

# »Es ist vor allen Dingen dieser Ton«

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

---

Britta Schneider

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

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

## 1 Sprechen mit Computern und die Affordanzen von Lautlichkeit

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

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

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

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

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

## 2 Theoretische Hintergründe und Forschungsstand – Sprachideologie, Posthumanismus, Lautlichkeit

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

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

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

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

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

---

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

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

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

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

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

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

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

---

<sup>3</sup> Die Vorstellung, dass Informationen vor dem Zeichen existieren und nur in Zeichen ›hineinge packt‹ werden, ist problematisch und wird vom Medienverständnis der Autorin nicht geteilt.

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

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

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

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

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

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

---

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

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

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

### **3 Sprachideologien zu Lautlichkeit im Umgang mit stimmgesteuerten Computern**

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

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

### 3.1 Problematisierungen von Befehlsintonation

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

#### *Excerpt 1<sup>5</sup>*

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

---

5 Siehe Appendix für Transkriptionskonventionen.

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

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

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

*Excerpt 2*

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

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

### 3.2 Erziehung zum ›richtigen‹ Ton

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

---

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

*Excerpt 3*

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

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

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

*Excerpt 4*

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

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

---

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

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

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

*Excerpt 5*

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

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

### 3.3 Lautliches Sprechen und die Entwicklung von Sozialbeziehungen

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

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

*Excerpt 6*

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

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

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

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

### 3.4 Die Rolle von Lautlichkeit in oraler Kultur

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

#### *Excerpt 7*

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

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

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

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

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

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

#### **4 Die vergessene Materialität sprachlicher Zeichen in der europäischen Moderne**

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

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

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

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

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

---

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

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

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

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

## 5 Zusammenfassung und Ausblick

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

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

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

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

## Literaturangaben

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

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

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

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

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

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

## Appendix

### Transkriptionskonventionen

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

