

Astrid Adler / Maria Ribeiro Silveira

# WELCHE SPRACHEN WERDEN IN DEUTSCHLAND GESPROCHEN?

## SPRACHE IN ZAHLEN: FOLGE 4

Das Leibniz-Institut für Deutsche Sprache (IDS) hat in Kooperation mit dem Deutschen Institut für Wirtschaftsforschung (DIW) eine deutschlandweite Repräsentativerhebung unter 4.380 Personen zum Sprachrepertoire der Menschen in Deutschland und zu ihren Spracheinstellungen durchgeführt: die *Deutschland-Erhebung 2017*. Eine Einführung in die *Deutschland-Erhebung 2017* findet sich im ersten Teil dieser Serie in Heft 4/2020. In dieser Folge berichten wir darüber, welche Sprachen in Deutschland gesprochen werden.

In den letzten Folgen unserer Reihe haben wir Zahlen zu den in Deutschland gesprochenen Dialekten vorgestellt (siehe Folge 2 und 3). In der aktuellen Folge geht es nun darum, welche Sprachen in Deutschland gesprochen werden. Die nationale Amtssprache in Deutschland ist Deutsch; daneben gibt es die gesetzlich verankerten Regional- und Minderheitensprachen (Niederdeutsch, Dänisch, Friesisch, Romanes, Sorbisch). Hinzu kommen weitere Sprachen wie beispielsweise Russisch oder Türkisch, die von vielen Menschen in Deutschland gesprochen werden. Gesicherte Zahlen über die Sprecherinnen und Sprecher dieser Sprachen gibt es nicht, da es bislang in Deutschland keine amtliche Sprachstatistik gibt.

In der *Deutschland-Erhebung 2017* sind bei den Fragen zum Sprachrepertoire nicht nur Fragen zu den Dialekten, sondern auch Fragen zu den Muttersprachen, den zuhause gesprochenen Sprachen und weiteren Sprachkenntnissen enthalten. Die Frage zu den Muttersprachen lautet wie folgt: „Welche Sprache bzw. welche Sprachen würden Sie als Ihre Muttersprache bezeichnen?“ Die Befragten konnten mehrere Antworten geben; somit können auch mehrsprachige Sprachrealitäten abgebildet werden. 2,2% der Befragten nutzen diese Möglichkeit und geben zwei und mehr Sprachen als Muttersprachen an.

Die grobe Verteilung der Antworten ist – wie man sicherlich erwarten würde – sehr deutlich (siehe die Ergebnisse der *Deutschland-Erhebung 2017* im Hochbalken links in Abb. 1): Die große Mehrheit gibt Deutsch als Muttersprache an. 85,3% nennen nur Deutsch als Muttersprache und 1,8% Deutsch und eine bzw. mehrere weitere Sprachen als Muttersprachen. 12,9% der Befragten nennen eine oder mehrere andere Sprachen als Deutsch als Muttersprachen. Als andere Sprachen werden sehr viele verschiedene Sprachen angegeben, die meisten davon werden von weniger als zehn Befragten genannt (das entspricht 1,5% der genannten anderen

Sprachen; siehe rechts im kreisförmigen Balkendiagramm in Abb. 1). Lediglich sieben Sprachen werden von dreißig und mehr Befragten als Muttersprachen angegeben (das entspricht 4% der Nennungen für andere Sprachen). Demnach sind nach Deutsch die häufigsten Muttersprachen in Deutschland Russisch, Türkisch, dann mit deutlichem Abstand Polnisch, Englisch und Italienisch und schließlich Spanisch und Rumänisch. Diese Verteilung – wenige Sprachen werden von vielen Menschen genannt und viele Sprachen von wenigen Menschen – ist typisch, macht die Erfassung und Aufarbeitung der Daten jedoch schwierig. Statistisch gültige Aussagen können für kleine Datenmengen nicht getroffen werden. Kleine Teilmengen werden deshalb oftmals zu größeren Mengen zusammengefasst. Bei dieser Verallgemeinerung gehen die Details der Teilmengen allerdings verloren. Die Stichprobe der *Deutschland-Erhebung 2017* ist für Erhebungen dieses Typs groß; für eine adäquate Darstellung der

Die Autorinnen sind Mitarbeiterinnen des Programmberichts Sprache im öffentlichen Raum am Leibniz-Institut für Deutsche Sprache, Mannheim.

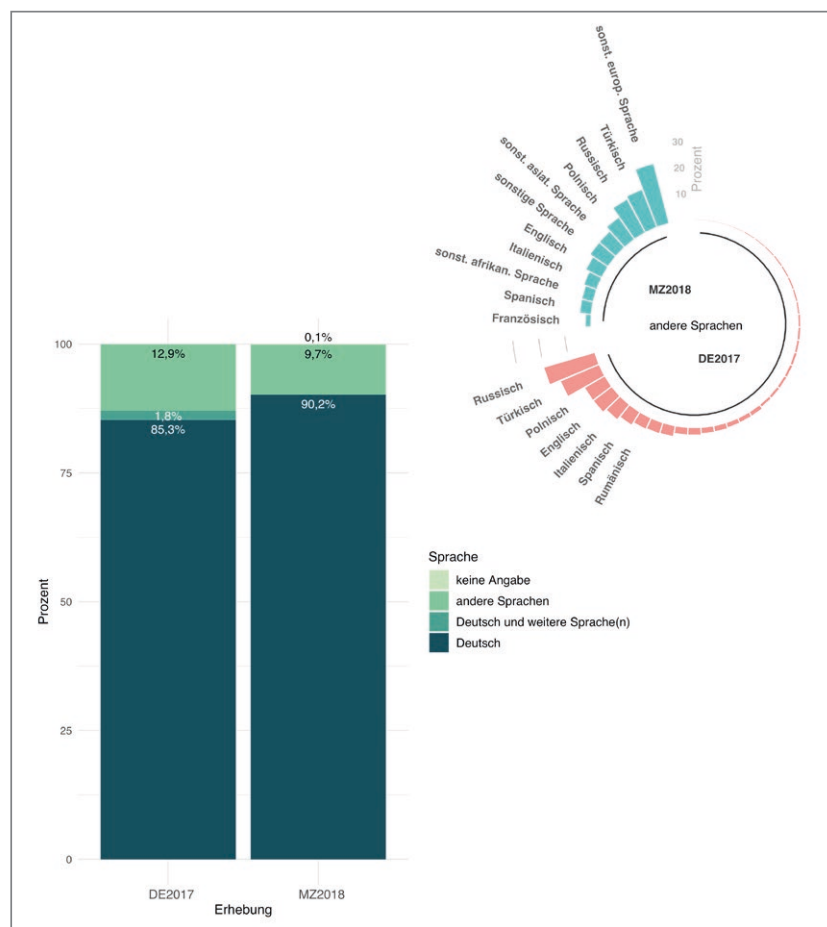


Abb. 1: Sprachen in Deutschland (*Deutschland-Erhebung 2017*, Mikrozensus 2018)<sup>1</sup>

weiteren Muttersprachen, die insgesamt lediglich einen relativ kleinen Teil dieser großen Stichprobe ausmachen, ist sie allerdings nicht groß genug. Gesicherte Aussagen über Teilstichproben dieser Teilstichprobe (Einzelsprachen innerhalb der Teilstichprobe „Muttersprache(n) nicht Deutsch“) sind nicht möglich. Für einen gesicherten Einblick in die Verteilung innerhalb dieser Teilstichprobe bedarf es einer deutlich größeren Gesamtstichprobe.

Die größte Haushaltsbefragung in Deutschland ist der Mikrozensus. Dafür wird eine einprozentige repräsentative Stichprobe befragt (d. h. etwa 810.000 Personen in ca. 370.000 Haushalten). Diese offizielle amtliche Statistik kann für das beschriebene Problem jedoch keine Abhilfe schaffen. Zwar wäre die Größe der Stichprobe ausreichend, allerdings weist das zur Erhebung der Sprache eingesetzte Erhebungssitem große Mängel auf. Im deutschen Mikrozensus wurde 2017 erstmals seit vielen Jahrzehnten wieder eine Frage zur Sprache der Bevölkerung eingesetzt. Die Frage lautet „Welche Sprache wird in Ihrem Haushalt vorwiegend gesprochen?“ Problematisch an diesem Erhebungssitem sind u. a. die Befragungseinheit Haushalt, die Beantwortung der Frage durch nur eine Person pro Haushalt und die Einschränkung auf nur eine Antwortmöglichkeit innerhalb einer begrenzten Auswahl an Antworten; und am problematischsten ist, dass anhand des Items keine mehrsprachigen Sprachrepertoires abgebildet werden können (vgl. Adler 2019). Die Ergebnisse der Sprachfrage des Mikrozensus erlauben entsprechend kein Abbild der sprachlichen Realität der Bevölkerung und müssen also mit großer Vorsicht interpretiert werden. Es ist davon auszugehen, dass die Angaben insbesondere für andere Sprachen als Deutsch verzerrt sind. Die Ergebnisse des Mikrozensus 2018 ergeben folgendes Verhältnis (siehe den Hochbalken rechts in Abb. 1): In 90,2% der Haushalte wird vorwiegend Deutsch gesprochen, in 9,7% der Haushalte wird vorwiegend eine andere Sprache gesprochen. Die häufigste Nennung nach Deutsch ist die generische Sammelkategorie „sonstige europäische Sprache“ mit 1,0% aller Nennungen bzw. 21,4% der gültigen Nennungen, die nicht Deutsch sind (siehe rechts im kreisförmigen Balkendiagramm in Abb. 1). Eine weitere, genauere Aufschlüsselung dieser Kategorie ist aufgrund des Designs des Erhebungssitems nicht möglich; die Informationsgüte dieser und der weiteren generischen Sammelkategorien ist somit sehr eingeschränkt. Bei den Einzelsprachen, die das Erhebungsdesign ermöglicht anzugeben,

sind insbesondere Türkisch und Russisch, dann Polnisch und Arabisch mit einer nennenswerten Häufigkeit in den Ergebnissen vertreten. Aufgrund der genannten Mängel kann anhand dieser Daten über die genaue Verteilung der Sprachen in Deutschland – insbesondere über die Mehrheitsposition des Deutschen hinaus, also über die Sprachen, die nicht Deutsch sind – jedoch keine zufriedenstellende Aussage getroffen werden.

Nimmt man die Ergebnisse beider Erhebungen zusammen (und bedenkt dabei die bedingte Aussagekraft der Ergebnisse des Mikrozensus), so zeigt sich ganz deutlich die klare Dominanz des Deutschen. Der Umfang der weiteren Sprachen ist im Vergleich dazu eher klein: 14,7% geben in der *Deutschland-Erhebung 2017* eine andere Sprache als Deutsch als Muttersprache bzw. eine ihrer Muttersprachen an, und laut Mikrozensus wird in 9,7% der Haushalte vorwiegend eine andere Sprache als Deutsch gesprochen.

Diese Statistiken sollen also den Ist-Zustand der Sprachen in Deutschland abbilden. In der *Deutschland-Erhebung 2017* wurden explizit auch Meinungen zu Sprachen abgefragt (dazu siehe Folge 1 und Adler/Ribeiro Silveira i. Vorb.). In diesem Zusammenhang wurde ein Erhebungssitem eingesetzt, bei dem die Befragten den Anteil an Menschen in Deutsch-

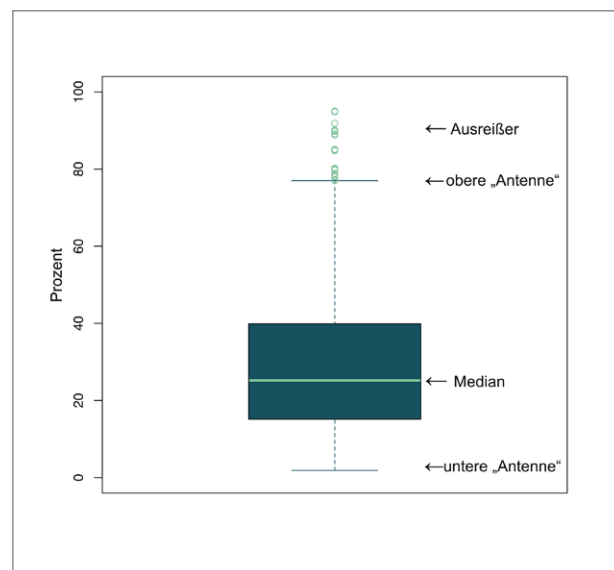


Abb. 2: Geschätzter Anteil anderer Muttersprachen als Deutsch (*Deutschland-Erhebung 2017*; N=1439)

land schätzen sollten, die eine nicht-deutsche Muttersprache haben. Die Frage lautet „Was schätzen Sie, wie viel Prozent der in Deutschland lebenden Bevölkerung hat eine andere Muttersprache als Deutsch?“ Die Ergebnisse dieser Frage sind in Abbildung 2 dargestellt.

Die geschätzten Werte liegen breit gestreut mit Werten zwischen 2% (untere „Antenne“) und 78% (obere „Antenne“) und einigen Ausreißern bis 95%. Die Hälfte der Befragten (dunkelgrüne Box) schätzt Werte zwischen 15% und 40%; der Median liegt bei 25% (der Mittelwert bei 30,3%). Im Mittel liegt die Schätzung der Befragten also höher als die aktuelle Verteilung – zumindest laut der bislang vorhandenen Daten dazu. Offensichtlich wird der Anteil anderer Muttersprachen überschätzt.<sup>2</sup> Wie die Darstellung der Werte zeigt, sind die Meinungen der Befragten breit gestreut. Anhand weiterer Analysen kann geprüft werden, ob es Gruppen von Befragten gibt, die sich in ihrer Schätzung unterscheiden. Es geht also darum, Faktoren zu identifizieren, die im Zusammenhang mit der Schätzung stehen, z. B. soziodemographische Faktoren wie Geschlecht, Alter und Bildungsabschluss oder andere Einstellungen wie das politische Interesse<sup>3</sup> oder die Lebenszufriedenheit<sup>4</sup> (vgl. Adler/Plewania i. Ersch.). Abbildung 3 illustriert das Ergebnis einer solchen Analyse und zeigt, welche Faktoren im Zusammenhang mit der abgefragten Schätzung stehen.

### Lesehinweis zum Regressionsmodell

Durchgeführt wurde eine lineare Regression mit den Ergebnissen der Schätzung als abhängige Variable. Auf der y-Achse der Abbildung werden die Variablen gelistet, für die im Modell geprüft wird, ob und was für einen Einfluss sie auf die abhängige Variable haben. Die signifikanten Variablen – d.h. Variablen, die in einem bedeutsamen Zusammenhang mit der abhängigen Variable stehen, – sind entsprechend der Höhe ihrer Signifikanz mit einer gewissen Anzahl von Asterisken gekennzeichnet (wie allgemein üblich, wird hier die folgende Notation entsprechend des Signifikanzniveaus ( $p$ ) verwendet: \* =  $p < 0,05$  ist signifikant, \*\* =  $p < 0,01$  ist hoch signifikant, \*\*\* =  $p < 0,001$  ist höchst signifikant). Bei kategorialen Variablen (z.B. Geschlecht) wird bei der Berechnung eine Ausprägung der Variable als Referenzkategorie verwendet (z.B. ist bei Geschlecht die Referenzkategorie „weiblich“; die Referenzkategorie ist jeweils angegeben). Kontinuierliche Variablen (z. B. Alter) sind aufsteigend sortiert.

Schlagen die Variablen von der mittleren Linie nach links aus (in blau), bedeutet das, dass die Schätzung im Vergleich zur Referenzkategorie (bzw. nach aufsteigender kontinuierlicher Variable) eher niedriger ist, für Variablen(ausprägungen) nach rechts (grün) ist es andersherum, die Schätzung fällt höher aus. Lesebeispiel: Befragte aus den nördlichen Bundesländern schätzen den Anteil im Vergleich zu Befragten aus den mittleren Bundesländern eher niedriger ein. Dieser Zusammenhang ist allerdings nicht signifikant. Befragte aus den südlichen Bundesländern schätzen den Anteil im Vergleich zu Befragten aus den mittleren Bundesländern auch eher niedriger ein. Dieser Zusammenhang ist signifikant.

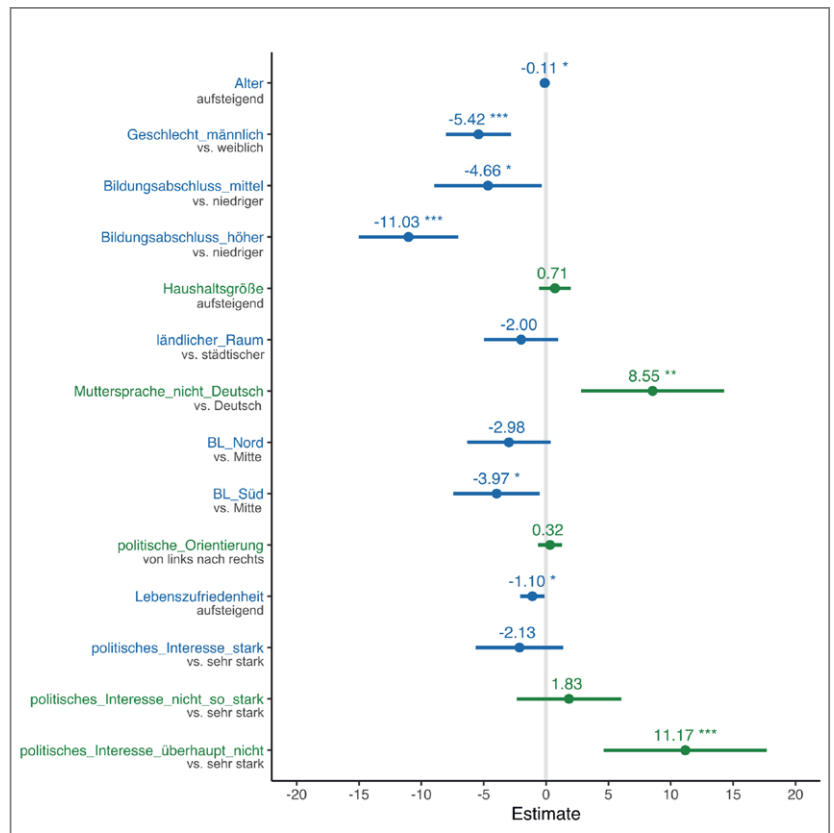


Abb. 3: Regressionsmodell mit Einflussfaktoren für den geschätzten Anteil anderer Muttersprachen als Deutsch (Deutschland-Erhebung 2017)<sup>5</sup>

Demnach unterscheiden sich die Schätzungen der Befragten insbesondere in Abhängigkeit ihres Geschlechts, des Bildungsabschlusses und ihres politischen Interesses, aber auch nach Alter, nach Muttersprache, zum Teil nach Wohnort und nach Lebenszufriedenheit. Männer schätzen den Anteil an Menschen in Deutschland mit einer anderen Muttersprache als Deutsch niedriger als Frauen. Bezüglich des Bildungsabschlusses liegt die Schätzung der Befragten niedriger, je höher ihr Abschluss ist. Je älter die Befragten sind, desto niedriger fällt ihre Schätzung aus. Das politische Interesse der Befragten steht in einem starken Zusammenhang mit der Schätzung.<sup>6</sup>

In Abbildung 4 wird der Zusammenhang zwischen dem politischen Interesse und der Schätzung im Detail dargestellt. Die Verteilung der geschätzten Werte wird jeweils für die Teilstichproben nach der Ausprägung des angegebenen politischen Interesses gezeigt. Unten ist die Verteilung derjenigen zu sehen, die angeben, sich sehr stark politisch zu interessieren. Ganz oben ist der Graph für diejenigen Befragten zu sehen, die angeben, sich überhaupt nicht politisch zu

interessieren.<sup>7</sup> Die Verteilung dieser beiden Gruppen unterscheidet sich deutlich: Während die sehr stark politisch interessierten Befragten den Anteil anderer Muttersprachen als Deutsch in Deutschland recht mehrheitlich eher niedrig einschätzen, schätzen die überhaupt nicht politisch interessierten Befragten den Anteil höher. Die höchste Spitze der Verteilung liegt für letztere bei etwa 28%; daneben gibt es eine weitere, wenn auch flachere Spitze, bei etwa 70%. Nimmt man die Schätzungen der beiden anderen Teilstichproben (nicht so starkes politisches Interesse und starkes politisches Interesse) hinzu, kann man die Tendenz für den Zusammenhang dieser beiden Variablen wie folgt zusammenfassen: Je stärker das politische Interesse ist, desto niedriger fällt die Schätzung des Anteils anderer Muttersprachen in Deutschland aus.

Zu den Sprachen in Deutschland gibt es wenige gute Zahlen. Die vorhandenen Zahlen zeigen deutlich, dass Deutsch mit Abstand am meisten gesprochen wird. Über die darüber hinaus in Deutschland gesprochenen Sprachen gibt es bislang keine guten Statistiken. Ihr Anteil liegt wohl zwischen 9,7% und 14,7%. Eben diesen Anteil schätzen die Menschen in Deutschland jedoch deutlich höher ein, nämlich auf im Mittel etwa 30,3% Prozent. Die Einschätzung dieses Anteils hängt zusammen mit soziodemographischen Faktoren wie Geschlecht, Alter und Bildungsabschluss, aber auch mit Einstellungen der Befragten wie dem politischen Interesse und der Lebenszufriedenheit. ■

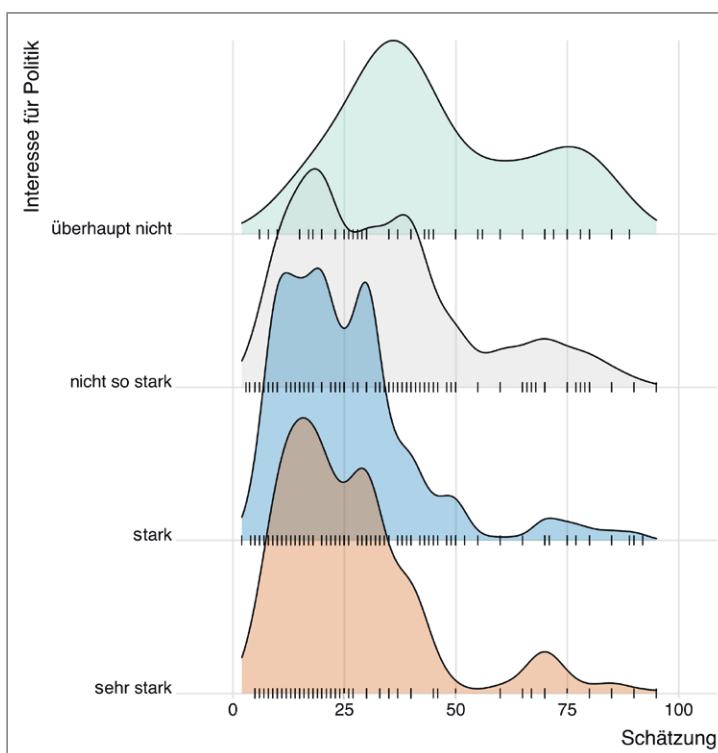


Abb. 4: Zusammenhang politisches Interesse und geschätzter Anteil anderer Muttersprachen als Deutsch (Deutschland-Erhebung 2017)

## Literatur

- Adler, Astrid/Ribeiro Silveira, Maria (i. Vorb.): Warum wir so wenig über in Deutschland gesprochene Sprachen wissen. Spracheinstellungen als Erkenntnisbarriere. In: Diskurs Kindheits- und Jugendforschung (4/2021), Themenheft „Perspektiven von Kindern und Jugendlichen auf sprachliche Diversität und Sprachbildungsprozesse“.
- Adler, Astrid/Plewnia, Albrecht (i. Ersch.): Was denken linguistische Laien über Sprache? Ergebnisse einer aktuellen Repräsentativerhebung zu Spracheinstellungen in Deutschland. In: Hoffmeister, Toke/Hundt, Markus/Naths, Saskia (Hg.): Laien, Wissen, Sprache. Theoretische, methodische und domänenspezifische Perspektiven. (= Sprache und Wissen (SuW) 50).
- Adler, Astrid (2019): Sprachstatistik in Deutschland. In: Deutsche Sprache 3, 19, S. 197-219.
- Folge 1 = Adler, Astrid/Ribeiro Silveira, Maria (2020): Spracheinstellungen in Deutschland – Was die Menschen in Deutschland über Sprache denken. Sprache in Zahlen: Folge 1. In: SPRACHREPORT 4/2020, S. 16-24.
- Folge 2 = Adler, Astrid/Ribeiro Silveira, Maria (2021b): Welche Dialekte werden in Deutschland gesprochen? Sprache in Zahlen: Folge 2. In: SPRACHREPORT 1/2021, S. 52-53.
- Folge 3 = Adler, Astrid/Ribeiro Silveira, Maria (2021a): Welche Dialekte werden in der Familie weitergegeben? Sprache in Zahlen: Folge 3. In: SPRACHREPORT 2/2021, S. 42-44.
- Petschel, Anja/Will, Anne-Kathrin (2020): Migrationshintergrund – ein Begriff, viele Definitionen. Ein Überblick auf Basis des Mikrozensus 2018. In: WISTA 5, S. 78-90.

## Anmerkungen

- <sup>1</sup> In der Abbildung werden für die *Deutschland-Erhebung 2017* lediglich die Sprachen explizit ausgewiesen, die von mehr als 30 Befragten angegeben werden.
- <sup>2</sup> Es gibt einen Wert in Deutschland, der auf einem ähnlichen Niveau liegt: der Migrationshintergrund. Für dieses – durchaus umstrittene – statistische Konstrukt wurde im Jahr 2018 ein Wert von 25,5% ermittelt. Dieser Wert schwankt allerdings sehr stark in Abhängigkeit davon, wie man den Migrationshintergrund genau definiert (siehe Petschel/Will 2020).
- <sup>3</sup> Nach dem politischen Interesse wird in der Umfrage des SOEP (siehe Folge 1) wie folgt gefragt: „Einmal ganz allgemein gesprochen: Wie stark interessieren Sie sich für Politik? Sehr stark, stark, nicht so stark oder überhaupt nicht?“
- <sup>4</sup> Nach der Lebenszufriedenheit wird in der Umfrage des SOEP (siehe Folge 1) wie folgt gefragt: „Nun möchten wir Sie noch nach Ihrer Zufriedenheit mit Ihrem Leben insgesamt fragen. Wie zufrieden sind Sie gegenwärtig, alles in allem, mit Ihrem Leben?“ (0 „ganz und gar unzufrieden“ bis 10 „ganz und gar zufrieden“).
- <sup>5</sup> Die Variable über den Wohnort der Befragten liegt in drei Ausprägungen vor. Der Einteilung liegt das Bundesland des Wohnorts zugrunde. BL\_Nord beinhaltet die Bundesländer Niedersachsen, Bremen, Schleswig-Holstein, Hamburg, Sachsen-Anhalt, Mecklenburg-Vorpommern, Brandenburg und Berlin, BL\_Mitte umfasst Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, Hessen, Thüringen und Sachsen und BL\_Süd besteht aus Baden-Württemberg und Bayern.
- <sup>6</sup> Die allgemeine Verteilung der Antworten auf die Frage nach dem politischen Interesse (siehe Endnote 3) ist wie folgt: 15,4% sehr stark, 38,1% stark, 37,0% nicht so stark, 9,5% überhaupt nicht.
- <sup>7</sup> Die Striche unter der jeweiligen Kurve bilden die Frequenz der jeweiligen Nennungen ab. ■