

Astrid Adler

Sprachstatistik in Deutschland

Abstract

Seit 2017 wird im deutschen Mikrozensus eine Frage zur Sprache der Bevölkerung gestellt. Die letzte Spracherhebung in einem deutschen Zensus datiert aus dem Jahr 1939; entsprechend gibt es aktuell keine aussagekräftigen Sprachstatistiken in Deutschland. Die neue Sprachfrage des Mikrozensus weist jedoch erhebliche Mängel auf; offensichtlich wurde sie als Stellvertreterfrage zur Messung kultureller Integration konzipiert. Im vorliegenden Text werden die Fragen diskutiert und ihre ersten Ergebnisse analysiert. Daran anschließend werden andere Varianten von Sprachfragen dargestellt, dabei wird insbesondere auf die vorbildlichen Sprachfragen im kanadischen Zensus eingegangen. Abschließend wird die Sprachfrage der Deutschland-Erhebung 2018 des IDS inklusive ihrer Ergebnisse vorgestellt; die Deutschland-Erhebung 2018 stellt neben dem Mikrozensus bislang die einzige repräsentative Spracherhebung in Deutschland dar.

Since 2017 the German micro census has contained a question about the language of the population. The last language survey in a German census dates from 1939, so there are currently no dependable language statistics in Germany. However, the new language question in the Microcensus has considerable shortcomings. It was clearly conceived as a proxy measure of cultural integration. The present article discusses the question and analyses its first results. It then reports on other variants of language questions, with particular reference to the exemplary language questions in the Canadian census. Finally, it looks at the language question in the Germany Survey 2018 undertaken by the Institute for the German Language (IDS), together with its results. Alongside the micro census, the Germany Survey 2018 has so far been the only representative language survey in Germany.

1. Sprache und Statistik

Mit der Erhebung der Sprache bzw. der Sprachen der Bevölkerung wurde in einigen preußischen Volksregistern schon in der ersten Hälfte des 19. Jahrhunderts begonnen (vgl. Labbé 2003). Im 19. und frühen 20. Jahrhundert etablierte sich eine solche Frage nach der Sprache als fester Bestandteil der preußischen Volksbefragungen. Damit waren damals vor allem sprachpolitische Ziele verbunden; so sollte beispielsweise die Effektivität der kommunalen Sprachpolitik überprüft werden (vgl. ebd., siehe auch Glück 1979). Diese Tradition endete 1939; seither wurde in Volksbefragungen in Deutschland nicht mehr nach Sprache gefragt. Erst im Jahre 2017 wurde eine neue Frage zur Sprache der Bevölkerung in den deutschen Mikrozensus aufgenommen; diese Frage wurde 2018 wiederholt, 2019 geht sie in die dritte Runde.

Aus sprachwissenschaftlicher Sicht ist es sicherlich begrüßenswert, dass es jetzt endlich statistische Daten zur Sprache der Bevölkerung geben wird. Dies wurde schon oft als Desiderat formuliert (vgl. z. B. Kemper 2010). Für Angaben zur Sprache der Bevölkerung musste man sich bisher mit Rückschlüssen aus anderen erfassten Daten behelfen (z. B. der Staatsangehörigkeit); sprachliche Gegebenheiten können jedoch nur unzureichend über solche Umwege erfasst werden. Noch 2012 bemerkt Stickel, dass es in Deutschland keinerlei Sprachzensus gebe und ein solcher „in absehbarer Zeit nicht zu erwarten ist“ (Stickel 2012, S. 229). Dass eine Spracherhebung im Zensus so lange gefehlt hat, ist durchaus bemerkenswert. Einerseits werden in Deutschland viele Daten zu den verschiedensten Themen erhoben, und andererseits ist das Deutsche sprachwissenschaftlich auch recht gut erfasst. Es gibt sogar repräsentative Daten zu den Spracheinstellungen der Menschen in

Deutschland (vgl. Gärtig/Plewnia/Rothe 2010; Adler/Plewnia 2018). Nur eine aussagekräftige Sprachstatistik, die sozusagen Sprache und Statistik miteinander verknüpft, hat es bislang nicht gegeben. Das hängt wohl auch damit zusammen, dass es in Deutschland keine kohärente Sprachpolitik gibt. Zu einer Sprachpolitik gehören solide Sprachstatistiken zur Beobachtung und Steuerung dazu. Das Fehlen einer solchen Sprachpolitik in Deutschland liegt sicherlich auch daran, dass aus nichtlinguistischer Perspektive kein Bedarf dafür gesehen wird, weil die deutsche Sprache das gesamte (schriftliche) öffentliche Leben in Schulen, Politik, Rechtswesen und Verwaltung dominiert. Obwohl es in Deutschland tatsächlich viele Sprecher anderer Sprachen als Deutsch gibt, ist Deutschland konzeptionell und institutionell einsprachig und wird auch so wahrgenommen (vgl. Adler/Beyer 2018). Es gibt zwar vier autochthone Minderheiten- und eine Regionalsprache in Deutschland, die über die Europäische Charta der Regional- oder Minderheitensprachen anerkannt und geschützt sind (Dänisch, Friesisch, Romanes, Sorbisch und Niederdeutsch), und außerdem noch die Deutsche Gebärdensprache, die ebenfalls gesetzlich anerkannt ist (über das Behindertengleichstellungsgesetz), in großen Teilen der Öffentlichkeit und in der laienlinguistischen Wahrnehmung spielen diese Sprachen aber mehrheitlich kaum eine Rolle. Neben diesen autochthonen gibt es zahlreiche allochthone Minderheiten mit nennenswerten Sprechergruppen. Die Sprachen dieser Gruppen haben keinen offiziellen Status, sind aber im Alltag durchaus präsent und zum Teil auch im offiziellen Geschehen verankert (z. B. mit dem herkunftssprachlichen Unterricht in der Schule, über mehrsprachiges Informationsmaterial der Verwaltungen u. ä.). Dennoch gibt es bislang keine belastbaren Zahlen über die Sprecher dieser Sprachen in Deutschland. Das heißt also auch, dass jeglicher Steuerungs- und Förderprozess in diesem Bereich ohne Datengrundlage durchgeführt wird.

Die Kenntnis über Sprecherzahlen ist eine wesentliche Voraussetzung nicht nur für sprachpolitische Vorhaben, sondern auch für bildungspolitische und gesellschaftspolitische Maßnahmen. Das Wissen beispielsweise über die Anzahl von Sprechern einer Minderheitensprache kann dazu genutzt werden, die Förderpolitik entsprechend auszurichten und geeignete Maßnahmen zu identifizieren. Der Nutzen von Sprecherzahlen für die Schulpolitik liegt auf der Hand. Bis heute gibt es dazu auf nationaler Ebene allerdings keine Zahlen, obwohl die Bundesländer seit Jahren Schulstatistiken erheben (siehe dazu Abschn. 2.3). Auch gesellschaftspolitisch wäre ein Wissen über die in Deutschland gesprochenen Sprachen und die Zahl ihrer Sprecher relevant. Denn Sprache ist beispielsweise eines der Merkmale, die im Diskriminierungsverbot des Grundgesetzes ausdrücklich genannt werden.

So erfreulich das amtliche Interesse an Sprachstatistiken aus linguistischer Sicht ist, muss man zugleich konstatieren, dass es mutmaßlich nicht in erster Linie sprachwissenschaftliche Überlegungen waren, die zur Aufnahme dieser Frage in den Mikrozensus geführt haben. Der wesentliche Auslöser dürfte im aktuellen soziopolitischen Geschehen zu suchen sein. Insbesondere mit dem Jahr 2015 und der sogenannten „Flüchtlingskrise“ in Deutschland werden wieder Fragen über die Zusammensetzung der Bevölkerung, deren nationale Identität und deren Sprache (neu) verhandelt. Die Implementierung einer Sprachfrage im deutschen Mikrozensus im Jahr 2017 muss zweifellos in diesem Kontext gesehen werden.

2. Die Sprachfrage im deutschen Mikrozensus

2.1 Der Mikrozensus und die darin gestellten Fragen

Der deutsche Mikrozensus erhebt regelmäßig Daten zur in Deutschland lebenden Bevölkerung. Einmal im Jahr wird dafür eine kleine, aber für die deutsche Wohnbevölkerung repräsentative Stichprobe befragt. Der Mikrozensus ist (neben dem Zensus, der entsprechend der EU-Richtlinien im Zehnjahresrhythmus stattfinden soll) die umfangreichste Befragung privater Haushalte in Deutschland. Befragt wird eine Stichprobe von etwa 810.000 Menschen in 400.000 privaten Haushalten. Üblicherweise werden die ausgewählten Haushalte durch ehrenamtliche Interviewer persönlich besucht und befragt.

Der Mikrozensus umfasst inzwischen einen sehr umfangreichen Fragekatalog: 2018 enthielt der Fragebogen 220 Fragen; 1957 waren es lediglich 52 (vgl. Anders 2018). Für die meisten Fragen im Fragebogen gilt eine gesetzlich verankerte Auskunftspflicht. Einige wenige Fragen sind freiwillig; dies ist jeweils explizit vermerkt, außerdem sind die entsprechenden Fragen im Fragebogen farblich abgesetzt. Die Fragen im Mikrozensus 2018 beziehen sich auf folgende Themen (in der Reihenfolge des Fragebogens): familiäre Situation im Haushalt (mit 19 Fragen), Beruf, Erwerbstätigkeit bzw. Arbeitssuche (100 Fragen), Kindertagesbetreuung (4 Fragen), Ausbildung (27 Fragen), Staatsangehörigkeit und Aufenthaltsdauer (27 Fragen), Einkommen und Lebensunterhalt (17 Fragen), Informations- und Kommunikationstechnologien im Haushalt (3 Fragen), Wohnsituation des Haushalts (22 Fragen) und Beteiligung an der Erhebung (2 Fragen).

Die grobe Struktur der Fragen sieht vor, dass die meisten Fragen von jeweils bis zu fünf Personen beantwortet werden können (ggf. müssen weitere Bögen verwendet werden). Dafür ist für jede dieser fünf Personen jeweils eine Spalte mit Antwortfeldern ((Freifeld-) Kästchen oder Boxen zum Ankreuzen) vorgesehen, die wie folgt beschriftet sind: „1. Person, 2. Person, 3. Person, 4. Person, 5. Person“. Einige Fragen richten sich nicht an alle Befragten; in diesem Fall gibt es jeweils farbliche Sprungverweise im Fragebogen. Soll die Frage nur von einer Person pro Haushalt beantwortet werden, dann werden nicht jeweils fünf Spalten mit entsprechendem Kästchen oder Freifeld angeboten, sondern lediglich eine Spalte.¹ In der Informationsbroschüre für die Befragten des Mikrozensus gibt es einen Hinweis darauf, wie genau die Abfrage für die verschiedenen Haushaltsmitglieder funktioniert: „Es ist nicht unbedingt nötig, dass alle Haushaltsmitglieder bei der Befragung anwesend sind. Die Antworten können von einer volljährigen Person stellvertretend für alle Haushaltsmitglieder erteilt werden.“ (vgl. Statistische Ämter des Bundes und der Länder (Hg.) 2018, S. 11). Das heißt also, dass nicht nur Fragen, die explizit lediglich nur an eine Person auf Haushaltsebene gestellt werden, durch nur eine Person beantwortet werden können, sondern im Prinzip gilt das auch für alle Fragen, die auf Personenebene eigentlich von den einzelnen Haushaltsmitgliedern beantwortet werden sollten.

Die Fragen im Mikrozensus sehen im Prinzip drei Antworttypen vor: im ersten Fall muss die entsprechende Antwort angekreuzt werden, im zweiten Fall muss ein Zahlenwert oder ein Kürzel eingegeben werden,² oder aber es muss im dritten Fall in einem Freifeld eine

¹ Im Fragebogen 2018 werden z. B. auf Seite 3 dezidiert Fragen auf Haushaltsebene gestellt (vgl. Frage 2 oder Frage 3). Für diese Fragen ist lediglich eine Antwort vorgesehen.

² Z. B. die Anzahl der Kinder, eine Jahreszahl oder ein Kürzel, das für einen Monat steht; dabei kann es sich auch um Ziffern oder Schlüssel handeln, die bei der Frage in einer Liste angeboten werden.

Antwort frei und selbst formuliert werden.³ Die bei manchen der Fragen angebotenen Listen mit vorgegebenen Antwortmöglichkeiten unterscheiden sich stark hinsichtlich ihrer Länge, einige sind sehr knapp gehalten, andere dagegen sehr ausführlich.⁴

Außerdem gibt es zwei Antworttypen bezüglich der Anzahl der zugelassenen Antworten: Entweder wird nur eine Antwort zugelassen oder aber mehrere (Einfachantwort vs. Mehrfachantworten). Für die Mehrheit der Fragen im Mikrozensus sollen und dürfen die Befragten nur eine Antwort geben. Das gilt auch für Fragen, bei denen mehrere Antworten denkbar und durchaus plausibel wären. Eine Mehrfachantwort wird bei diesen Fragen durch einen entsprechenden, kursiv gesetzten, Hinweis eingeschränkt: „Wenn mehrere Gründe zutreffen, geben Sie bitte den Hauptgrund an“ oder „Bitte wählen Sie den Hauptgrund aus.“⁵ Allerdings gibt es vereinzelt im Mikrozensus auch Fragen, bei denen ausdrücklich mehrere Antworten möglich sind.⁶

2.2 Die Position der Sprachfrage und ihre Formulierung

Die Frage zur Sprache der Bevölkerung findet sich erstmalig im Mikrozensusfragebogen 2017. Sie wird in den Folgejahren 2018 und 2019 wiederholt. Im Fragebogen 2017 (und ebenso in den Fragebögen für 2018 und 2019) befindet sie sich in der zweiten Hälfte im Block mit Fragen zu „Staatsangehörigkeit und Aufenthaltsdauer“ (ab S. 50 des Fragebogens). Diese Platzierung deutet darauf hin, dass es hier nicht in erster Linie darum geht, Sprache als soziodemographisches Merkmal zu erheben, sondern eher als Stellvertreterkategorie zur Messung von kultureller Integration und als einen weiteren Baustein für das Konstrukt „Migrationshintergrund“. Viele der Fragen in diesem Abschnitt müssen nicht von allen Befragten beantwortet werden. Anhand von Filtern, die über die Beantwortung der jeweils vorherigen Frage bestimmt werden, werden die Befragten durch die zu beantwortenden Fragen geleitet. Zehn der 26 Fragen, darunter auch die zur Sprache, gehen allerdings ungefiltert an alle Befragten. Die Sprachfrage ist dabei je nach Filterung die zweite oder fünfte Frage. Über die anderen Fragen werden die Staatsangehörigkeit, der Typ von Staatsangehörigkeit, der Geburtsort und die Aufenthaltsdauer in Deutschland erhoben, und (je nach Filter) jeweils auch von den Eltern der Befragten. Diese Fragen bilden die Grundlage zur Einteilung der Befragten in die artifiziell generierte Kategorie „Migrationshintergrund“ mit ihren diversen Unterkategorien (vgl. z. B. Will 2016, 2018).

Die Frageformulierung lautet: „Welche Sprache wird in Ihrem Haushalt vorwiegend gesprochen?“; gesetzt ist sie wie folgt (vgl. Abb. 1):

³ Vgl. im Fragebogen 2018 Fragen 35 und 40.

⁴ Vgl. im Fragebogen 2018 etwa Liste 5 zu Frage 47, Liste 11 oder Liste 12.

⁵ Vgl. im Fragebogen 2018 Frage 26.

⁶ So werden etwa bei der Frage danach, warum man keine bezahlte Tätigkeit gesucht hat (vgl. im Fragebogen 2018 Frage 102), den Fragen nach der Art der Arbeitssuche (vgl. im Fragebogen 2018 Fragen 110 und 111) und den Fragen nach spezifischen Arten von Einkünften (vgl. im Fragebogen 2018 Fragen 178, 179 und 180) mehrere Antworten zugelassen. Diese Fragen, bei denen es sich jeweils um Antworten zum Ankreuzen handelt, sind ergänzt um den kursiv gesetzten Zusatz „Kreuzen Sie bitte alles Zutreffende an.“ bzw. Varianten davon.

159 Welche Sprache wird in Ihrem Haushalt vorwiegend gesprochen? 1. Person

Deutsch 01

Nicht Deutsch, und zwar ...

... Arabisch 02

... Englisch 03

... Französisch 04

... Italienisch 05

... Polnisch 06

... Russisch 07

... Spanisch 08

... Türkisch 09

... eine sonstige europäische Sprache 10

... eine sonstige afrikanische Sprache 11

... eine sonstige asiatische Sprache 12

... eine sonstige Sprache 13

154 Bei Frage 154 ist es ausreichend, wenn eine Person stellvertretend für alle Personen, die gemeinsam einen Haushalt in der Wohnung bilden, die Frage beantwortet. Für alle weiteren Personen im Haushalt geht es mit Frage 155 weiter. Welche Sprache wird in Ihrem Haushalt vorwiegend gesprochen? 1. Person

Deutsch 01

Nicht deutsch, sondern ...

... arabisch 02

... englisch 03

... französisch 04

... italienisch 05

... polnisch 06

... rumänisch 14

... russisch 07

... spanisch 08

... türkisch 09

... eine sonstige europäische Sprache 10

... eine sonstige afrikanische Sprache 11

... eine sonstige asiatische Sprache 12

... eine sonstige Sprache 13

Abb. 1: Die Sprachfrage im deutschen Mikrozensus 2017 (links) und 2019 (rechts)⁷

Die Antwortmöglichkeiten werden in einer geschlossenen Liste im Anschluss an die Frage präsentiert. Diese Liste ist deutlich in zwei Teile gegliedert: Die erste Nennung ist „Deutsch“, alle weiteren Nennungen werden unter dem Titel „Nicht Deutsch, und zwar ...“ bzw. „Nicht deutsch, sondern ...“ gruppiert. Als nicht deutsche Sprachen können angegeben werden: Arabisch, Englisch, Französisch, Italienisch, Polnisch, (Rumänisch), Russisch, Spanisch, Türkisch, eine sonstige europäische Sprache, eine sonstige afrikanische Sprache, eine sonstige asiatische Sprache und – als letztes – eine sonstige Sprache. Das Set an Antworten beinhaltet keine Möglichkeit, eine eigene Antwort zu formulieren. Bei der Sprachfrage müssen die Befragten also aus diesen vorgegebenen Antworten eine ihnen adäquat erscheinende Antwort wählen. Für die Antwort ist allerdings nur eine Spalte mit einem Antwortkästchen vorgesehen, diese ist mit dem Zusatz „1. Person“ betitelt. Die Information über die vorwiegend im Haushalt gesprochene Sprache ist also lediglich auf Haushaltsebene und nicht auf individueller Ebene vorgesehen.⁸ Das bedeutet, eine einzige Person beantwortet stellvertretend für alle anderen im Haushalt lebenden Personen die Sprachfrage. Wer genau diese einzige Person ist, die die Angabe auf diese Frage machen darf, ist nicht klar; vermutlich liegt dieser Konstruktion das gedankliche Konzept des „Haushaltsvorstands“ zugrunde. Die Sprachfrage unterscheidet sich dadurch stark von allen anderen Fragen dieses Themenblocks, für die jeweils fünf Antwortspalten vorgesehen sind. Die Einschränkung auf die Haushaltsebene ist nur dann sinnvoll, wenn man davon ausgeht, dass die individuellen Angaben der einzelnen Haushaltsmitglieder ohnehin übereinstimmen würden.

Die Sprachfrage wird im Mikrozensus 2018 und 2019 (siehe Abb. 1) in mehr oder weniger identischer Form wiederholt. Es gibt lediglich kleinere Änderungen: 2018 wurde ein expliziter Hinweis auf den Filter hinzugefügt: „Bei Frage 155 ist es ausreichend, wenn eine Person stellvertretend für alle Personen, die gemeinsam einen Haushalt in der Wohnung bilden, die Frage beantwortet. Für alle weiteren Personen im Haushalt geht es mit

⁷ www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/HaushalteMikrozensus/HaushalteFamilien2010300177004.pdf?__blob=publicationFile (Stand: 4.2.2019) und www.statistik-berlin-brandenburg.de/datenerheb/dateien/mz.pdf (Stand: 4.2.2019).

⁸ Ähnlich verhält es sich z. B. bei der Frage nach dem Haushaltseinkommen (vgl. Frage 182).

Frage 156 weiter.“ Der Zusatz macht die Einschränkung der Beantwortung dieser Frage auf lediglich eine Person pro Haushalt explizit. Möglicherweise wurde der Sachverhalt in der Version von 2017 als nicht hinreichend klar empfunden, sodass eine Klarstellung notwendig war (dazu siehe Abschn. 3.2, insbesondere zum Anteil der fehlenden Angaben). 2019 wurde die Frage wie 2018 wiederholt, allerdings wurde eine Antwortoption hinzugefügt, nämlich die Einzelsprache Rumänisch.

2.3 Die Sprachfrage als Stellvertreterfrage

Zur Entstehung und Intention der Sprachfrage im deutschen Mikrozensus lassen sich nur einige wenige Metainformationen finden. Einige davon stehen im Mikrozensusgesetz, das die Grundlage für die Sprachfrage bildet, und dem offiziellen Entwurfstext dazu; weitere Hinweise gibt es im Bericht zu den Ergebnissen des Mikrozensus (vgl. Statistisches Bundesamt (Hg.) 2018). Zunächst zum Gesetz: Das neue Mikrozensusgesetz (MZG bzw. Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und die Arbeitsmarktbeteiligung sowie die Wohnsituation der Haushalte) beruht auf einem Gesetzesentwurf der Bundesregierung aus dem Sommer 2016. Am 7. Dezember 2016 wurde das Gesetz beschlossen, am 1. Januar 2017 trat es in Kraft.⁹ Dieses Mikrozensusgesetz ist nun als erstes seiner Art ohne Befristung gültig. In diesem Gesetz wird beschrieben, dass die Sprachfrage ein Bestandteil des Fragenteils zu Staatsangehörigkeit und Migrationshintergrund ist. Die Sprachfrage ist also explizit nicht Bestandteil der im engeren Sinne demographischen Angaben wie etwa Geschlecht, Geburtsjahr und Familienstand. Das ist durchaus bemerkenswert.

Aufschlussreich ist diesbezüglich auch die Eigendarstellung der im Mikrozensus ermittelten Themen. In der Informationsbroschüre zum Mikrozensus werden als Themenbereiche die folgenden genannt: „Angaben zur Person (z. B. Geschlecht, Geburtsjahr, Familienstand, Staatsangehörigkeit); Erwerbstätigkeit, Beruf, Arbeitsuche; Kindertagesbetreuung, Schule, Studium; Aus- und Weiterbildung; Lebensunterhalt, Einkommen; Altersvorsorge; Wohnsitz und Erwerbsbeteiligung“ (Statistische Ämter des Bundes und der Länder (Hg.) 2018, S. 15). In dieser Aufzählung ist der Themenbereich „Staatsangehörigkeit und Aufenthaltsdauer“, immerhin mit 27 Fragen einer der größeren Frageblöcke, beinahe komplett ausgespart. Lediglich auf das Merkmal „Staatsangehörigkeit“ wird verwiesen. Dieses wird allerdings in den demographischen Angaben zur Person gelistet, wo es im Mikrozensusfragebogen gar nicht steht.

Der in den Gesetzestexten genannte Kontext und die thematische Einbettung der Sprachfrage machen deutlich, worum es geht und worum es nicht geht: Es geht nicht um die deutschsprachigen Deutschen, also die mutmaßliche Mehrheit der in Deutschland lebenden Bevölkerung, und um das Erfassen der (zum Teil auch komplexen) Sprachrepertoires dieses Personenkreises, sondern es geht um jene Personen, die einen sogenannten „Migrationshintergrund“ haben. Deutlich wird das auch in den Erläuterungen zum entsprechenden Gesetzesentwurf (18/9418, S. 32):

Die Zuwanderung aus dem Ausland und die Integration von Migranten ist ein wichtiges politisches Thema. Die im Rahmen des Mikrozensus erhobenen Angaben sind wesentlicher Bestandteil der Integra-

⁹ Vgl. Bundesgesetzblatt Jahrgang 2016 Teil I Nr. 59, ausgegeben zu Bonn am 13. Dezember 2016 (Online: www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&start=%252F%252F%252A%255B%2540attr_id=%2527bgbl116s0390.pdf%2527%255D#__bgbl__%2F%2F%255B%40attr_id%3D%27bgbl116s2826.pdf%27%5D__1551176929351, Stand: 26.2.2019).

tionsberichterstattung. Die Angaben zur Einbürgerung ermöglichen Rückschlüsse auf die formale Integrationsbereitschaft von Migranten. [...] Die Erfassung der im Haushalt vorwiegend gesprochenen Sprache ergänzt die Informationen zum Migrationshintergrund und ist für die Einschätzung verschiedener Dimensionen der Integration von Bedeutung. Es werden differenzierte Analysen zum Stand der Integration ermöglicht. Insbesondere die kulturelle Integration steht in enger Verbindung mit der im Haushalt gesprochenen Sprache. Die Aufnahme des Merkmals ermöglicht Vergleiche mit anderen Statistiken, u. a. der Schulstatistik sowie der Kinder- und Jugendhilfestatistik. (Hervorhebungen A. A.)

Der Text macht den Fokus der Sprachfrage explizit: Es geht um Fragen der Integration. Genauer geht es darum, Bestandteile der Integrationsberichterstattung zu erheben, d. h. die formale Integrationsbereitschaft. Die erhobene Sprache soll zur Ergänzung der Information zum Migrationshintergrund dienen, der selbst eine eher opake Kategorie ist (vgl. Will 2016, 2018). Am Sprachgebrauch soll demnach eine Dimension von Integration abgelesen werden. An diesem Text sind insbesondere die beiden letzten Sätze hervorzuheben: Im vorletzten Satz steht die Behauptung, dass kulturelle Integration und die im Haushalt gesprochene Sprache in enger Verbindung miteinander stehen. Diese Behauptung wird weder belegt noch epistemisch abgetönt (das liegt sicherlich auch an der Textsorte). Der Behauptung entspricht, dass die kulturelle Integration von Migranten oder Menschen mit Migrationshintergrund an der im Haushalt, also der zuhause gesprochenen Sprache gemessen werden soll. Letztlich soll also die kulturelle Integration erhoben werden. Da dies jedoch nicht direkt vorgenommen werden kann (bislang weiß zumindest niemand, wie das gehen sollte), soll die Sprache als Stellvertreterkategorie, auch Proxy genannt, fungieren; also als ethnokultureller Indikator. Nun ist die aufgeführte Behauptung, obgleich sie zunächst naheliegend erscheinen mag,¹⁰ aus sprachwissenschaftlicher Sicht durchaus diskussionswürdig. Es gibt keine klare Korrelation zwischen der zuhause gesprochenen Sprache und der kulturellen Integriertheit. Das heißt, die Annahme, dass nur derjenige, der zuhause die Mehrheitssprache spricht, kulturell gut in die Mehrheitsgesellschaft integriert ist, greift zu kurz. Hinzu kommt, dass eigentlich nicht ganz klar ist, was unter dem Konzept von kultureller Integration verstanden werden soll. Relativ klar ist jedenfalls, dass die sprachliche Wirklichkeit eine deutlich komplexere ist, als es eine solch einfache Gleichung abbilden könnte. Selbst für einsprachig deutsche Haushalte, also für die wahrscheinlich am häufigsten vorkommende Form, ist fraglich, ob diese Annahme immer zutrifft. Für mehrsprachige Haushalte ist es mit Sicherheit komplizierter. Der Behauptung dürfte eine laienlinguistisch übliche Einsprachigkeitsperspektive bzw. Einsprachigkeitsideologie zugrunde liegen (vgl. z. B. Gogolin 1997). Als Einsprachigkeitsideologie kann man eine durch eine einsprachige Sicht gefärbte Perspektive auf die komplexe, und eben auch mehrsprachige, (Sprach-)Realität bezeichnen. Demnach wäre Einsprachigkeit der Normalfall oder sogar die Norm; jegliche Abweichung davon – also jede Form von Mehrsprachigkeit – wird entweder gar nicht erst wahrgenommen oder nur als markierter Sonderfall interpretiert. In einer solchen Einsprachigkeitsperspektive können Mehrsprachigkeit und ihre Vielfalt offensichtlich nur schwer konzeptualisiert werden. Ein mehrsprachiges Repertoire etwa, das sich auf Kommunikationsdomänen verteilt, ist dann kaum vorstellbar. Gleiches gilt für Konstellationen, bei denen zuhause mehrere Sprachen gleichermaßen oft gesprochen werden.

¹⁰ Es soll hier selbstverständlich nicht in Frage gestellt werden, dass es einen Zusammenhang zwischen Kultur und Sprache gibt (vgl. u. a. Lameli 2013). Allerdings weiß man noch nicht genau, wie sich Sprache, Kultur und Identität von Mehrsprachigen miteinander verhalten, insbesondere unter dem Aspekt der Integration.

Die Sprache soll im Mikrozensus also als Stellvertreterkategorie erhoben werden und zur Messung eines anderen Phänomens dienen, nämlich der kulturellen Integration. Eine solche Proxy-Funktion hatte Sprache bemerkenswerterweise bereits in den Volksbefragungen des 19. Jahrhunderts.¹¹ Zu dieser Zeit etablierte sich die Bevölkerungsstatistik, und die Methode der Volksbefragung wurde verfeinert. Der preußische Statistiker Böckh beispielsweise plädierte damals für das Erfragen der Muttersprache, da diese in seinem Konzept der Nationalitätenstatistik eine tragende Rolle spielte. Seiner Meinung nach stand die Muttersprache im einem engen Zusammenhang mit der nationalen Identität der Bevölkerung. Von letzterer aber glaubte er, dass sie zu labil sei, um sie direkt zu erfragen. Aus diesem Grund wurde die Muttersprache als indirekte Kategorie benötigt (siehe dazu auch Leuschner 2004). Dieser Ansatz scheint gar nicht so weit von der Intention entfernt zu sein, die offensichtlich hinter der aktuellen Sprachfrage steht.

Der letzte Satz des Ausschnitts aus dem Gesetzestext schließlich verweist auf einen weiteren möglichen Aspekt: die Variable „Sprache“ soll anscheinend mit anderen Statistiken in Verbindung gesetzt werden. Das bedeutet also, dass Sprache bereits in anderen Statistiken ein operatives Merkmal ist, offensichtlich etwa in der Schulstatistik und in der Kinder- und Jugendhilfestatistik. Die Sprache der Schüler und zum Teil auch der Eltern wird in der Schulstatistik insbesondere als Migrationsmerkmal für die Definition des Migrationshintergrundes eingesetzt. Betrachtet man nun allerdings, wie Sprache in den Schulstatistiken erhoben wird, wird sehr schnell klar, dass ein Vergleich kaum möglich sein wird (siehe hierzu die detaillierte Analyse von Kemper 2015, 2017). Das liegt daran, dass sich die Schulstatistik in Deutschland vor allem durch ihre Heterogenität auszeichnet: Die Schulstatistiken der einzelnen Bundesländer unterscheiden sich nicht nur untereinander, sondern auch von den Vorgaben der Kultusministerkonferenz. Lediglich dreizehn Bundesländer erfassen überhaupt die Sprache der Schüler, jedoch in jeweils sehr verschiedener Form (vgl. Kemper 2017, S. 147 ff. und 153 ff.). Die Unterschiede sind vielfältig: Sie finden sich etwa in der Frageformulierung und in den vorgegebenen Antworten. Sicher ist, ein Vergleich der Angaben zu Sprache ist unmöglich. Daraus folgt auch, dass es eine nationale Schulstatistik mit Informationen zu Sprache und Migrationshintergrund aktuell nicht gibt und auch in absehbarer Zeit nicht geben wird. Ein sinnvoller Vergleich mit Daten aus dem Mikrozensus ist also ebenfalls nicht möglich.

3. Kritik an der Sprachfrage des Mikrozensus

3.1 Analyse der Sprachfrage

Jede Frage stellt eine Variante aus einem Set an möglichen Varianten dar. Auch die Sprachfrage des Mikrozensus ist eine von vielen Möglichkeiten, die Sprache der Bevölkerung zu erheben.¹² Eine Analyse der Sprachfrage im deutschen Mikrozensus zeigt jedoch, dass die gestellte Fragenvariante keine besonders glückliche Wahl darstellt. Das liegt im Wesentlichen an folgenden kritischen Aspekten: 1) der Formulierung der Fragestellung mit der Einschränkung auf lediglich eine mögliche Antwort, 2) der Beantwortung durch nur eine Person stellvertretend pro Haushalt, 3) der Auswahl der Antwortmöglichkeiten, 4) dem

¹¹ Etwa in Österreich, Ungarn, Preußen und Russland (vgl. Arel 2002, S. 96).

¹² Es ist nicht klar, wie das Statistische Bundesamt zu genau dieser Frageformulierung gelangt ist. Üblicherweise werden im Vorhinein verschiedene Versionen einer Frage pilotiert (so etwa in England, vgl. Sebba 2017); für den deutschen Mikrozensus ist das nicht nachvollziehbar.

Fehlen einer offenen Antwortmöglichkeit und 5) der Position im Migrationsblock (i. e. im Themenblock mit Fragen zu „Staatsangehörigkeit und Aufenthaltsdauer“).¹³

1) Die Formulierung der Frage macht es unmöglich, Mehrsprachigkeitskonstellationen abzubilden. Das ist gerade deshalb bemerkenswert, da insbesondere Migranten, also Mehrsprachige, im Fokus der Frage stehen. Die Formulierung der Frage bezieht zwar im Prinzip die Möglichkeit ein, dass in einem Haushalt mehrere Sprachen gesprochen werden. Möglich ist jedoch lediglich die Nennung einer einzigen Sprache. Die Form der Antwort – eine Einfachantwort – entspricht dem vorherrschenden Muster der Fragen im Mikrozensus mit einer sehr starken Präferenz für Einfachantworten; Fragen, bei denen mehrere Antworten möglich sind, gibt es nur wenige (siehe oben Abschn. 2.1). Ein solcher Antworttyp wäre sicherlich gerade für die Sprachfrage wünschenswert gewesen. Die Einschränkung auf nur eine Antwort hat nämlich zur Folge, dass die Sprachrepertoires von mehrsprachigen Menschen nicht adäquat abgebildet werden können. Das führt allerdings dazu, dass diese Sprachfrage keinerlei Ergebnisse zum Anteil der mehrsprachigen Bevölkerung in Deutschland erbringt. Prinzipiell hätte eine Sprachfrage genau dazu genutzt werden können (siehe unten Abschn. 4.2 und 5). Die Möglichkeit dazu wurde im Mikrozensus allerdings vertan.

Nun ist in Deutschland wahrscheinlich die Mehrheit der Bevölkerung einsprachig Deutsch. Es gibt aber eben auch einige Menschen, die mehr als eine Muttersprache haben, also mehrsprachig sind. Die Sprachen dieser Menschen verteilen sich vermutlich auf vielfältig unterschiedliche Weise auf die verschiedenen sprachlichen Kommunikationsdomänen, d. h. in privaten Domänen (wie zuhause) wird möglicherweise eine andere Sprache gesprochen als in öffentlichen Domänen (wie bei der Arbeit).¹⁴ Außerdem gibt es auch mehrsprachige Haushalte, in denen zuhause mehrere Sprachen gesprochen werden, möglicherweise sogar etwa gleich häufig. Das bedeutet aber nicht zwangsläufig, dass alle Haushaltsmitglieder alle diese Sprachen tatsächlich gleich häufig sprechen. Die Verwendung der Sprachen kann in einem mehrsprachigen Haushalt für jedes Haushaltsmitglied durchaus unterschiedlich sein. So ist es etwa sehr wahrscheinlich, dass in mehrsprachigen Haushalten die gesprochenen Sprachen generationsbedingt verschieden sind. Dadurch aber, dass die Sprachfrage auf die Nennung nur einer Sprache beschränkt ist, muss eine Entscheidung darüber getroffen werden, welche aller möglicherweise im Haushalt vorhandenen Sprachen die „vorwiegend“ gesprochene Sprache sein soll. Eine Abbildung von mehrsprachigen sozialen Realitäten ist bei der Beantwortung dieser Sprachfrage nicht möglich. Die Sprachfrage erzwingt die Reduktion auf eine virtuelle Einsprachigkeit: mehrdimensionale, mehrsprachige Sprachrepertoires werden zwangsläufig auf eindimensionale, einsprachige Konstellationen eingeebnet und reduziert. Die Ergebnisse der Sprachfrage können also die gesamte sprachliche Realität gar nicht abbilden, sondern nur einen arbiträren Teil davon. Daraus folgt auch, dass diese Teilinformationen eine nur sehr eingeschränkte Gültigkeit aufweisen. Es ist offensichtlich, dass dies gerade bei einer Frage, die auf die mehrsprachige Bevölkerung abzielt und über deren Sprachverwendung als Stellvertreterkategorie auf ihre kulturelle Integration schließen möchte, besonders ungünstig ist.

¹³ Eine detaillierte Auswertung der Frage im Mikrozensus findet sich in Adler (2018).

¹⁴ Darauf deuten etwa Daten aus einer Repräsentativerhebung von 2008 hin; vgl. die Auswertung dieser Daten in Adler (2018) in Abschnitt 4.

2) Zu dieser Einschränkung kommt hinzu, dass die Frage lediglich auf Haushaltsebene erhoben wird. Die im Prinzip sehr große Stichprobe des Mikrozensus von etwa 810.000 Befragten wird dadurch einerseits auf die deutlich kleinere Ebene der Haushalte mit 400.000 eingeschrumpft. Andererseits führt das zu inhaltlich problematischen Verkürzungen bei mehrsprachigen Haushalten. Die Sprachverwendung und deren Wahrnehmung durch die individuellen Haushaltsmitglieder können durchaus verschieden sein. Für Einpersonenhaushalte ist dieser Aspekt im Prinzip unproblematisch.¹⁵ Allerdings kann für diese Haushalte sowieso bezweifelt werden, dass es sinnvoll ist, nach einer vorwiegend im Haushalt gesprochenen Sprache zu fragen. Für mehrsprachige Mehrpersonenhaushalte gilt: Das Zulassen lediglich einer individuellen Angabe stellvertretend für alle potenziell möglichen Varianten der sprachlichen Realität ist notwendigerweise arbiträr. Die Relevanz der Antwort wird dadurch also doppelt eingeschränkt: einmal durch die Reduktion der Antworten auf eine Angabe und außerdem durch die Beschränkung der Antwort auf Haushaltsebene.

3) Ein weiterer Mangel liegt in den vorgegebenen Antwortmöglichkeiten. Es wird ein relativ knappes, eingeschränktes Set an möglichen Antworten angeboten. Die erste Antwortmöglichkeit ist „Deutsch“. Diese ist ganz offensichtlich als Default markiert: Sie wird als erstes genannt und auch von allen folgenden Nennungen dadurch abgegrenzt, dass letztere gruppiert und zusätzlich mit dem Titel „Nicht Deutsch“ versehen werden. Der zweite Teil der „nicht deutschen“ Antwortmöglichkeiten besteht einerseits aus Einzelsprachen und andererseits aus generischen Sammelkategorien. Zunächst zu letzteren: angegeben werden die drei allgemeineren Sammelkategorien „eine sonstige europäische Sprache“, „eine sonstige afrikanische Sprache“ und „eine sonstige asiatische Sprache“. Die Sammelkategorien sind anscheinend geografisch geordnet.¹⁶ Die Liste wird abgeschlossen mit der generischen bzw. unspezifischen Antwortmöglichkeit „eine sonstige Sprache“ (als ultimative Restkategorie, vgl. Adler 2018, S. 7 und 9).

Die Einteilung und Abgrenzung dieser generischen Kategorien ist aber nur scheinbar einfach. Tatsächlich eignen sich geografische Kategorien nicht besonders gut zur Klassifizierung von Sprachen. So ist beispielsweise nicht klar, wie Sprachen eingeordnet werden sollen, die auf mehreren Kontinenten gesprochen werden. Portugiesisch beispielsweise wird nicht nur in Portugal gesprochen, sondern auch unter anderem in Angola (Afrika) oder – mit erheblich größerer Sprecherzahl – in Brasilien (Südamerika). Sollte Portugiesisch dann als „(sonstige) europäische Sprache“ bezeichnet werden oder als „(sonstige) afrikanische Sprache“ oder gar als „sonstige Sprache“?¹⁷ Schwierig ist beispielsweise auch die Einordnung des Türkischen: Handelt es sich um eine europäische oder um eine asiatische Sprache? Hier hilft allein, dass es als Einzelnennung bereits vorgegeben ist. Für

¹⁵ Laut der Ergebnisse des Mikrozensus 2017 sind 41,8% der privaten Haushalte in Deutschland (insgesamt 41.304.000) Einpersonenhaushalte; daneben gibt es 33,5% 2-Personenhaushalte, 12,0% 3-Personenhaushalte, 9,3% 4-Personenhaushalte und 3,4% Haushalte mit 5 und mehr Personen. Die Anzahl der Einpersonenhaushalte weist seit 1962 eine im Prinzip steigende Tendenz auf. (vgl. www.destatis.de/DE/ZahlenFakten/Indikatoren/LangeReihen/Bevoelkerung/lrbev05.html;jsessionid=5126D2C4D75E0C37820B695A81FDB86E.InternetLive1, Stand: 14.1.2019).

¹⁶ Im Prinzip könnte es sich auch um eine Einordnung nach Sprachfamilien handeln.

¹⁷ Prinzipiell könnte es sogar als „(sonstige) asiatische Sprache“ bezeichnet werden, wenn man seinen offiziellen Status in Macau als Grundlage heranziehen würde.

Sprecher des Kurdischen bleibt eine Einordnung weiterhin schwierig. Unklar oder mindestens unglücklich ist außerdem die Einordnung der Minderheitensprachen Deutschlands. Dänisch beispielsweise muss als anerkannte Minderheitensprache in eine der übrigen Kategorien „sonstige europäische Sprache“ eingeordnet werden. Was ist mit Niederdeutsch: Wird es von seinen Sprechern zu den europäischen Sprachen gezählt oder dem Deutschen zugesprochen? Ähnlich schwierig ist es für die deutsche Gebärdensprache: Deutsch, „sonstige europäische Sprache“ oder „sonstige Sprache“? Diese Einordnungsschwierigkeiten gelten bereits für die Beantwortung der Frage, für die Auswertung der Ergebnisse kommen sie erneut zum Tragen. Denn die einzelnen Antworten, die die generischen Sammelkategorien bilden, können im Anschluss nicht mehr aufgeschlüsselt werden; sie sind für die Auswertung sozusagen verloren. Das kommt für die hier vorhandenen Sammelkategorien zu ihrer nicht vorhandenen Trennschärfe hinzu. Für die Ergebnisse gilt dementsprechend, dass die Aussagekraft und Gültigkeit dieser Sammelkategorien sehr fragwürdig sind.

Die vorgegebenen acht Einzelsprachen sind folgende: Arabisch, Englisch, Französisch, Italienisch, Polnisch, Russisch, Spanisch und Türkisch; ab 2019 kommt Rumänisch als weitere Einzelsprache hinzu. Die Sprachen sind alphabetisch gereiht, lassen sich aber prinzipiell zwei Typen zuordnen, den europäischen Bildungs- und Prestigesprachen einerseits und den Migrantensprachen andererseits.¹⁸ Über die genannten und entsprechend die eben nicht genannten Sprachen kann man vorhandene gesellschaftspolitische Muster und Sprachideologien erkennen (welche Sprachen sind es wert, genannt zu werden, und welche nicht). Man könnte auch auf Bourdieu (1991) verweisend in den genannten Sprachen und deren Anordnung im Antwortset Machtverhältnisse gespiegelt sehen (vgl. auch Busch 2010, 2016 für eine entsprechende Analyse der Sprachfrage im österreichischen Zensus). Demnach scheinen beispielsweise die als Einzelsprachen gelisteten Sprachen eine gewisse Relevanz zu haben; andere Sprachen dagegen nicht, wie etwa Griechisch, die Sprachen des Balkans oder überhaupt die in Deutschland anerkannten autochthonen Minderheitensprachen (diese freilich mit sehr kleinen Sprecherzahlen).

4) Die Frage anhand der vorgesehenen Antwortmöglichkeiten zu beantworten, ist also nicht trivial: Das liegt insbesondere am begrenzten Set an benennbaren Einzelsprachen und den unklaren Sammelkategorien. Für die Beantwortung muss jedoch eine Kategorie ausgewählt werden, unabhängig davon, wie adäquat sie der Befragte findet. Das liegt zum einen an der Auskunftspflicht und zum anderen daran, dass die Formulierung einer eigenen Antwort nicht möglich ist. Denn eine offene Antwort dürfen die Befragten bei dieser Frage nicht formulieren. Das entspricht der allgemeinen Präferenz der Fragen im Mikrozensus. Es gibt allerdings auch einige wenige Fragen, die diese Möglichkeit vorsehen (siehe Abschn. 2.1). Eine solche Option bei dieser Frage anzubieten, wäre sicherlich sinnvoll, insbesondere auch deshalb, weil diese Frage seit langem zum ersten Mal wieder gestellt wurde.

¹⁸ Die Sprecherzahl, die für diese Sprachen angenommen werden kann, kann kaum der einzige Grund für diese Sprachenauswahl gewesen sein. Das zeigt ein Blick in das Ausländerzentralregister (das allerdings nicht Sprachen, sondern Staatsangehörigkeiten ausweist und damit in Bezug auf Sprachen in den meisten Fällen grobe Vereinfachungen produziert; vgl. Adler 2018). Demnach müsste es etwa viele Sprecher des Türkischen, aber im Vergleich nur sehr wenige Sprecher des Französischen geben. Und auf dieser Grundlage müssten auch noch eine Reihe weiterer Sprachen in Betracht gezogen werden.

5) All die genannten mangelhaften Aspekte der Sprachfrage im Mikrozensus schränken die Gültigkeit der Ergebnisse um ein Vielfaches ein. Hinzu kommt die oben bereits besprochene Position der Sprachfrage im Migrationsblock. Diese Position produziert bei den Befragten möglicherweise einen bestimmten Erwartungsdruck, sodass sehr wahrscheinlich die soziale Erwünschtheit zum Tragen kommt.

3.2 Ergebnisse der Sprachfrage im Mikrozensus 2017

Das Statistische Bundesamt hat die Ergebnisse zur Sprachfrage im Sommer 2018 in einem Bericht zur Bevölkerung mit Migrationshintergrund veröffentlicht (vgl. Statistisches Bundesamt (Hg.) 2018, S. 484–492). Darin werden in Tabelle 18 die gesamten Ergebnisse zur Sprachfrage auf neun Seiten dargestellt.¹⁹ Sie werden jeweils insgesamt und detailliert nach bestimmten Teilstichproben aufgelistet. Die Angaben beziehen sich bereits auf die gesamte Bevölkerung, d. h. die Mikrozensusergebnisse wurden hochgerechnet auf die Gesamtbevölkerung. Da die Frage auf Haushaltsebene gestellt und beantwortet wurde, ist die Einheit der Ergebnisse jeweils ein Haushalt (d. h. die Ergebnisse beziehen sich nicht auf Personen). Die relevante Bezugsgröße für die genannten Sprachen sind somit 41.304.000 Haushalte. Für diese werden die Ergebnisse jeweils betrachtet nach Migrationsstatus, nach Haushaltsstruktur und nach Herkunftsland. Die entsprechende Tabelle ist relativ komplex. Hinzu kommt, dass aufgrund der vielfach gestaffelten Kreuzungen insgesamt sehr viele Zellen entstehen. Davon sind viele jedoch gar nicht mit Inhalten gefüllt.²⁰ Lediglich knapp über die Hälfte der Zellen (52 %) enthalten Zahlen. Hinzu kommt, dass ein ganzer Spaltentyp bereits für die fehlenden Angaben steht, d. h. im Prinzip sind auch diese als inhaltslos zu bewerten.

Zur Darstellung der Ergebnisse wird hier eine einfachere Übersichtstabelle vorgelegt (vgl. Tab. 1). Darin sind die gesamten Antworten auf die Sprachfrage für alle Haushalte nach drei verschiedenen Aspekten aufgeschlüsselt: zunächst in absoluten Zahlen, dann in zwei verschiedenen prozentualen Anteilen: erstens bezogen auf die Gesamtgröße der Haushalte, zweitens für die Gruppe der nicht deutschen Sprachen bezogen auf die gesamte Anzahl dieser Gruppe.

Die deutliche Mehrheit der deutschen Bevölkerung, nämlich 87 % der gesamten Haushalte, spricht demnach zuhause vorwiegend Deutsch. Eine andere Sprache als Deutsch nennen 9,1 % der Haushalte. Ein sehr großer Anteil, nämlich 3,9 % (1.623.000 Haushalte), macht für diese Sprachfrage allerdings gar keine Angabe. Damit entfällt der höchste Wert nicht auf eine der angebotenen Einzelsprachen, sondern auf diejenigen, die keine Antwort gegeben haben.

¹⁹ Die Ergebnisse gibt es auch im Excel-Format unter www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/MigrationIntegration/Migrationshintergrund2010220177005.xlsx?__blob=publicationFile (Stand: 4.2.2019).

²⁰ Für die Sprachfrage gibt es zwei Typen dieser inhaltslosen Felder. Einerseits gibt es Felder, für die keine Angabe gemacht wird, da der Zahlenwert nicht sicher genug ist, und andererseits gibt es Felder, für die einfach kein Zahlenwert vorhanden ist. In den Tabellen werden folgende Symbole verwendet: / wenn der Wert nicht sicher genug ist und – wenn nichts vorhanden ist (vgl. Statistisches Bundesamt (Hg.) 2018, S. 3). 39 % aller Zellen von Tabelle 18 sind Zellen des ersten Typs, 9 % sind Zellen des zweiten Typs.

Sprache	absolute Häufigkeit	in Prozent	
<i>Deutsch</i>	35.918.000	87,0	
ohne Angabe	1.623.000	3,9	
„nicht Deutsch“	3.763.000	9,1	
davon			innerhalb „nicht Deutsch“
„sonstige europäische Sprache“	782.000	1,9	20,8
<i>Russisch</i>	527.000	1,3	14,0
<i>Türkisch</i>	522.000	1,3	13,9
<i>Polnisch</i>	336.000	0,8	8,9
<i>Arabisch</i>	303.000	0,7	8,0
„sonstige asiatische Sprache“	283.000	0,7	7,5
„sonstige Sprache“	285.000	0,7	7,6
<i>Englisch</i>	216.000	0,5	5,7
<i>Italienisch</i>	173.000	0,4	4,6
„sonstige afrikanische Sprache“	144.000	0,3	3,8
<i>Spanisch</i>	133.000	0,3	3,5
<i>Französisch</i>	61.000	0,1	1,6
d. h. als Einzelsprachen insgesamt	2.271.000	5,5	60,3
d. h. als Sammelkategorien insgesamt	1.494.000	3,6	39,7
insgesamt	41.304.000		

Tab. 1: Ergebnisse der Sprachfrage im deutschen Mikrozensus 2017 (gereiht nach den Angaben in Statistisches Bundesamt (Hg.) 2018)

Auch wenn die absolute Zahl sehr hoch ist, ist der relative Anteil an fehlenden Werten zunächst kein sehr hoher Wert. Da es sich beim Mikrozensus im Prinzip aber um eine Befragung mit gesetzlicher Auskunftspflicht handelt, und die Sprachfrage keine freiwillige Frage war, ist das Ausmaß der fehlenden Angaben doch bemerkenswert. Eine mögliche Erklärung dafür könnte sein, dass die Mehrheit der Fragen im Block „Staatsangehörigkeit und Aufenthaltsdauer“ sich vorrangig an jeweils kleinere Teile der Befragten richtet. Vielleicht wurde die Frage deshalb fälschlicherweise nicht von allen beantwortet. Allerdings sollten die eventuell unübersichtlichen Filter keine solchen Effekte bewirken, da der Mikrozensus in großer Mehrheit durch Interviewer durchgeführt wird. Diese werden geschult und führen den gleichen Fragebogen mehrmals durch. Für die zweite Durchführung der Sprachfrage wurde jedenfalls eine Veränderung im Fragebogen vorgenommen, die die Filterführung explizit macht (siehe oben Abschn. 2.2).

Besonders hoch sind die Ausfälle bei der Sprachfrage in Haushalten, deren Mitglieder einen Migrationshintergrund haben. Das bedeutet also, dass unter den fehlenden Angaben vermutlich einige zu anderen Sprachen gewesen wären. Die Validität der Antworten wird dadurch weiter eingeschränkt. Das gilt insbesondere deshalb, weil die relevanten gültigen Einzelnennungen zu den nicht deutschen Sprachen jeweils kleiner sind als die fehlenden Angaben.

Die häufigste Nennung innerhalb der nicht deutschen Antworten liegt bei lediglich 1,9%; es handelt sich um die Sammelkategorie „eine sonstige europäische Sprache“. Streng genommen verhält es sich auch mit dieser Antwort ähnlich wie mit den fehlenden Angaben: Im Prinzip sind diese Antworten eher wertlos, denn die Details dieser Sammelkategorie können im Nachhinein nicht mehr aufgeschlüsselt werden. Überhaupt machen die Sammelkategorien einen großen Anteil innerhalb der nicht deutschen Antworten aus: zwei Fünftel der nicht deutschen Antworten sind Sammelantworten (1.494.000 Haushalte bzw. 3,6%), die kaum weitere Rückschlüsse erlauben. Bei einigen davon lässt sich zwar eine grobe Spezifizierung vornehmen (europäisch, asiatisch, afrikanisch), bei einem Fünftel dieser Antworten allerdings ist selbst das nicht möglich („eine sonstige Sprache“).

Unter den genannten Einzelsprachen der Kategorie „nicht Deutsch“ werden vor allem die großen Migrantensprachen Russisch, Türkisch, Polnisch und Arabisch genannt. Insgesamt geben aber nur jeweils 1,3% der Haushalte Russisch oder Türkisch als vorwiegend im Haushalt gesprochene Sprache an (i. e. 527.000 Haushalte Russisch und 522.000 Haushalte Türkisch), 0,8% Polnisch (336.000 Haushalte) und 0,7% Arabisch (303.000 Haushalte). Diese Angaben scheinen sehr niedrig. Gerade für die ersten drei Sprachen würde man erwarten, dass sie weit häufiger genannt würden. Alle weiteren vorgegebenen Einzelsprachen (Englisch, Italienisch, Spanisch und Französisch) werden kaum genannt.

4. Varianten von Spracherhebungen

4.1 Mögliche Varianten einer Sprachfrage

Die Frage nach der Sprache der Bevölkerung, wie sie im deutschen Mikrozensus gestellt worden ist, ist eine mögliche Variante neben vielen anderen. Die Analyse der Frage und der Ergebnisse hat allerdings gezeigt, dass die im Mikrozensus gewählte Form nicht besonders gut ist. Prinzipiell gibt es insbesondere hinsichtlich dreier Aspekte verschiedene Möglichkeiten für eine Sprachfrage.²¹ Der erste grundlegende Punkt ist das gewählte Sprachkonzept. Zweitens geht es um die Anzahl von Sprachen, die die Frageformulierung vorsieht, und darum, wie viele Antworten entsprechend möglich sind. Drittens ist relevant, in welcher Form die Antworten vorgegeben sind: als geschlossene Liste, als offenes Textfeld oder aber als eine Mischung dieser beiden Optionen.

4.2 Sprachfragen im kanadischen Zensus

Der kanadische Zensus erhebt die Sprache der Bevölkerung auf recht geschickte und vorbildliche Weise. In Kanada haben Spracherhebungen eine lange Tradition. Der nationale Zensus erhebt die Sprache der Bevölkerung auf relativ komplexe Weise: Es wird nicht nur eine Frage gestellt, sondern vier (z. B. im Zensus von 2016; vgl. Abb. 2):

²¹ Vgl. hierzu auch die Empfehlungen der Vereinten Nationen (United Nations Economic Commission for Europe (Hg.) 2006, 2015). Einen Überblick über Sprachfragen in Zensus weltweit erhält man beispielsweise in Arel (2002) und in Humbert/Coray/Duchêne (2018).

7. Can this person speak English or French well enough to conduct a conversation?
Mark one circle only.

- 1: English only
- 2: French only
- 3: Both English and French
- 4: Neither English nor French

8. a) What language does this person speak most often at home?

- 1: English
- 2: French
- 3: Other language – specify

b) Does this person speak any other languages on a regular basis at home?

- 1: No
- 2: Yes, English
- 3: Yes, French
- 4: Yes, Other language – specify

9. What is the language that this person first learned at home in childhood and still understands?
If this person no longer understands the first language learned, indicate the second language learned.

- 1: English
- 2: French
- 3: Other language – specify

Abb. 2: Fragen zur Sprache der Bevölkerung im kanadischen Zensus 2016²²

Der Fokus der Fragen liegt dabei jeweils auf den offiziellen Staatssprachen Englisch und Französisch. Mehrsprachigkeit und Sprachenvielfalt werden von den Fragen aber prinzipiell mitgedacht und können entsprechend angegeben werden. Das Set an Fragen ist so zusammengestellt, dass es möglich ist, die vorhandenen (auch komplexeren) Sprachrepertoires der Bevölkerung recht detailliert nachzuzeichnen.

Die erste Frage erhebt die Kompetenz in den beiden offiziellen Staatssprachen Englisch und Französisch. Die zweite Frage elizitiert die vorwiegend gesprochene Haushaltssprache. Die Formulierung der Frage entspricht der Sprachfrage des deutschen Mikrozensus. Als Antwort können zunächst die geschlossen angebotenen offiziellen Sprachen Englisch und Französisch ausgewählt werden; als dritte, offene Option kann eine andere Sprache genannt werden. Die Kritik an der Formulierung der Frage im deutschen Mikrozensus trifft hier allerdings nicht zu, da erstens auf Personenebene gefragt wird, zweitens keine fragwürdige Liste vorgelegt wird, sondern eine eigene Antwort formuliert werden kann, und weil drittens mehrsprachige Konstellationen über die Gesamtheit des Fragensets abgebildet werden können. Der zweite Teil der zweiten Frage bzw. die dritte Frage elizitiert alle anderen im Haushalt auch noch gesprochenen Sprachen. Es gibt vier Antwortmöglichkeiten auf diese Frage: erstens die Verneinung der Frage, zweitens und drittens die Angabe der offiziellen Sprachen Englisch und Französisch und viertens wieder die offene Option, eine andere Sprache anzugeben. Als letzte Frage des Sets gibt es dann noch eine

²² www12.statcan.gc.ca/nhs-enm/2016/ref/questionnaires/questions-eng.cfm (Stand: 1.7.2019).

Frage nach der Spracherwerbshistorie der Befragten. Es ist somit auch möglich, Sprachen anzugeben, die zwar zunächst erworben, später aber aufgegeben wurden. Die Antwortoptionen sind wie bei den anderen Fragen die folgenden: die zwei geschlossenen Antwortmöglichkeiten Englisch und Französisch und drittens die offene Option zur Angabe jeder anderen Sprache. Durch die Kombination der Fragen können die Sprachrepertoires der Bevölkerung recht detailliert – auch erwerbshistorisch – abgebildet werden. Es kann dabei nicht nur die Kompetenz der offiziellen Sprachen beschrieben werden. Vielmehr kann auch dargestellt werden, welche diversen Sprachen die kanadische Bevölkerung überhaupt (zu Hause) spricht bzw. früher sprechen konnte.

Interessant ist auch die Position der Sprachfragen im kanadischen Zensus. Sie unterscheidet sich deutlich von derjenigen der Frage im deutschen Mikrozensus. Im kanadischen Zensus werden die Fragen relativ früh im einleitenden, allgemeinen Teil des Fragebogens gestellt; sie werden als Fragen 7 bis 9 direkt nach den Fragen zum Geschlecht, zum Geburtstag, zum Familienstand und zu den familiären Beziehungen zwischen den im Haushalt gemeinsam lebenden Personen geführt. Die Sprachfragen reihen sich damit ein in die soziodemographischen Basisinformationen über die Bevölkerung. Auch im kanadischen Zensus gibt es einen Block mit Fragen zu Staatsangehörigkeit und verwandten Themen (z. B. Geburtsort, Staatsangehörigkeit etc.). Diese soziokulturellen Informationen²³ folgen im kanadischen Fragebogen allerdings erst nach den Fragen zu *Activities of Daily Living* (i. e. Fragen zu Gesundheitsproblemen, z. B. Seh- und Höreinschränkungen).

Während der Sprachfrage im deutschen Mikrozensus anscheinend eine Einsprachigkeitsideologie zugrunde liegt, ist das beim kanadischen Amt für Statistik offensichtlich nicht der Fall. Mehrsprachigkeit und Sprachenvielfalt werden hier offen und direkt angesprochen. So heißt es im Titel eines Beitrags auf der Seite von Statistics Canada/Statistique Canada, bei dem es um die Sprachenfragen im Zensus geht, explizit „Linguistic diversity and multilingualism in Canadian homes“, also Sprachenvielfalt und Mehrsprachigkeit in kanadischen Haushalten.²⁴ Ein weiterer Hinweis auf die offenkundig positive Einstellung gegenüber Mehrsprachigkeit und Sprachenvielfalt findet sich in den Erläuterungen zu den Zensusfragen. Darin heißt es zu den Sprachfragen, dass diese dazu dienen sollen, ein Profil der sprachlichen Vielfalt von Kanadas Bevölkerung bereitzustellen und den aktuellen Status und die Entwicklung der diversen kanadischen Sprachgruppen zu bewerten. Verknüpft wird diese Angabe mit dem Wunsch, den Bedarf nach englisch- und französischsprachigen Diensten (*services*) zu erheben.

Die Sprachfragen aus dem kanadischen Zensus erweisen sich als eine sehr gute Möglichkeit, die Sprache der Bevölkerung zu erheben. Sie stellen somit einen positiven Gegenentwurf zur Art und Weise dar, wie Sprache im deutschen Mikrozensus erhoben wird und wie die Ergebnisse gehandhabt werden.

²³ Sociocultural information, vgl. www12.statcan.gc.ca/nhs-enm/2016/ref/questionnaires/questions-eng.cfm (Stand: 8.11.2018).

²⁴ Vgl. www12.statcan.gc.ca/census-recensement/2016/as-sa/98-200-x/2016010/98-200-x2016010-eng.cfm (Stand: 8.11.2018).

5 Eine andere Spracherhebung in Deutschland

5.1 Die Sprachfrage in der Deutschland-Erhebung 2018

Eingangs wurde zwar konstatiert, dass die Datenlage zur Sprachensituation in Deutschland insgesamt nicht besonders gut ist. Allerdings gibt es in der Sprachwissenschaft durchaus einige Bemühungen um Abhilfe. Das Leibniz-Institut für Deutsche Sprache hat im Herbst/Winter 2018 in Kooperation mit dem Deutschen Institut für Wirtschaftsforschung (DIW) im Rahmen der Innovationsstichprobe des Sozio-oekonomischen Panels (SOEP-IS) des DIW eine Repräsentativerhebung durchgeführt, in der auch nach Sprachkompetenzen gefragt wurde (vgl. Adler/Plewnia 2018; Richter/Schupp 2012), die Deutschland-Erhebung 2018. Diese Erhebung fand also in etwa zeitgleich mit der ersten Erhebung der Sprachdaten im Mikrozensus statt. Die repräsentative Deutschland-Erhebung 2018 besteht aus zwei Teilen, einem direkten Interview (CAPI, *computer assisted personal interview*, $N = 4.339^{25}$ in 1.784 Haushalten) zu Sprach- und Dialektkompetenz sowie zu Sprach- und Dialektbewertungen und einem Onlinefragebogen (CAWI, *computer assisted web interview*, $N = 1.439$) zu einem breiten Spektrum weiterer sprachbezogener Fragen.

In der Deutschland-Erhebung 2018 wurde unter anderem auch die Erstsprache erhoben. Die Frage lautete: „Welche Sprache bzw. welche Sprachen würden Sie als Ihre Muttersprache bezeichnen?“²⁶ Anders als im Mikrozensus wurde für die Frage hier der alltags-sprachliche Terminus „Muttersprache“ gewählt. Es wurde explizit nach mehreren Sprachen gefragt; entsprechend war es den Befragten auch möglich, mehrere Antworten zu geben. Dadurch können auch mehrsprachige Konstellationen abgebildet werden. Außerdem war die Frage offen gestellt. Für die Interviewer gab es zwar eine Precodeliste mit etlichen Antwortmöglichkeiten, allerdings ohne Sammelkategorien. Für die Befragten war die Liste nicht einsehbar, sie spielte für sie bei der Beantwortung keine Rolle. Die frei formulierten Antworten wurden für die Datenauswertung aufgearbeitet. Anders als im Mikrozensus wurden alle Personen der Stichprobe befragt und nicht nur lediglich eine Person stellvertretend pro Haushalt. Zwar ist die Stichprobe im Vergleich zum Mikrozensus eher klein, für die in der Sprachwissenschaft üblichen Größenordnungen handelt es sich bei der Deutschland-Erhebung 2018 allerdings bereits um einen sehr großen Datensatz. Da die Personen auch Haushalten zugeordnet werden können, ist es möglich, die erhobenen Sprachen sowohl auf Personenebene als auch auf Haushaltsebene auszuwerten und die Ergebnisse dann zu vergleichen.

5.2 Ergebnisse der Sprachfragen in der Deutschland-Erhebung 2018

Von den 4.339 Befragten nennen 3.816 Deutsch als ihre Muttersprache, das ist mit 87,9% die große Mehrheit. 14,4% der Befragten ($n = 628$) geben eine andere Muttersprache an. Bemerkenswert ist, dass es für diese Frage keine fehlenden Angaben gibt. Das ist auch ein Zeichen dafür, dass die Frage gut funktioniert hat und offensichtlich einfach zu beantworten war. Die Ergebnisse der Frage sind in der folgenden Tabelle detailliert dargestellt:

²⁵ Die Daten sind mit einem vorläufigen Gewichtungsfaktor gewichtet.

²⁶ Die Frage nach der Muttersprache bzw. den Muttersprachen ist Teil eines umfangreicheren Sets an Fragen zur Sprachkompetenz der Befragten. Abgefragt wurde beispielsweise auch die Dialektkompetenz. Diese Fragen wurden im direkten Interview der Deutschland-Erhebung 2018 gestellt.

Sprache	absolute Häufigkeit	in Prozent
<i>Deutsch</i>	3.816	87,9
ohne Angabe	–	
nicht Deutsch	628	14,4
davon		davon
<i>Russisch</i>	104	17
<i>Türkisch</i>	99	16
<i>Polnisch</i>	81	13
<i>Italienisch</i>	56	9
<i>Englisch</i>	43	7
<i>Spanisch</i>	34	5
<i>Griechisch</i>	27	4
<i>Rumänisch</i>	17	3
<i>Tamil</i>	16	3
<i>Kroatisch</i>	15	2
<i>Niederländisch</i>	14	2
<i>Ungarisch</i>	14	2
<i>Albanisch</i>	12	2
<i>Serbisch</i>	12	2
<i>Portugiesisch</i>	11	2
<i>Französisch</i>	8	1
<i>Arabisch</i>	8	1
<i>Tschechisch</i>	7	1
<i>Japanisch</i>	6	1
<i>Ghana/Ewe/Twi</i>	5	1
<i>Vietnamesisch</i>	5	1
<i>Mazedonisch</i>	3	0,5
<i>Montenegro</i>	3	0,5
<i>Serbokroatisch</i>	3	0,5
<i>Slowenisch</i>	3	0,5
<i>Bulgarisch</i>	2	0,3
<i>Finnisch</i>	2	0,3
<i>Libanesisch</i>	2	0,3
<i>Luxemburg</i>	2	0,3
<i>Thailändisch</i>	2	0,3
<i>Ukrainisch</i>	2	0,3
weitere Sprachen	11	2
insgesamt	4.339	

Tab. 2: Ergebnisse der Sprachfrage (Muttersprache(n)) in der Deutschland-Erhebung 2018²⁷

²⁷ Weitere Sprachen werden jeweils wie folgt genannt: *Aramäisch, Aserbaischan, Georgisch, Indonesisch, Kurdisch, Latein, Lettisch, Marokanisch, Mischung Russisch/Deutsch, Philippinisch, Plattdeutsch.*

Am häufigsten werden neben Deutsch folgende Sprachen als Muttersprachen angegeben: Russisch (mit 17% innerhalb der Sprachen, die nicht Deutsch sind), Türkisch (16%) und Polnisch (13%). Es folgen Italienisch (9%), Englisch (7%), Spanisch (5%) und Griechisch (4%). Alle anderen Sprachen werden von nur sehr wenigen Befragten genannt; darunter sind viele Sprachen, die insbesondere in osteuropäischen Ländern gesprochen werden (z. B. Rumänisch, Ungarisch, Tschechisch), und Sprachen des Balkans (z. B. Kroatisch, Serbisch, Albanisch, Serbokroatisch). Würde man einige dieser Nennungen zu größeren Gruppen zusammenführen, würden sie innerhalb der Tabelle deutlich weiter oben geführt werden, als sie es so als Einzelnennungen sind.

Unter den Antworten, die andere Sprachen als Deutsch bezeichnen (14,4%), befinden sich 39 Sprachen oder sich auf Sprachen beziehende Kategorien (wie etwa Ländernamen). Das zeigt, dass innerhalb dieser kleinen Menge eine recht große Vielfalt herrscht. Viele Sprachen werden von nur einem, zwei oder sehr wenigen Befragten angegeben. Das entspricht wahrscheinlich der sozialen Realität,²⁸ allerdings liegt darin auch ein Problem. Die Deutschland-Erhebung ist zwar relativ groß, aber die bezüglich der Mehrsprachigkeit und Sprachenvielfalt interessante Teilstichprobe macht nur einen kleinen Teil dieser großen Stichprobe aus. Will man diesen kleinen Teil genauer betrachten, also weiter in kleinere Gruppen zerschneiden, werden diese Untergruppen zwangsläufig noch kleiner. Entsprechend ist die Validität dieser Daten (d. h. hier z. B. auch deren Reihung) mit großer Vorsicht zu betrachten.

Die Frage ließ Mehrfachantworten zu, deshalb können nun diese Mehrfachnennungen und deren Kombinationen betrachtet werden: 97 Befragte geben zwei Muttersprachen an, 5 Befragte nennen drei Muttersprachen, 3 Befragte nennen vier Muttersprachen. Davon sind es 86 Befragte, die Deutsch und eine weitere Muttersprache angeben, 5 Befragte nennen Deutsch und zwei weitere Muttersprachen, und 3 Befragte Deutsch und drei Muttersprachen. Die meisten davon nennen neben Deutsch Russisch (27 Nennungen), Englisch (13), Türkisch (12), Italienisch (10) und Spanisch (7).

In der Deutschland-Erhebung 2018 wurden die Sprachfragen an Personen und nicht an Haushalte gestellt. Die Befragten lassen sich aber Haushalten zuordnen. Das bedeutet auch, dass man bei Mehrpersonenhaushalten die Angaben der Haushaltsmitglieder miteinander vergleichen kann. Das betrifft einerseits die Angaben zu den Muttersprachen und andererseits auch jene zu den zuhause gesprochenen Sprachen.²⁹ Letzteres ist insbesondere deshalb interessant, da eine solche Auswertung einen weiteren Hinweis darauf liefern kann, inwiefern die Angabe einer Person stellvertretend für einen gesamten Haushalt aussagekräftig ist. Die Analyse der Angaben über die zuhause gesprochenen Sprachen nach Haushalten ermöglicht also eine weitere Bewertung der Güte der Mikrozensus Sprachfrage.

Tatsächlich finden sich innerhalb der Haushalte, deren Mitglieder die Frage nach den zuhause gesprochenen Sprachen beantwortet haben, einige Dissonanzen: In einem Haushalt mit zwei Personen beispielsweise, von denen beide als Muttersprache nur Polnisch

²⁸ Die Verteilung scheint folgendermaßen zu sein: Viele Menschen in Deutschland sprechen nur eine Sprache, nämlich Deutsch; daneben sprechen nur wenige Menschen viele andere Sprachen.

²⁹ Allerdings ist die letzte der beiden Auswertungen nur eingeschränkt möglich, da nicht allen Befragten der Deutschland-Erhebung 2018 die Frage nach den zuhause gesprochenen Sprachen gestellt wurde. Die Frage wurde lediglich den Befragten gestellt, die zuvor nicht Deutsch als Muttersprache angegeben hatten.

angeben, nennt eine Person als zuhause gesprochene Sprache Polnisch, die andere gibt Deutsch und Polnisch an. Ähnliche verschiedene Angaben finden sich bei Haushalten mit Personen mit anderen Muttersprachen wie etwa Griechisch, Italienisch, Russisch oder Türkisch. Ein weiteres Beispiel: In einem Haushalt mit vier Haushaltsmitgliedern geben alle Tamilisch als einzige Muttersprache an. Die Angaben zu den zuhause gesprochenen Sprachen sind allerdings verschieden. Zwei dieser Personen nennen Tamilisch als einzige zuhause gesprochene Sprache, eine nur Deutsch, und die vierte gibt Deutsch und Tamilisch an. In einem anderen Haushalt mit fünf Personen, von denen alle Türkisch als einzige Muttersprache angeben, nennen drei ausschließlich Deutsch als zuhause gesprochene Sprache, eine Person nennt Türkisch, und die fünfte nennt Deutsch und Türkisch. Diese Beispiele belegen die oben formulierte Annahme, dass die Mitglieder eines gleichen Haushalts nicht zwangsläufig die gleichen Angaben über die zuhause gesprochene Sprache machen.

Hinzu kommt, dass von den Befragten, die die Frage nach zuhause gesprochenen Sprachen beantworten, 20,3 % (123 Befragte) mehr als eine Sprache nennen. Es gibt also einen beträchtlichen Teil, der zuhause mehrere Sprachen spricht; außerdem geben quasi alle diese Befragten als eine dieser gesprochenen Sprachen Deutsch an (i. e. 122 Befragte). Das unterstreicht auch, dass bei einer erzwungenen Einzelantwort, die eine andere Sprache als Deutsch ist, nicht zwangsläufig davon ausgegangen werden kann, dass im Haushalt nicht auch Deutsch gesprochen wird.³⁰ Obgleich sich die Auswertung der zuhause gesprochenen Sprachen in der Deutschland-Erhebung 2018 nur auf eine kleine Teilstichprobe bezieht, liefern diese Ergebnisse weitere Hinweise darauf, dass die Ergebnisse der Sprachfrage im Mikrozensus nur bedingt aussagekräftig sein können.

5.3 Vergleich der Ergebnisse der Spracherhebungen in Deutschland

Die Ergebnisse beider Erhebungen – des Mikrozensus einerseits und der Deutschland-Erhebung 2018 andererseits – sind natürlich nur eingeschränkt vergleichbar. Das liegt daran, dass den erhobenen Sprachen nicht die gleiche Frage zugrunde liegt. Legt man die Ergebnisse beider Erhebungen nebeneinander, kann man allerdings sehen, dass das grobe Muster vergleichbar ist: Die Mehrheit der Menschen in Deutschland spricht Deutsch. Dann aber gibt es einen großen Unterschied: Im Mikrozensus gibt es 3,9 % fehlende Angaben, in der Deutschland-Erhebung 2018 dagegen keine. Das ist deshalb so schwerwiegend, weil das zweite inhaltlich relevante Ergebnis in einer ähnlichen Größenordnung liegt. Es geht um den Anteil aller Sprachen, die nicht Deutsch sind. Hier stehen 9 % im Mikrozensus 14 % in der Deutschland-Erhebung 2018 gegenüber. Die knapp 4 % fehlenden Angaben würden hier schon etwas ausmachen.

Innerhalb dieser anderen Sprachen unterscheiden sich die Ergebnisse zum Teil recht deutlich. Allerdings sollten diese Unterschiede mit Bedacht bewertet werden, da es sich hier um vergleichsweise kleine Teilstichproben (bei der Deutschland-Erhebung 2018) bzw. um ein bedenkliches Fragedesign (beim Mikrozensus) handelt. Die zahlenmäßig auffälligsten Unterschiede zu den Ergebnissen des Mikrozensus betreffen einerseits Russisch, das in der Deutschland-Erhebung 2018 die am häufigsten genannte Sprache ist; da hier nicht die Haushaltssprache, sondern die Muttersprache gefragt wurde, muss das keinen Widerspruch darstellen. Zum anderen betrifft das Arabisch, das hier erheblich seltener genannt

³⁰ Es ist natürlich des Weiteren selbstverständlich, dass die Angabe einer anderen Sprache als Deutsch als im Haushalt vorwiegend gesprochene keine Schlüsse auf die Deutschkenntnisse zulässt.

wird; das dürfte in erster Linie damit zusammenhängen, dass die Stichprobe des SOEP-IS in dieser Hinsicht noch nicht gut balanciert ist, wie überhaupt die kleinen Zahlen nur als Tendenzen zu deuten sind; hier müsste die Stichprobe deutlich größer sein. Vor allem aber haben wir durch das Fragesetting die Möglichkeit zu einer weitreichenden Differenzierung; beispielsweise sieht man, dass die „sonstigen europäischen Sprachen“ Rumänisch, Niederländisch und Griechisch jeweils noch vor Französisch liegen, oder dass etwa die Balkansprachen zusammen eine zahlenmäßig relevante Gruppe bilden.

6. Sprache und nationale Identität

Abschließend kann man also Folgendes festhalten: Inzwischen gibt es auch in Deutschland Sprachstatistiken. Allerdings sind diese mit großer Vorsicht zu betrachten. Die Auswertung der Sprachfrage des Mikrozensus hat gezeigt, dass die Ergebnisse weder als objektive Sprachstatistik noch als Indikator für ethnokulturelle Integration dienen können. Die eigentlich banale Feststellung, wie wichtig es ist, eine gute und geeignete Frage zu stellen, hat sich hierbei deutlich bewahrheitet. Aus dem prinzipiell großen Spektrum an möglichen Sprachfragen zeigt etwa der kanadische Zensus ein exemplarisches Vorgehen beim Erheben des Sprachrepertoires seiner Bevölkerung. Für Deutschland kann zu diesem Zeitpunkt lediglich die Deutschland-Erhebung 2018 objektiv erhobene Daten zur Sprache der Bevölkerung vorweisen. Allerdings gilt hier, dass die Stichprobe – insbesondere im Vergleich zum Mikrozensus – recht klein ist. Gerade um aussagekräftig Daten über die mehrsprachigen Sprecher zu erhalten, ist eine große Stichprobe aber unerlässlich. Denn ebendiese Sprecher sind unter der deutschen Wohnbevölkerung in der Minderheit.

Es hat sich außerdem gezeigt, dass der Mikrozensus als eine vermeintlich objektive Zählung nicht frei von Ideologien und Einstellungen ist. Einen weiteren Beleg dafür liefert die Pressemitteilung, die das Statistische Bundesamt zu den Ergebnissen der Sprachfrage vorgelegt hat.³¹ Der darin gewählte Sprachduktus (nicht deutsche Sprachen werden wiederholt als „ausländische Sprachen“ bezeichnet), die Auswahl der dargelegten Ergebnisse sowie deren Darstellungsweise (z. B. die Bezugsgröße bei prozentualen Angaben)³² machen aus dem Text, der im Prinzip eine objektive Darstellung objektiver Ergebnisse sein sollte, einen stark subjektiv eingefärbten Text. Laienlinguistische Wahrnehmungen der Sprachsituation in Deutschland (i. e. als eine vor allem einsprachige) und Einstellungen gegenüber anderen in Deutschland gesprochenen Sprachen haben wohl die Sprachfrage des Mikrozensus und deren Handhabung beeinflusst. Dass es auch anders geht, zeigt wiederum der kanadische Zensus mit seiner Erhebung der Sprache und seinem Umgang damit.

In diesem Zusammenhang sind sicherlich weitaus grundlegendere Fragen relevant, etwa welche Sprachen als Teil der deutschen Gesellschaft angesehen werden und welche nicht. Gerade diese Fragestellungen sind wesentlich, wenn es bei der Intention einer Frage um ethnokulturelle Identität gehen soll. Gilt es, die ethnokulturelle Integration zu messen, dann geht es letztlich auch um die ethnokulturelle bzw. nationale Identität.

³¹ www.destatis.de/DE/PresseService/Presse/Pressemitteilungen/2018/08/PD18_282_12511.html (Stand: 4.2.2019).

³² Für Türkisch beispielsweise könnten demnach entweder 1,8% oder 17% und für Russisch entweder 1,6% oder 15% angegeben werden, in der Pressemitteilung werden jeweils die größer wirkenden Zahlen gewählt.

Eines ist also ganz offensichtlich: Die Sprachfrage im Mikrozensus muss verbessert werden. Dabei muss es nicht gleich eine ausführliche Erhebung wie im kanadischen Zensus sein. Eine gut formulierte Frage mit sinnvollen Antwortoptionen, inklusive einer offenen, und mit der Möglichkeit, mehrere dieser Antwortoptionen zu nennen, sollte aber das Minimum sein. All diese Varianten sind im deutschen Mikrozensus bei anderen Fragen vorhanden und sollten demnach also problemlos auch bei der Sprachfrage umsetzbar sein.

Literatur

- Adler, Astrid (2018): Die Frage zur Sprache der Bevölkerung im deutschen Mikrozensus 2017. Arbeitspapier. Mannheim: Institut für Deutsche Sprache. (Online: <https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/7318>, Stand: 26.2.2019).
- Adler, Astrid/Beyer, Rahel (2018): Languages and language politics in Germany/Sprachen und Sprachpolitik in Deutschland. In: Stickel, Gerhard (Hg.): National language institutions and national languages. Contributions to the EFNIL Conference 2017 in Mannheim. Budapest: Hungarian Academy of Science, Research Institute for Linguistics. S. 221–242.
- Adler, Astrid/Plewnia, Albrecht (2018): Möglichkeiten und Grenzen der quantitativen Spracheinstellungsforschung. In: Lenz, Alexandra N./Plewnia, Albrecht (Hg.): Variation – Normen – Identitäten. Berlin/Boston: De Gruyter. S. 63–98. (= Germanistische Sprachwissenschaft um 2020, 4).
- Anders, Carsten (2018): 60 Jahre Mikrozensus. Ein Rück- und Ausblick. In: Statistisches Monatsheft Baden-Württemberg 2. S. 3–9. (Online: www.statistik-bw.de/Service/Veroeff/Monatshefte/PDF/Beitrag_18_02_01.pdf, Stand: 4.2.2019).
- Arel, Dominique (2002): Language questions in censuses: backward or forward-looking? In: Kertzer, David I./Arel, Dominique (Hg.): Census and identity. The politics of race ethnicity and language in national censuses. Cambridge: Cambridge University Press. S. 92–120. (= New perspectives on anthropological and social demography 1).
- Bourdieu, Pierre (1991): Language and symbolic power. ed. and introd. by John B. Thompson. transl. by Gino Raymond and Matthew Adamson. Cambridge: Polity Press.
- Busch, Brigitta (2010): ... und Ihre Sprache? Über die Schwierigkeiten, eine scheinbar einfache Frage zu beantworten. In: Stichproben. Wiener Zeitschrift für kritische Afrikastudien 10, 19. S. 9–33.
- Busch, Brigitta (2016): Categorizing languages and speakers. Why linguists should mistrust census data and statistics. (Online: www.academia.edu/20770728/WP189_Busch_2016_Categorizing_languages_and_speakers_Why_linguists_should_mistrust_census_data_and_statistics, Stand: 1.3.2018). (= Working Papers in Urban Language & Literacies 189).
- Gärtig, Anne-Kathrin/Plewnia, Albrecht/Rothe, Astrid (2010): Wie Menschen in Deutschland über Sprache denken. Ergebnisse einer bundesweiten Repräsentativerhebung zu aktuellen Spracheinstellungen. Mannheim: Institut für Deutsche Sprache. (= amades 40).
- Glück, Hartmut (1979): Die preußisch-polnische Sprachenpolitik. Eine Studie zur Theorie und Methodologie der Forschung über Sprachenpolitik, Sprachbewußtsein und Sozialgeschichte am Beispiel der preußisch-deutschen Politik gegenüber der polnischen Minderheit von 1914. Hamburg: Buske.
- Gogolin, Ingrid (1997): The „monolingual habitus“ as the common feature in teaching in the language of the majority in different countries. In: *Per Linguam*, 13, 2. S. 38–49.
- Humbert, Philippe/Coray, Renata/Duchêne, Alexandre (2018): Compter les langues: histoire, méthodes et politiques des recensements de population. Une revue de la littérature. Rapport du Centre scientifique de compétence sur le plurilinguisme. Fribourg: Institut de plurilinguisme.
- Kemper, Thomas (2010): Migrationshintergrund – eine Frage der Definition! In: *Die Deutsche Schule* 102, 4. S. 315–326. (Online: www.pedocs.de/volltexte/2012/5151/pdf/Kemper_2010_Migrationshintergrund_Frage_der_Definition_D_A.pdf, Stand: 4.2.2019).
- Kemper, Thomas (2015): Bildungsdisparitäten von Schülern nach Staatsangehörigkeit und Migrationshintergrund. Eine schulformspezifische Analyse anhand von Daten der amtlichen Schulstatistik. Münster

- u. a.: Waxmann. (Online: www.pedocs.de/volltexte/2017/14643/pdf/Kemper_2015_Bildungsdisparitaeten.pdf, Stand: 4.2.2019).
- Kemper, Thomas (2017): Die schulstatistische Erfassung des Migrationshintergrundes in Deutschland. In: *Journal for Educational Research Online (JERO)* 9, 1. S. 144–168. (Online: www.j-e-r-o.com/index.php/jero/article/download/737/308, Stand: 4.2.2019).
- Labbé, Morgane (2003): Dénombrer les nationalités en Prusse au XIXe siècle: entre pratique d’administration locale et connaissance statistique de la population. In: *Annales de démographie historique* 105, 1. S. 39–61.
- Lameli, Alfred (2013): Strukturen im Sprachraum. Analysen zur arealtypologischen Komplexität der Dialekte in Deutschland. Berlin/Boston: De Gruyter. (= Linguistik – Impulse & Tendenzen 54).
- Leuschner, Torsten (2004): Richard Böckh (1824–1907): Sprachenstatistik zwischen Nationalitätsprinzip und Nationalstaat. In: *Historiographia Linguistica* 31, 2. S. 389–421.
- Richter, David/Schupp, Jürgen (2012): SOEP Innovation Sample (SOEP-IS). Description, Structure and Documentation. In: *SOEPpapers* 463. (Online: www.diw.de/documents/publikationen/73/diw_01.c.407141.de/diw_sp0463.pdf, Stand: 7.12.2017).
- Sebba, Mark (2017): ‘English a foreign tongue’. The 2011 Census in England and the misunderstanding of multilingualism. In: *Journal of Language and Politics* 16, 2. S. 264–284.
- Statistische Ämter des Bundes und der Länder (Hg.) (2018): Informationen zum Mikrozensus., Wiesbaden: Statistisches Bundesamt. (Online: www.it.nrw/sites/default/files/atoms/files/mikrozensus-infobroschue_re_fuer_befragte_2019.pdf, Stand: 15.1.2019).
- Statistisches Bundesamt (Hg.) (2018): Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund. Ergebnisse des Mikrozensus 2017. Fachserie 1, Reihe 2. (Online: www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/MigrationIntegration/Migrationshintergrund2010220177004.pdf?__blob=publicationFile, Stand: 4.2.2019).
- Stickel, Gerhard (2012): Deutsch im Kontext anderer Sprachen in Deutschland heute. Daten und Einschätzungen. Unter Mitarbeit von Julia Weinheimer. In: Eichinger, Ludwig M. et al. (Hg.): *Sprache und Einstellungen. Spracheinstellungen aus sprachwissenschaftlicher und sozialpsychologischer Perspektive. Mit einer Sprachstandserhebung zum Deutsch von Gerhard Stickel*. Tübingen: Narr. S. 227–321. (= Studien zur Deutschen Sprache 61).
- United Nations Economic Commission for Europe (Hg.) (2006): *Conference of European Statisticians Recommendations for the 2020 Census of Population and Housing*. Prepared in cooperation with the Statistical Office of the European Communities (EUROSTAT). New York/Genf: United Nations.
- United Nations Economic Commission for Europe (Hg.) (2015): *Conference of European Statisticians Recommendations for the 2010 Census of Population and Housing*. New York/Genf: United Nations.
- Will, Anne-Kathrin (2016): 10 Jahre Migrationshintergrund in der Repräsentativstatistik. Ein Konzept auf dem Prüfstand. In: *Leviathan – Berliner Zeitschrift für Sozialwissenschaft* 44, 1. S. 9–34.
- Will, Anne-Kathrin (2018): Der Migrationshintergrund im Mikrozensus. Wie werden Zuwanderer und ihre Nachkommen in der Statistik erfasst? Informationspapier für den Mediendienst Integration. (Online: https://mediendienst-integration.de/fileadmin/Dateien/Informationspapier_Mediendienst_Integration_Migrationshintergrund_im_Mikrozensus_Aktualisierung_2018.pdf, Stand: 4.2.2019).

Dr. Astrid Adler
 Leibniz-Institut für Deutsche Sprache
 Postfach 101621
 68016 Mannheim
 E-Mail: adler@ids-mannheim.de