command to start the Super Alexa Mode is based on a sequence that is more commonly known for its use with game controllers: "up up, down down, left right, left right, A, B, start" is actually a sequence of keystrokes that activates advantages in video gaming. It has been a well-known 'classical' Easter egg in the gamer scene since the 1980s and works in many different games (Baumann 2023). Kurt, as connoisseur of the function, first establishes his host Lukas' level of awareness regarding this 'Easter egg' (l. 155); the latter's reaction in line 157 indicates his complete lack of knowledge. With his request "wait" (l. 159), Kurt then projects the subsequent demonstration of the feature, for which he first reduces the volume of the music playing via voice command (l. 160) and then utters the atypical voice command (l. 167). The VUI ratifies the input with a confirmation of activation of the Super Alexa Mode (l. 169–177), which is then cancelled proactively by the

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

system a little later (l. 180–184), presumably as part of the feature. By naming the feature in the beginning of the excerpt (l. 155 and 159) and by raising the volume for the presentation, Kurt frames the function and its demonstration as something "atypical" – beyond the usual set of commands used by household members. An asymmetry is thereby staged between Kurt, the well-informed guest who is familiar with the presented feature, and the household members Lukas and Alex to whom it is as yet unknown.

This leads to the question of how the devices participate in social practice, as discussed by Reeves and Porcheron (2023) and Hector (in preparation). To explore in what sense and to what extent the devices 'participate' – and are *treated as* participants by the co-present humans – it is helpful to revisit debates on the role of anthropomorphization that have been ongoing ever since technology began to become embedded in everyday life. The late 1990s and early 2000s saw the rise of a paradigm called "Computers as Social Actors" (CASA), which basically argued that humans mindlessly transfer habits from human-to-human communication to interactions with other entities perceived to exhibit a certain degree of 'life of their own', as computers may do, especially if their design is anthropomorphic (Nass and Moon 2000, 98). Reeves and Porcheron (2023) have interrogated these concepts intensively and challenge the notion that conversational AI systems, such as voice assistants, 'participate' in social interactions in the same ways that humans do. They argue that to assume so overlooks the fundamental role of social situations in making 'AI interactions' meaningful. Indeed, the significance of the ongoing social situation for an accomplishment of 'understanding' and 'meaning-making' was already demonstrated by Harold Garfinkel in 1967 in his research on the early chatbot ELIZA, which deployed rather simple sequence-orientated scripts to simulate an interaction based on connectable utterances (see Eisenmann et al. 2023, 6). Hence, instead of conceiving of voice assistants as social actors, Reeves and Porcheron (2023, 581) suggest that dialogues with these systems are better understood as regulated exchanges among participants within organized social (group) contexts, in which anthropomorphic utterances can be a linguistic resource.

The notion of participation is thereby not merely about direct interaction with the system, but focuses more intently on the broader social dynamics and the collaborative efforts of users to incorporate these technologies into their everyday lives. The situated, emergent nature of participation in VUI dialogues come to the fore. This is also in line with the findings of Hector (in preparation), who adopts the way participation is defined by Stefan Hirschauer (2004,

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

2016). Hirschauer argues from a praxeological standpoint that basically any material artifacts can be part of an action – with various different "levels of activity" (Hirschauer 2016, 49), ranging from active to passive on one spectrum and from active to prohibitive on another. Based on this model, combined with analyses of recordings of VUI dialogues in multi-party settings, Hector (in preparation) proposes from a linguistic standpoint that voice assistants can indeed be 'participants', but not in an equivalent way to human conversational partners. Human utterances seemingly directed 'to' the intelligent personal assistant following an input–output exchange are often, upon closer scrutiny, actually addressed to other humans as a form of co-address. Superficially, their linguistic contribution seems to be directed to the virtual interlocutor and can articulate polite formulas, evaluations, or follow-up requests. However, as no listening mode is activated by the device, at a pragmatic level these utterances should instead be seen as evaluations, frustration management, humor provocation, or other activities within a group of co-present speakers (see section 4.3 for an example).

The excerpt in Example 2 also shows that the embedding of smart speakers as participants in multi-party interactions can shed light on the negotiation of agency of technological actors, which – "especially in voice-based exchange with smart speaker technology" (Habscheid, Hector, and Hrncal 2023, 1) – is dynamically accomplished and intrinsically "bound to the local (linguistic) practices carried out by or rather involving contributions by participants with unequal resources for participating" (ibid.). The abovementioned meta-interaction space is highly relevant for this negotiation of agency between human and technical entities (Habscheid, Hector, and Hrncal 2023, 10). In line with the aforementioned praxeological perspectives, it makes sense to work with a praxeological conceptualization of agency, too. Krummheuer (2015) discusses how agency is dynamically constructed and negotiated within the interactional context, referring to concepts from Actor-Network-Theory (Latour 2005) as well as ethnomethodological conversation analysis. Rather than identifying it as an attribute inherent to either humans or technology, agency is viewed as a situationally emerging property of ongoing social interaction (see also Pentzold and Bischof 2019; Natale and Guzman 2022). In human–technology exchanges, participants might address the technology as a communicative partner, an actant, or even an "opponent", according to their situational needs and the unfolding interaction (Krummheuer 2015, 180). This concept of agency directs research on the linguistic unfolding of social practices with smart speakers towards the question of how users ascribe

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

and negotiate agency through their interactional practices. Our research has shown that agency occurs as a situational accomplishment that is linguistically negotiated between the participants (Habscheid, Hector, and Hrncal 2023), including the non-human 'participants'. The balance of agency can shift within very short time frames.

But how can we trace this in the excerpt presented above? Kurt, who is not a member of Lukas and Alex's household, claims agency by taking over the operation of the device, which is granted to him by the two flatmates who do not interfere. Here, agency is initially negotiated within the multi-party constellation of the human interactants. With the utterance "Alexa, up up down down left right left right B A start" (l. 164–167) directed towards the VUI, Kurt follows the script structure required to correctly execute his voice command, and in doing so he – at least to some degree – downgrades his own agency and attributes a certain level of agency to the technical device, which then realizes the output requested by Kurt's input. To some extent, the device itself then proactively indicates when the operation requested in Kurt's utterance has been redeemed to a satisfactory degree and is complete.[3]

4.3 The Linguistic Accomplishment of Social Usage Practices

In the process of 'domesticating' new technologies (Waldecker and Hector 2023), routine everyday practices are modified and new sociotechnical practices emerge. The latter initially serve to make the devices work and embed them into domestic spaces and infrastructures, which in turn are subject to processes of change and (re-)design. From the outset, users are challenged to overcome the 'resistance' of any new technology – often together with other people in co-presence – and to cope with the associated alienation (section 4.3.1). Over time, changed and new practices can become more or less deeply anchored in everyday domestic life. Insofar as all these social and sociotechnical cooperation processes are essentially linguistically and multimodally mediated, they can be investigated from a (media) linguistic perspective. This includes cases where attempts at domestication fail due to shortcomings of voice user interfaces (section 4.3.2).

3 It should be mentioned at this point that agency discernable "in front of" the device, in the interaction situation, can differ significantly from data practices at the "back end" of the device – if the latter are taken into account, the relative agency on the users' side is considerably diminished (Waldecker, Hector, and Hoffmann 2024).

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

Typically, users initially explore the practical potential of new technologies by testing and practicing. In the case of smart speakers, these processes are partly guided by the system through corresponding 'test scripts', but our observations show that users often go far beyond those in what they do (linguistically) in the early phase of appropriation (Habscheid 2023). While one approach is to cautiously probe the limits of the (linguistic) performance and flexibility of the systems, some users deliberately push the devices to their limits and beyond (Krummheuer 2010, 263) in order to then assert their superiority by ridiculing and exploiting bizarre utterances of the VUI to humorous effect (Krummheuer 2010, Chapter 9).

However, the appeal of such nonserious domestication practices can wear off over time. Furthermore, anyone who wants to make practical use of the technologies' capacity to increase convenience in everyday life is obliged to adapt to the limits of their linguistic communication capabilities rather than exacerbate them (Drösser 2020, 72). Smart speakers are capable, for example, of processing certain pair sequences. In contrast, utterances that can only be interpreted on the basis of conversational implicatures taking into account their sequential position irrespective of formal sequences, evidently represent an excessive challenge for the systems (at least those at the technical level we have investigated so far). This may incite amusement in an early phase of use (Habscheid 2023), but in the longer term users are more likely to accommodate their linguistic behavior towards that of the devices.

4.3.1 Early stage

The following excerpt (Example 3) is a transcript of a video excerpt from the documentation of an initial installation: Lukas (LF) and Alex (AK), two young men who live in a shared apartment (also protagonists in Example 2), are busy putting a smart speaker into operation:

Example 3: "You asked for mom"

```
421 LF:  §WILL ich das mit meinen kontakten verbinden,
          do I want to connect this to my
          contacts
    ak:  § enters the room and walks towards the sofa->
422 LF:   dann SPÄter,§
          later
    ak:              ->§
```

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

```
423 AK:  nein?
         no
424 AK:  §$((lacht))          $§
             laughs
    ak:  §sits down on the sofa§
    lf:   $smiles-------------$
425 p :  §(3.0)       §
    ak:  §grabs the laptop standing on the sofa, takes it
          onto his lap§
426 AK:  §außer du willst dass ich die ganze zeit damit
          RUMtrolle;
          unless you want me to troll with it all the
          time
    ak:  §leans forward and rummages through a few items
          on the coffee table->§
427 AK:  $alexa ruf meine MOM an;$§
          alexa call my mom
    lf:  $smiles-----------------$
    ak:                          ->§
428 AK:  +HAllo lukas mom;+
          hello lukas mom
    ak:  +looks at LF-----+
429 LF:  h°°
430 AL:  §du hast nach MAma gefragt,        §
          you asked for mom
    ak:  §leans back and operates the laptop§
431 AL:   [aber ich kann diesen namen in deinen
           kontakten] über die geräteliste nicht FINden;
           but I can't find this name in your contacts via
           the device list
432 AK:  $[((lacht laut und klatscht in die
            Hände)) ]$
              laughs loudly and claps hands
    lf:  $smiles-----$
433 AK:  ((lacht laut)) [((lacht))  ]
          laughs loudly  laughs
434 LF:                 [!UH!?      ]
```

In this initial installation phase, the users have not yet set up separate user accounts for the smart speaker. Typical work-sharing usage has not yet been established in this sense. Lukas is therefore hesitant to connect the shared stationary device with the contacts in his cloud (l. 421). Entering the room, Alex jokingly suggests that if Lukas did so, Alex could "troll around" with his flatmate's contacts (l. 426). He thus uses humor to raise the issue of intra-household surveillance: individual privacy may be relinquished to other household members if they use the same stationary smart speaker without differentia-

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

tion of accounts. Lukas adds weight to his joke with a hypothetical, quotation-like pattern of speech: "Alexa, call my Mom" (l. 427).

The practice of hypothetical quotation bears a formal resemblance to the initiation of a real dialogue with the smart speaker – which is indeed its effect. Alex's utterance "Alexa, call my Mom" is identified by the device as an invocation and linguistic input, but is not processed successfully: Instead of implementing the command, the system opens an insertion sequence with a kind of query: "You asked for Mom but I can't find this name in your contacts via the device list" (l. 430–431).

The fact that the smart speaker responds to an utterance not addressed to it as such and then processes it incorrectly may show the problem of the medial externalization of the household, but in this example case it mainly provides a cause for amusement: Alex exhibits his pleasure to the maximum by means of a linguistic staging procedure (Schmitt 2003) – loud, extended laughter accompanied by hand-clapping – while Lukas, also with a practice of staging, realizes an emphatically minimized format with a smile. While the entertainment value of such practices in the meta-interaction space is evident, they also indicate how users assure each other that they are in control of the system. It is then not the technology that "curates" social practices (Dolata 2019, 195), but rather the users who 'domesticate' the system by displaying their superiority.

4.3.2 Later stage

The following excerpt (Example 4) shows a usage situation from a later phase of domestication. Users are conducting a knowledge search, but the system proves to be extremely 'recalcitrant': not only does it repeatedly fail to provide the information sought, it also employs a dysfunctional and time-consuming presentation format that evidently causes frustration for users Robin (RL) and Lara (LS). Despite the disappointing experience, which is indicated mainly in the meta-interaction space (see section 4.2) through rejections (l. 201), comments (l. 222) and laughter (l. 218), Robin repeatedly initiates attempts to repair the dialogue, taking a trial-and-error approach. This is exhibited in abruptions (l. 204/205), new trials (l. 205), and reformulations of the request (l. 207, l. 234). Ultimately, use is discontinued with an expression of frustration (l. 242, l. 247) and a sweepingly negative evaluation ("You are useless", l. 254), which is answered by the assistance system ("I am still learning", l. 258):

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

Example 4: "When is the next bus?"

188	RL	hey GOOgle;
189		(0.8)
190	RL	WANN-
		when
191		[fährt der NÄCHste-]
		is the next
192	LS	[(halts MAUL.)]
		shut up
193	RL	bUs zum HAUPTbahnhof.
		bus to the main station
		(2.3)
194	GA	die BESte möglichkeit-
		the best option
195		um mit dem BUS zum nächsten
		dortmund- (.)
		to go by bus to the next dortmund
196		HUCKarde; (.)
		huckarde
197		HAUPTbahnhof-
		main station
198		an der adresse DORTmund zu kommen; (.)
		at the address dortmund
199		ist die linie NEUN um siebzehnuhrachtzehn von
		lückersweg.
		is line nine at seventeen ten from lückersweg
200		(0.2)
201	RL	NE-
		no
202		(0.3)
203	GA	DAS ist von-
		that is from
204		[hier etwa VIER min-]
		here about four min-
205	RL	[hey GOOgle;]
206		(0.3)
207	RL	wo ist die nächste bushalteSTELle;
208		(2.2)
209	GA	ich habe in deiner nähe mehrere einträge zum
		thema BUShaltestelle-
		I've found several entries on the topic 'bus
		stop'

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

```
210         [geFUNden (.) der-]
                        the
211   LS:   [(in deiner MAIL?)]
              in your mail
212   GA:   ERSte ist der hohler weg.
            first one is hohler weg
213         (0.3)
214   GA:   der ZWEIte ist küstenbecke a-
            the second one is küstenbecke A
215         SPARkassencenter;
            banking center
216         (0.4)
217   GA:   der DRITte ist küstenbecke c (.)
            sparkassencenter.
            the third one is küstenbecke C banking center
218   RL:   ((lacht))
             laughs
219         (0.6)
220   LS:   xxx;
221         (0.4)
222   RL:   ja (sind halt alles nicht die NÄCHsten xxx)-
            well, they are all just not the next (xxx)
223         (5.4)
224   RL:   (wo IS dat;)
             where is it
225         (1.5)
226   RL:   ((incomprehensible))
227         (1.1)
228   RL:   ((incomprehensible, loud noise))
229         (2.4)
230   RL:   hh°
231         (0.6)
232   RL:   hey GOOgle;
233         (0.4)
234   RL:   wAnn kommt der BUS;
            when is the bus coming
235         (2.3)
236   GA:   zum NÄCHsten hohler weg kommst du mit dem
            auto am schnellsten-
            the fastest way to get to the next hohler weg
            by car is
237   RL:   [((lacht))              ]
               laughs
238   GA:   [über golmer STRAße und-]
             via golmer straße and
```

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

```
239          [schmieDEstraße;]
              schmiedestraße
240    RL:   [hey GOOgle;    ]
241          (0.4)
242    RL:   stopp- (.)
243    GA:   die fahrt DAUert-
             the ride takes
244    GA:   [UNgefähr sechs-]
              about six
245    RL:   [hey GOOgle;    ]
246    GA:   [miNUT-]
247    RL:   [STOPP,]
248          (1.7)
249    GA:   okAY.
250          (0.4)
251    GA:   ABgebrochen.
             cancelled
252    RL:   hey GOOgle;
253          (0.6)
254    RL:   du bist NUTZlos;
             you are useless
255          (1.3)
256    GA:   naJA,
             well
257          (0.3)
258    GA:   ich LERne noch.
             I am still learning
```

Before the transcript starts, Robin's first attempt to obtain information about the bus connection had just failed; Lara's utterance (l. 192) reflects the frustration that is already beginning to set in. In the following minutes, the Google Assistant repeatedly produces voice outputs that do not match the desired query – mainly because an incorrect bus stop was selected as the starting point and the best stop is not identified even when queried directly. An opportunity for a local initiation of repair (Schegloff, Jefferson, and Sacks 1977, 374) does not present itself: There is no possibility for a *short* intervention or correction within the longer utterances of the VUI – the only option is to cancel whole oral text pieces. Additionally, unlike written search query results in a browser or smartphone navigation app, the temporality of speech production (Auer 2000) renders the information fleeting, inextricable, and difficult to compare. Identifying the source of trouble is therefore particularly difficult (see also Porcheron et al. 2018, 10; Garg, Cui, and Kapida 2021). Presets of the VUI, such as the selection of a car route instead of a bus route, cannot be viewed in advance in order to check or amend them before making the query,

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

which results in the continued verbalization of the unwanted information, as in l. 244. As the dialogue progresses, the user inputs become increasingly non-specific, more and more general. Whereas in one of the first voice inputs (uttered before the excerpt above) Robin specified very precisely when he wanted to travel from which bus stop to which destination, the last attempt is reduced to the question "When is the bus coming" (l. 234), with no local specification for start or end of the trip or for the time. This trajectory seems to result from repeated disappointments and failed trial-and-error attempts, leading from resignation to failure and a dwindling willingness to make any (linguistic) effort (see also Hector, in preparation).

5. From Smart Speakers to Smart Homes: An Outlook

With our analyses, we have been able to demonstrate that the social practices performed in interactions with smart speakers are already established everyday practices that are undertaken in changing medial conditions and hence undergo continual modification and transformation – consequently their accomplishment or non-accomplishment. Users have to sequentially cope with 'hybrid' dialogue systems that have been designed to simulate social interaction to a certain extent, yet also repeatedly deviate from it in ways that disrupt conversation. Thus, systems occasionally initiate the kind of phatic communication ("ich hoffe (.) du hattest einen TOLlen mittwoch./ I hope you had a great Wednesday") associated with small talk (Senft 2009), but sometimes they do so at inappropriate moments, in unsuitable contexts, and without the competence to respond appropriately to a follow-up move. In Example 4 ("When is the next bus?"), well-known problems from human–machine exchange as they have been described by Suchman (2007) occur: The dialogue design of the machine is rule-based, with the rules remaining opaque for the users, and the humans' situated utterances are only recognizable for the technology when they fit within the framework it 'anticipates'. These fundamental differences between the ways humans and machines process linguistic signals prevent the accomplishment of real "interaction", with openness and situation-boundedness presenting the most challenging characteristics of human interaction that machines need to cope with (Hector and Hrncal 2024). Resources that can increase the user-friendliness of interfaces, such as visualization and written elements, were not available for smart speakers at the time of our research – but since then, some more recent generations have incorporated screens.

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

What also becomes clear is that VUIs can sometimes be perceived as so bulky and dysfunctional that their domestication fails – despite the greatest linguistic effort. In cases when neither the user(s) nor the system achieve any situational agency – understood praxeologically (Krummheuer 2015) –, a recovery strategy can be observed on the part of the users: They assure one another of their assumed autonomy and superiority on a meta-level, for example through 'arrogant' jokes about or insults of the pseudo-social counterpart in its co-presence, such as the utterance "You are useless" in the last example. If one takes this user at his word, it becomes clear that the domestication of assistance systems can – in extreme cases – fail as a result of their voice interfaces' limitations.

Despite economic inefficiencies associated with VUIs (Amadeo 2022), their significance in consumer technology is apparently set to remain high due to the strategic ambitions of service providers such as Amazon, that appear to be aiming to establish VUIs as pivotal control hubs for smart home environments (Strüver 2023). As smart home technologies become increasingly sophisticated and data-intensive, the integration of advanced sensor technology into private residences is expected to grow, enhancing the capabilities and appeal of VUIs. Looking ahead, there are notable announcements suggesting the deployment of large language models (LLMs) within VUIs (see also Strüver, this volume). This development holds the potential to significantly enhance the conversational abilities and functionality of smart speakers. However, at least at the time of writing this chapter, experts point to hurdles for such technological evolution, at least for Amazon's smart speaker, resulting from path dependencies in the architecture (see Eric 2024). Additionally, there is speculation about discontinuation of Google's current smart speaker models, with a possible shift towards integrating newer voice assistant technology into tablets. Hence, linguistic practices may be as ephemeral as the technologies in relation to which they emerge – while at the same time, linguistic practices with VUIs demonstrate impressively the flexibility of competent speakers to adapt their practices for organization of talk to the organization of sociotechnical exchange.

Acknowledgements

We would like to express our sincere gratitude to Sarah Diehl and Franziska Niersberger-Guéye for their invaluable assistance with the transcription work and for their helpful comments on earlier drafts of this manuscript.

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project ID 262513311 – SFB 1187 Media of Cooperation. *Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 262513311 – SFB 1187 Medien der Kooperation.*

References

Amadeo, Ron. 2022. "Amazon Alexa is a 'colossal failure', on pace to lose $10 billion this year." *Ars Technica*, November 21, 2022. https://arstechnica.com/gadgets/2022/11/ amazon-alexa-is-a-colossal-failure-on-pace-to-lose-10-billion-this-year/.

Albert, Saul, Magnus Hamann, and Elizabeth Stokoe. 2023. "Conversational User Interfaces in Smart Homecare Interaction: A Conversation Analytic Case Study" In *Proceedings of 5th International Conference on Conversational User Interfaces*, edited by Minha Lee et al. New York: ACM. https://doi.org/10.1145/3571884.3597140.

Auer, Peter. 2000. "On line-Syntax – Oder: was es bedeuten könnte, die Zeitlichkeit der mündlichen Sprache ernst zu nehmen." *Sprache und Literatur* 31 (1): 43–56. https://doi.org/10.30965/25890859-031-01-90000005.

Ayaß, Ruth. 2012. "Introduction. Media Appropriation and Everyday Life." In *The Appropriation of Media in Everyday Life*, edited by Ruth Ayaß and Cornelia Gerhardt, 1–16. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.224.01aya.

Baldauf, Heike. 2002. *Knappes Sprechen*. Tübingen: Niemeyer.

Barthel, Mathias, Henrike Helmer, and Silke Reineke. 2023. "First users' interactions with voice-controlled virtual assistants: A micro-longitudinal corpus study." *Proceedings of the 27th Workshop on the Semantics and Pragmatics of Dialogue – SemDial*. PrePrint. Accessed August 22, 2024. https://mezzanine.um.si/wp-content/uploads/ Marilogue_Proceedings.pdf.

Baumann, Melanie. 2023. "Cheat-Code für den Super-Alexa-Modus. Amazon Alexa Cheat Code anwenden – so geht's". Accessed August 22, 2024. https://www.homeandsmart.de/alexa-cheat-code.

Blattner, Meera, Denise Sumikawa, and Robert Greenberg. 1989. "Earcons and Icons: Their Structure and Common Design Principles." *Human–Computer Interaction* 4 (1): 11–44. https://doi.org/ 10.1207/s15327051hci0401_1.

Brevini, Benedetta. 2023. "Myths, techno solutionism and artificial intelligence: reclaiming AI materiality and its massive environmental costs." In

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

Handbook of Critical Studies of Artificial Intelligence, edited by Simon Lindgren, 869–877. Chaltenham and Northampton: Edward Elgar.

Brock, Alexander and Peter Schildhauer, eds. 2017. *Communication Forms and Communicative Practices. New Perspectives on Communication Forms, Affordances and What Users Make of Them*. Frankfurt a.M.: Peter Lang.

Couldry, Nick, and Ulises Mejias. 2019. "Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject." *Television & New Media* 20 (4): 336–349. https://doi.org/10.1177/1527476418796632.

Crawford, Kate, and Vladan Joler. 2018. "Anatomy of an AI System: The Amazon Echo as an anatomical map of human labor, data and planetary resources." Accessed August 22, 2024. https://anatomyof.ai/.

Crawford, Kate. 2021. *Atlas of AI – Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale: Yale University Press.

Deppermann, Arnulf. 2018. "Instruction practices in German driving lessons: Differential uses of declaratives and imperatives." *International Journal of Applied Linguistics* 28 (2): 265–282. https://doi.org/10.1111/ijal.12198.

Deppermann, Arnulf, Helmuth Feilke, and Angelika Linke, eds. 2016a. *Sprachliche und kommunikative Praktiken*. Berlin and Boston: de Gryuter.

Deppermann, Arnulf, Helmuth Feilke, and Angelika Linke 2016b. "Sprachliche Praktiken. Eine Annäherung aus linguistischer Sicht." In *Sprachliche und kommunikative Praktiken*, edited by Arnulf Deppermann, Helmuth Feilke, and Angelika Linke, 1–23. Berlin and Boston: de Gryuter.

Dolata, Ulrich. 2019. "Plattform-Regulierung. Koordination von Märkten und Kuratierung von Sozialität im Internet." *Berliner Journal für Soziologie* 29: 179–206. https://doi.org/10.1007/s11609-020-00403-9.

Drösser, Christoph. 2020. *Wenn die Dinge mit uns reden. Von Sprachassistenten, dichtenden Computern und Social Bots*. Berlin: Dudenverlag.

Eisenmann, Clemens, Jakub Mlynář, Jason Turowetz, and Anne W. Rawls. 2023. "'Machine Down': making sense of human–computer interaction – Garfinkel's research on ELIZA and LYRIC from 1967 to 1969 and its contemporary relevance." *AI & Society*. https://doi.org/10.1007/s00146-023-01793-z.

Endter, Cordula, Florian Fischer, and Tobias Wörle. 2023. "Doing Home by Using Digital Assistive Technologies. On the Role of Meaning and Materiality in the use of Digital Assistive Technologies at Older People's Domesticity." *Digital Culture & Society* 9 (1): 179–200. https://doi.org/10.14361/dcs-2023-0109.

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

Eric, Mihail. 2024. "How Alexa Dropped the Ball on Being the Top Conversational System on the Planet." Accessed August 22, 2024. https://www.mihaileric.com/posts/how-alexa-dropped-the-ball-conversational-ai/.

Garfinkel, Harold. 1967. *Studies in Ethnomethodology.* Englewood Cliffs: Prentice Hall.

Garg, Radhika, Hua Cui, and Yash Kapadia. 2021. "'Learn, Use, and (Intermittently) Abandon': Exploring the Practices of Early Smart Speaker Adopters in Urban India." *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW2): 1–28. https://doi.org/10.1145/3479495.

Goodwin, Charles. 2018. *Co-Operative Action.* New York: Cambridge University Press.

Habscheid, Stephan. 2016. "Handeln in Praxis. Hinter- Und Untergründe Situierter Sprachlicher Bedeutungssituation." In *Sprachliche und kommunikative Praktiken*, edited by Arnulf Deppermann, Helmuth Feilke, and Angelika Linke, 127–51. Berlin and Boston: de Gruyter.

Habscheid, Stephan. 2023. "Socio-technical dialogue and linguistic interaction. Intelligent Personal Assistants (IPA) in the Private Home." *Sprache und Literatur* 51 (2): 167–196. https://doi.org/10.30965/25890859-05002020.

Habscheid, Stephan, Tim Hector, and Christine Hrncal. 2023. "Human and Non-Human Agency as Practical Accomplishment: Interactional Occasions for Ascription and Withdrawal of (Graduated) Agency in the Use of Smart Speaker Technology". *Social Interaction. Video-Based Studies of Human Sociality* 6 (1): 1–31. https://doi.org/10.7146/si.v6i1.137378.

Habscheid, Stephan, Tim Hector, Christine Hrncal, and David Waldecker. 2021. "Intelligente Persönliche Assistenten (IPA) mit Voice User Interfaces (VUI) als 'Beteiligte' in häuslicher Alltagsinteraktion. Welchen Aufschluss geben die Protokolldaten der Assistenzsysteme?" *Journal für Medienlinguistik* 4 (1): 16–53. https://doi.org/10.21248/jfml.2021.44.

Hanks, William F. 1996. *Language and Communicative Practices.* Oxford and Boulder: Westview Press.

Hausendorf, Heiko. 2015. "Interaktionslinguistik". In *Sprachwissenschaft im Fokus. Positionsbestimmungen und Perspektiven*, edited by Ludwig M. Eichinger, 43–70. Berlin and Boston: de Gruyter.

Hausendorf, Heiko. 2022. "'Telekopräsenz'. Interaktionslinguistische Anmerkungen zu einer Kommunikationsbedingung im Wandel." In *Brückenschläge: Linguistik an den Schnittstellen*, edited by Sarah Brommer, Kersten Sven Roth, and Jürgen Spitzmüller, 205–44. Tübingen: Narr Francke Attempto.

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

Hector, Tim. 2023. "Smart Speaker in der Praxis. Methodologische Überlegungen zur medienlinguistischen Erforschung stationärer Sprachassistenzsysteme." *Sprache und Literatur* 51 (2): 197–229. https://doi.org/10.30965/25890859-05002021.

Hector, Tim. Forthcoming. "Sprachassistenzsysteme und ihre Interfaces: eine medienlinguistische Analyse" In *Un/reale Interaktionsräume. Formen sozialer Ordnung im Spektrum medienspezifischer Interaktion,* edited by Clara Kindler-Mathôt, Didem Leblebici, Giacomo Marinsalta, Till Rückwart, and Anna Zaglyadnova. Bielefeld: transcript.

Hector, Tim. In preparation. *Smart Speaker als Gesprächsbeteiligte? Sprachliche Praktiken im Prozess der Domestizierung mit Voice-User-Interfaces.* Dissertation, University of Siegen, Faculty of Arts and Humanities – Faculty I. Defense: April 12, 2024.

Hector, Tim Moritz, and Christine Hrncal. 2020. "Intelligente Persönliche Assistenten im häuslichen Umfeld. Erkenntnisse aus einer linguistischen Pilotstudie zur Erhebung audiovisueller Interaktionsdaten." *Working Paper Series Media of Cooperation* 14: 1–14. http://dx.doi.org/10.25819/ubsi/1013.

Hector, Tim Moritz, and Christine Hrncal. 2024. "Sprachassistenzsysteme in der Interaktion". In *Sprache und digitale Kommunikation*, edited by Jannis Androutsopoulos und Friedemann Vogel, 309–28. Berlin and Boston: de Gruyter.

Hector, Tim, Franziska Niersberger-Gueye, Franziska Petri, and Christine Hrncal. 2022. "The 'Conditional Voice Recorder': Data practices in the co-operative advancement and implementation of data-collection technology." *Working Paper Series Media of Cooperation* 23: 1–15. https://doi.org/10.25819/ubsi/10168.

Hepp, Andreas, Wiebke Loosen, Stephan Dreyer, Juliane Jarke, Sigrid Kannengießer, Christian Katzenbach, Raina Malaka, Michaela Pfadenhauer, Cornelius Puschmann, and Wolfgang Schulz. 2023. "ChatGPT, LaMDA, and the Hype Around Communicative AI: The Automation of Communication as a Field of Research in Media and Communication Studies." *Human-Machine-Communication* 6: 41–63. https://doi.org/ 10.30658/hmc.6.4.

Hirschauer, Stefan. 2004. "Praktiken und ihre Körper. Über die materiellen Partizipanden des Tuns." In *Doing Culture. Neue Positionen zum Verhältnis von Kultur und sozialer Praxis,* edited by Julia Reuter and Karl H. Hörning, 73–91. Bielefeld: transcript.

Hirschauer, Stefan. 2016. "Verhalten, Handeln, Interagieren. Zu den mikrosoziologischen Grundlagen der Praxistheorie." In *Praxistheorie. Ein soziologi-*

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

sches Forschungsprogramm, edited by Hilmar Schäfer, 45–67. Bielefeld: transcript. https://doi.org/10.1515/9783839424049-003.

Krummheuer, Antonia. 2010. *Interaktion mit virtuellen Agenten? Zur Aneignung eines ungewohnten Artefakts.* Stuttgart: Lucius&Lucius.

Krummheuer, Antonia. 2015. "Technical Agency in Practice: The enactment of artefacts as conversation partners, actants and opponents." *PsychNology Journal* 13 (2–3): 179–202.

Latour, Bruno. 2005. *Reassembling the Social. An Introduction to Actor-Network-Theory.* Oxford: Oxford University Press.

Levina, Anastasia, Igor Ilin, Dayana Gututishvili, Kristina Kochetkova, and Andrea Tick. 2024. "Towards a Smart Hospital: Smart Infrastructure Integration." *Journal of Open Innovation: Technology, Market, and Complexity* 10 (3): 1–12.

Liesenfeld, Andreas, Alianda Lopez, and Mark Dingemanse. 2023. "Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators." *Proceedings of the 5th International Conference on Conversational User Interfaces*, 1–6. https://doi.org/10.1145/3571884.3604316.

Lind, Miriam, and Sascha Dickel. 2024. "Speaking, but having no voice. Negotiating agency in advertisements for intelligent personal assistants." *Convergence* 30 (3), 1008–1024. https://doi.org/10.1177/13548565231192100.

Lotze, Netaya. 2016. *Chatbots.* Frankfurt a.M.: Peter Lang.

Lotze, Netaya. 2020. "Künstliche Intelligenz im Dialog – Ein methodologisches Konzept zur Analyse von Mensch-Maschine-Interaktionen." In *Deutsch in Sozialen Medien*, edited by Konstanze Marx and Henning Lobin, 363–67. Berlin, Boston: de Gruyter.

Luckmann, Thomas. 1986. "Grundformen der gesellschaftlichen Vermittlung des Wissens: Kommunikative Gattungen." In *Kultur und Gesellschaft* (= *Kölner Zeitschrift für Soziologie und Sozialpsychologie.* Sonderheft 27), edited by Friedhelm Neidhardt, M. Rainer Lepsius, and Johannes Weiß, 191–211. Opladen: Westdeutscher Verlag.

Meiler, Matthias. 2019. "Zur praxeologischen Verhältnisbestimmung von Materialität, Medialität und Mentalität oder: Medien als Praxis." *Zeitschrift für Semiotik* 41 (1–2): 63–88. https://doi.org/10.14464/zsem.v41i1-2.709.

Merkle, Benedikt, and Tim Hector. Forthcoming. "Werkzeuge und Medienpraktiken. Intelligente persönliche Assistenten und das Paradigma objektorientierten Programmierens." In *Akustische Interfaces. Interdisziplinäre Perspektiven auf Schnittstellen von Technologien, Sounds und Menschen*, edited by

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

Christoph Borbach, Timo Kaerlein, Robert Stock, and Sabine Wirth. Wiesbaden: Springer Vieweg.

Mondada, Lorenza. 2014. "Instructions in the operating room. How the surgeon directs their assistant's hands." *Discourse Studies* 16 (2): 131–161. https://doi.org/10.1177/1461445613515325.

Mondada, Lorenza. 2016. "Conventions for multimodal transcription." Accessed June 15, 2024. https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/ mondada_multimodal_conventions.pdf.

Nass, Clifford, and Youngme Moon. 2000. "Machines and Mindlessness: Social Responses to Computers." *Journal of Social Issues* 56 (1): 81–103. https://doi.org/10.1111/0022-4537.00153.

Natale, Simone, and Henry Cooke. 2021. "Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces." *Media, Culture & Society* 43 (6): 1000–1016. https://doi.org/10.1177/0163443720983295.

Natale, Simone, and Andrea L. Guzman. 2022. "Reclaiming the human in machine cultures." *Media, Culture & Society* 44 (4): 627–637. https://doi.org/10.1177/01634437221099614.

Pentzold, Christian, and Andreas Bischof. 2019. "Making affordances real: socio-material prefiguration, performed agency, and coordinated activities in human–robot communication." *Social Media + Society* 5 (3): 1–11. https://doi.org/10.1177/2056305119865472.

Phan, Thao. 2017. "The Materiality of the Digital and the Gendered Voice of Siri." *Transformations* 29: 23–33. https://www.transformationsjournal.org/.

Phan, Thao. 2019. "Amazon Echo and the Aesthetics of Whiteness." *Catalyst: Feminism, Theory, Technoscience* 5 (1): 1–39. https://doi.org/10.28968/cftt.v5i1.29586.

Porcheron, Martin, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. "Voice Interfaces in Everyday Life." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems – CHI '18*. 1–12. https://doi.org/10.1145/3173574.3174214

Reckwitz, Andreas. 2003. "Grundelemente einer Theorie sozialer Praktiken." *Zeitschrift für Soziologie* 32: 282–301.

Reeves, Stuart, and Martin Porcheron. 2023. "Conversational AI: Respecifying Participation as Regulation". In *The SAGE Handbook of Digital Society,* edited by William Housley, Adam Edwards, Roser Beneito-Montagut, and Richard Fitzgerald, 573–92. London: Sage.

Roe, Jasper, and Mike Perkins. 2023. "'What they're not telling you about ChatGPT': exploring the discourse of AI in UK news media headlines." *Humani-*

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

ties and Social Sciences Communications 10: 1–9. https://doi.org/10.1057/s41599-023-02282-w.

Schegloff, Emanuel A., Gail Jefferson, and Harvey Sacks. 1977. The Preference for Self-Correction in the Organization of Repair in Conversation. *Language* 53 (2): 361–382.

Schatzki, Theodore R. 2002. *The Site of the Social. A Philosophical Account of the Constitution of Social Life and Change.* University Park, PA: Pennsylvania State University Press.

Schiff, Daniel S., and Rinat B. Rosenberg-Kima. 2023. "AI in Education: Landscape, Vision and Critical Ethical Challenges in the 21st Century." In *Handbook of Critical Studies of Artificial Intelligence*, edited by Simon Lindgren, 804–14. Cheltenham, Northampton: Edward Elgar.

Schmitt, Reinhold. 2003. "Inszenieren: Struktur und Funktion eines gesprächsrhetorischen Verfahrens." *Gesprächsforschung* 4: 186–250. http://www.gespraechsforschung-online.de/.

Schneider, Jan Georg. 2017. "Medien als Verfahren der Zeichenprozessierung: Grundsätzliche Überlegungen zum Medienbegriff und ihre Relevanz für die Gesprächsforschung". In *Gesprächsforschung* 18: 34–55. http://www.gespraechsforschung-online.de/.

Schüttpelz, Erhard. 2017. "Infrastructural Media and Public Media". *Media in Action* 1(1): 13–61. http://dx.doi.org/10.25819/ubsi/7935.

Schüttpelz, Erhard, and Christian Meyer. 2017. "Ein Glossar zur Praxistheorie. ‚Siegener Version' (Frühjahr 2017)". *Navigationen* 17 (1): 155–163. https://doi.org/10.25969/mediarep/1762.

Senft, Gunter. 2009. "Phatic communion" In *Culture and language use*, edited by Gunter Senft, Jan-Ola Östman, and Jef Verschueren, 226–33. Amsterdam John Benjamins. https://doi.org/10.1075/hoph.2.20sen.

Selting, Margret, Peter Auer, Dagmar Barth-Weingarten, Jörg Bergmann, Pia Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzlufft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock, and Susanne Uhmann. 2011. "A system for transcribing talk-in-interaction: GAT 2." *Gesprächsforschung* 12: 1–51. http://www.gespraechsforschung-online.de/.

Silverstone, Roger, Eric Hirsch, and David Morley. 1992. "Information and communication technologies and the moral economy of the household." In *Consuming Technologies. Media and Information in Domestic Spaces*, edited by Roger Silverstone and Eric Hirsch, 13–28. London: Routledge.

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - (cc) BY-NC

Strengers, Yolande, and Jenny Kennedy. 2020. *The smart wife. Why Siri, Alexa, and other smart home devices need a feminist reboot.* Cambridge: The MIT Press.

Strüver, Niklas. 2023. "Frustration Free: How Alexa Orchestrates the Development of the Smart Home." *Digital Culture & Society* 9 (1): 99–123. https://doi.org/10.14361/dcs-2023-090106.

Suchman, Lucy. 2007. *Human-Machine Reconfigurations: Plans and Situated Actions. 2nd Edition*. Cambridge: Cambridge University Press.

Waldecker, David and Tim Hector. 2023. "A Praxeological Approach Towards the Domestication of Connected Media Technologies. Introduction." *Digital Culture & Society* 9 (1): 5–22. https://doi.org/10.14361/dcs-2023-0102.

Waldecker, David, Tim Hector, and Dagmar Hoffmann. 2024. "Intelligent Personal Assistants in practice. Situational agencies and the multiple forms of cooperation without consensus." *Convergence* 30 (3), 975–91. https://doi.org/10.1177/13548565231189584.

Wittgenstein, Ludwig. 1984. "Philosophische Untersuchungen." In *Ludwig Wittgenstein. Tractatus logico-philosophicus. Tagebücher 1914–1916. Philosophische Untersuchunngen. Werkausgabe Band 1.* Frankfurt a.M.: Suhrkamp.

Woods, Heather. 2021. "Smart homes: domestic futurity as Infrastructure." *Cultural Studies* 35 (4–5): 876–899. https://doi.org/10.1080/09502386.2021.1895254.

Zoeppritz, Magdalena. 1985. "Computer Talk?" *Technical Report TN 85.05.* Heidelberg: IBM Germany Heidelberg Scientific Center.

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access -

https://doi.org/10.14361/9783839472002-008 https://www.inlibra.com/de/agb - Open Access - CC BY-NC