

Germany's micro census of 2017:  
The return of the language question.

a working paper by Astrid Adler

## 1. Introduction<sup>1</sup>

For the first time in almost eighty years, a German census has reincorporated a question eliciting information about the population's language. Presumably, this question will be repeated in years to come during the annual German micro census, a representative survey that has been carried out since 1957 (it is comparable to the American Community Survey). Around 830,000 people in approximately 370,000 households (i.e. approximately 1% of the total population) take part in this survey.<sup>2</sup> The micro census elicits sociodemographic data, data on personal and marital status, housing, education and the work situation. The respondent is obliged to answer by law. The mere fact that a language question reappeared in a German micro census is to be welcomed, especially from a linguistic standpoint. Unfortunately though, the question has some shortcomings that can already be identified before the publication of results. Usually, the German Office of Statistics (*Statistisches Bundesamt*) publishes first results in the second half of the following year.

There are countries that include questions on the population's language in their census and there are others where this is not or no longer the case. Questions on language were part of several population censuses in the 19<sup>th</sup> century (e.g. in Belgium, Austria, Prussia and Switzerland). At the time the construction of the (respective) population's national identity was of great importance in Europe. Accordingly, population statistics were developed to showcase these national identities and other relevant concepts of nation building. Statisticians from several countries formed an international association of statistics (cf. e.g. Labbé 2009) that served as a platform to discuss questions regarding the content and methodology of population statistics. The questions under debate also concerned the language and identity of the population, issues that remain relevant today, e.g. for the current language question in the German micro census. Therefore, the following section of this paper will give an overview about general aspects and issues of language questions in censuses, most of which took place in the 19<sup>th</sup> century. A discussion of the current language question in the German micro census 2017 follows in Section 3.

## 2. Language questions in censuses

To survey a population's language one first needs to decide on how to adequately express the language concept that underlies the wording of the question. There was a lot of debate on labelling issues throughout the 19<sup>th</sup> century: Should one ask for the "conversational language" (*Umgangssprache*), for the "family language" or for the "mother tongue" (cf. Arel 2002)?<sup>3</sup> Richard Böckh for example, a Prussian statistician, advocated using the label "mother tongue" in the question, also because it played a major role in his conceptualisation of national statistics (cf. Leuschner 2004). In his opinion someone's "mother tongue" was closely related to his or her national identity; and while, in his opinion, "national identity" was a concept too unstable to be asked for, the

---

<sup>1</sup> The author thanks the "Programmbereich Sprache im öffentlichen Raum", especially Albrecht Plewnia, Rahel Beyer, and David Hünlich for special assistance on the English version of the text.

<sup>2</sup> Cf. Statistisches Bundesamt (no specified year)

<sup>3</sup> For a detailed description of language questions in censuses see e.g. Humbert/Coray/Duchêne 2018, this is also the most recent compendium about language questions in censuses worldwide, cf. also Duchêne/Humbert 2018, i.e. the Special Issue of the International Journal of the Sociology of Language 2018, Number 252; Kertzer/Arel 2002 and Arel 2002; Lieberman 1966 and de Vries e.g. 2006 also provide overviews on the subject.

concept of the “mother tongue” was stable enough and well suited to serve as a proxy, i.e. an indirect category.

With regard to a language question's wording it is not only relevant to look at the language concept but also the question's design including the proposed answers. The latter should be analysed carefully as they may reveal undercurrents of ideology and language politics. For example, when a question offers a closed list of languages as possible answers then their selection and order indicates which languages are seen as worthy of selection, and which are not. It is also essential which languages are being labelled and what labels they receive in the question as well as in the analysis of the question, e.g. as single languages or as part of a general category (e.g. “European languages”). The label can already imply an evaluation – even if it is unintentional. In the analysis of the Austrian census of 2001, for example, the labels used to refer to general categories were clearly judgmental or insensitive, e.g. Native American languages were labelled as “Indian” languages (“Indianersprache”), the language Hindi was labelled as “Indian” (“Indisch”) and Westafrican languages were called “indigenous” with a clearly colonial undertone (“westafrikanische Eingeborenen-sprachen”; cf. Busch 2010:14, Busch 2016:6f.). The decision on whether to propose a certain language as an answer category should of course depend on their probable frequency of occurrence. If in Germany, for example, there is allegedly no one or only very few people speaking Nahuatl then the probability of a respondent entering such a language as an answer in the survey is rather small. Respectively, it is quite reasonable to not propose the language as a possible answer. However, if contrary to all expectations there is someone who wants to give this answer, it should be possible, e.g. in an open answer box. Altogether, this shows how by using words it is possible to exercise power in the sense of Bourdieu (1991), i.e. to create social realities, to form collective identities or to ignore them.

Because language is strongly related to people's identity, it is also strongly associated with social groups and their perception. From a social and political point of view, this is quite important as these groups, i.e. their size and proportion, become highly visible when they enter the census. Thus, questions on language are an instrument to register language as a characteristic of someone's collective identity; therefore they play a substantial role in the perception of minority and majority constellations. Through the language questions in a census, minorities and majorities (i.e. as social groups or as parts of other social groups) can be strengthened or weakened. Also, public perception or creation of groups can be triggered, accelerated or suppressed. Via the census count the group as such becomes visible, and its portion of the population becomes evident. Depending on the particular interests of the concerned social group and on whether the results are beneficial or detrimental, this can be desirable – or just the opposite. In Burundi, Mauritania and Pakistan, for example, ethnic categories were excluded and census counts prohibited for fear that a certain social group might be represented as a minority and it may consequently lose its political power (cf. Kertzer/Arel 2002:23-24; a current example is the US-census with a discussion on the question on citizenship which has been put forward by the Trump-administration). However, for certain language minorities for instance in the case of Low German, it can be beneficial from a language policy point of view to regularly survey the number of their speakers, e.g. to monitor the vitality of their language or to review measures of revitalisation (cf. Adler et al. 2016; also for Basque cf. e.g. Urla/Burdick 2018). The examples illustrate the potential of such counts or censuses for conflicts of interests, e.g. when the concerns of involved minorities and majorities collide. This again can be a reason for exploiting or even manipulating the question, the counts and their analyses (e.g. in Macedonia around 1900, cf. Kertzer/Arel 2002:19ff.).

Usually, a question on language focusses on several languages, i.e. multilingualism is crucial. From a statistical point of view, multilingualism represents a difficulty, for it cannot be measured as easily as monolingualism. While monolinguals can be assigned to only one answer and the corresponding analysis is therefore relatively straightforward, for multilinguals who give several answers to a question on language this is not the case. The responding individual is connected to several responses. Therefore the analysis is not as easy, e.g. to crosstab variables or characteristics respectively. Possible reactions to this statistical problem are the prohibition of multiple answers or to propose such answers but to then reduce them to one single answer in the process of the analysis, i.e. to de facto only evaluate one answer (for example in the analysis of the Austrian census 2001, c.f. Busch 2010).

Statistics represent counts and measures of objects, and they generate knowledge about these objects. Correspondingly, statistics are being perceived as means to produce knowledge and truth (cf. Urla 1993). A population census is a special kind of count or measure, because it surveys a large portion of the population or even all of it and it therefore generates a supreme type of statistical data. For this reason, any census has special status and is seen as "[t]he ultimate register of the 'truth'" (Arel 2002:94). Population censuses are considered to legitimately produce truth, and are also exploited as a means to yield legitimacy (cf. Arel 2002:115: "The census is the prized instrument to "officialize" the territorial distribution of language communities."). This also means that a concern of these counts not only lies in the mere count but also in the "right", i.e. most favourable, results. If the outcome is unclear then oftentimes corresponding questions are withdrawn. In Lebanon, for example, there has not been a census since 1932 because it is obvious that results would not correspond with the expectations of reality and therefore might cause political agitation. Similarly, in Belgium the census does not ask a question on language since 1960, because the last results were cause for trouble (cf. Lévy 1960, Kertzer/Arel 2002:31, Arel 2002:105f.).

Possibilities to influence or even manipulate the counts exist on almost all levels of the process (cf. e.g. Kertzer/Arel 2002:27ff., Arel 2002). For example, respondents can influence the results by the kind of answer they give (though the refusal to answer is sometimes liable to prosecution). In several countries there have been repeated campaigns intended to sway the answers to census questions into a certain direction. In the 1990s, in Canada, a new answer category to the question on ethnic identity emerged for this reason: the Toronto Star's campaign "Call Me Canadian!" (cf. Kertzer/Arel 2002:16) called upon respondents to fill "Canadian" into the open answer box because the category had not been provided in the questionnaire of previous years. Many followed the campaign's appeal, and because open answers with a certain number of respondents have to be added to the set of proposed answers, the new category Canadian was then included in the 1996 census (in the 1991 census 3.3 % of respondents specified "Canadian" in the open-answer box, in 1996 24.1 % ticked the respective category). Based on such and similar incidents the fuzziness of censuses and referenda is sometimes criticised (cf. Lévy 1960). The census in Scotland represents another example; this time, the influence was not exercised by those who give the answers, but by those who are responsible for the census' design. Approximately three months before the census that included a language question, the Scottish government launched a website to "inform" the population, and to facilitate the "right" answer ([www.ayecan.com](http://www.ayecan.com), cf. Sebba 2016). Such information campaigns can have good and legitimate reasons, but in any case, they will influence results. The examples show that the social and political context and the corresponding social desirability are relevant to a question on language; in particular, when the named languages are visible as a feature of collective identity of a certain group which plays a role in current debates. An even greater influence probably comes from those

who determine the language question's design and therefore ultimately highlight certain issues and facilitate certain results.

This shows how important it is to document all processes and discussions that lead to the emergence and the development of a language question, beginning with the decision to pose such a question in the first place (i.e. the history of the development of a language question) and ending with the evaluation of the question (i.e. the decisions involved in the reduction of complexity in the question's evaluation). Only this kind of documentation makes it possible to retrace, assess and review the whole process or at least part of it. The decision to pose a question on language or not already represents such a process (e.g. Switzerland's census has been continuously posing such a question, in France there has never been a question on language and in Belgium the language question was removed from the census in 1960). The development of the respective language questions in the English and Scottish census have been well documented and are available. On the basis of documentation one can reconstruct decisions which are partly disputable (cf. Sebba 2016, 2017). Until this day, such documentation does not exist for the German language question put into the micro census in 2017. There merely is the underlying legislative text, which is publicly available (see below, section 3.9).

The manner in which results are presented is also highly significant. It should be self-evident that a language question should be evaluated with special diligence. Nevertheless, the Office for National Statistics in England released an erroneous statistic where the dominant language of the household had been labelled as "main language" (cf. Sebba 2017). By way of how results are selected and presented, the "census designers" influence the discussions taking place in public. As time goes by, certain results may no longer match the context of a question but they may still be represented (e.g. by the media) as legitimate truth because of their status as census' results.

### 3. The language question in the German micro census 2017

It has been approximately eighty years since the last time a question on the population's language was part of a German census: in 1939, the census asked for the participants' "mother tongue". In the early 19<sup>th</sup> century the language of the population – at the time termed as the "family language" – was noted in registers in parts of Prussia (e.g. 1825 in Königsberg and Oppeln, 1830 in Posen and Westpreußen; cf. Labbé 2003). Back then, the intention mainly was to examine the effectiveness of community politics (cf. Labbé 2003, see also Glück 1979). In 1861, the Royal Prussian Statistical Bureau of Statistics in Berlin (*Königlich Preußisches Statistisches Bureau*; renamed in 1905 as *Königlich Preußisches Statistisches Landesamt*, and in 1918 as *Preußisches Statistisches Landesamt*) carried out a unified census including a question on the population's "mother tongue" in all of Prussia. After that, in the censuses until 1890, there no longer were questions on language. From 1890 on, the question on the population's "mother tongue" was then frequently repeated (in the censuses of 1900, 1905, 1910, 1925, 1933, 1939; for the questionnaires, i.e. *Zählkarten*, and individual results see the detailed version of this text, Adler in prep.; cf. also *Königlich Statistisches Bureau 1893ff.*). The question in these censuses was not altered much. However, in 1905 and 1910 as well as in 1925 a question on proficiency in German was added for participants with another first language than German. This additional question mirrors the language policies of the time, i.e. how other language minorities and the multilingual population are being dealt with. For the regional administrations the additional question was important as they could assess the effectiveness of

efforts to germanize the population with another native language than German. The changes in language policies are also reflected in the explanations of the questions as well as in the evaluations of the results and their presentation.

### 3.1 The question on language and the wording of the language question

The language question in Germany's micro census 2017 reads as follows: "Welche Sprache wird in Ihrem Haushalt vorwiegend gesprochen?" ("Which language is being spoken predominantly in your household?"). The question is part of a set of questions on citizenship and the duration of stay in Germany which follows a set of questions on education. This thematic set of migration-related questions begins with a question determining whether the participant's place of birth lies within or outside the borders of Germany.<sup>4</sup> The latter answer then serves as a filter leading to more questions, e.g. about the actual place of birth.<sup>5</sup> The string of questions following the filter ends right before the language question, which means that apparently all respondents have to answer it regardless of their origin.

Figure 1 shows the exact format and wording of the question and the proposed set of answers (the underlying pdf-document is a sample of the questionnaire which could be found on the website of the statistical office of Hamburg and Schleswig-Holstein until recently):

---

<sup>4</sup> There are different versions of the questionnaire online. They only differ minimally in the aspects discussed here. One of the versions lists two questions at the beginning, namely "Have you been born in Germany?" and "Is your place of birth located inside of the current state territory of Germany?", another version simply asks "Have you been born inside of the current state territory of Germany?".

<sup>5</sup> The questions are the following: „In which current state have you been born?“, "When did you move (for the first time) to Germany?" and "What was the (main) motive to move to Germany?", for the latter two there is an explanation at the beginning: "If you are not born inside of the current territory of Germany." The answers to these three questions can respectively be given by up to five respondents.

1. Person

159 Welche Sprache wird in Ihrem Haushalt vorwiegend gesprochen ?

Deutsch	01	<input type="checkbox"/>
Nicht Deutsch, und zwar		
... Arabisch	02	<input type="checkbox"/>
... Englisch	03	<input type="checkbox"/>
... Französisch	04	<input type="checkbox"/>
... Italienisch	05	<input type="checkbox"/>
... Polnisch	06	<input type="checkbox"/>
... Russisch	07	<input type="checkbox"/>
... Spanisch	08	<input type="checkbox"/>
... Türkisch	09	<input type="checkbox"/>
... eine sonstige europäische Sprache	10	<input type="checkbox"/>
... eine sonstige afrikanische Sprache	11	<input type="checkbox"/>
... eine sonstige asiatische Sprache	12	<input type="checkbox"/>
... eine sonstige Sprache	13	<input type="checkbox"/>

Figure 1 The question on language in Germany's micro census 2017<sup>6</sup>

However, there also was a version of this questionnaire where the language names in the list of possible answers were not capitalized. An official final version of the micro census' questionnaire could not be located.

The present question does not ask about the "mother tongue" but about the "language predominantly spoken in the household" without explicitly calling it the "household language", i.e. how one (predominantly) speaks at home. There are other concepts that could have been asked for, such as the "main language" (e.g. in England's census in 2011), "mother tongue" (e.g. in Finland and historically in Prussia) or "conversational language" (i.e. *Umgangssprache*, e.g. until 2011 in Austria). By choosing the concept of the "household language" other concepts such as the "mother tongue" or "conversational language" that may be difficult to explain and comprehend can be avoided. An objective and generally more intelligible notion like "household language" and its description are certainly not a bad choice since the question appears on the German micro census for the first time for most participants.

It is not clear, however, whether any general explanations were provided or whether there were specific ones for the language question, as it was the case for some language questions in the censuses of the early 20<sup>th</sup> century. One has to distinguish between explanations that are only meant for the interviewers and explanations that directly address the respondents. For a question that had not appeared on German censuses for almost 80 years, an explanation would have been helpful. There can be several reasons why it was not done. Perhaps the question was thought to be easy, or the influence of explanations on the response was deliberately avoided.

### 3.2 The list of possible responses

The possible answers are presented in a closed format and they are graphically divided into two parts. They are closed-format questions because all answer options are presented in a fixed list and there is no open possibility to express an individual response (for the lack of an open answer, see below 3.3; for the discussion of the difference between closed- and open-format questions especially for language categories cf. Plewnia/Rothe 2012:27-33). The first possible answer is "Deutsch"

<sup>6</sup> Cf. exemplary questionnaires of the Statistical offices of Hamburg and Schleswig-Holstein: [https://www.statistik-nord.de/fileadmin/Dokumente/Erhebungen/Mikrozensus/MZ-Fragebogen\\_2017\\_Muster.pdf](https://www.statistik-nord.de/fileadmin/Dokumente/Erhebungen/Mikrozensus/MZ-Fragebogen_2017_Muster.pdf), last accessed on 18.1.2018.

('German'); according to its position right at the beginning of the answer options and the graphical division from all other possible answers its character as a default response is emphasized. The following answers are listed below the words "Nicht Deutsch, und zwar ..." ('not German, namely ...'). This division can be interpreted as an evaluative judgement in that German is the normal case or default and all other languages are the opposite.

Another significant problem with the question immediately follows from this arrangement: the question does not allow for multilingualism, since only one single household language can be named. This raises several questions about the social desirability of multilingualism and multilingual speakers. Respondents speaking several languages at home (e.g. German and another language like German and Polish, or two or more languages other than German such as Turkish and Kurdish) are forced to make an unnatural and incorrect decision (for further explanation on this subject cf. sections 3.4 and 3.6 as well as section 4).

The following languages are alphabetically (i.e. neutrally) listed as 'not German': Arabic, English, French, Italian, Polish, Russian, Spanish and Turkish followed by three general language categories "another European language", "another African language" and "another Asian language." The list concludes with an undifferentiated category for any remaining languages under the label "another language". This option is closed so that participants cannot further specify their answer. The languages listed by name in the answer options include typical migrant languages, i.e. languages such as Russian or Turkish spoken mainly by foreign migrants and citizens with a migrant background (*Personen mit Migrationshintergrund*) living in Germany. The list also contains languages such as English and French that are not amongst the most frequently spoken native or heritage languages of the migrant population in Germany, but which enjoy more social prestige. The current list with its structure, its answer options and combination of languages is problematic for a number of reasons that deserve consideration.

The German Federal Statistical Office (*Statistisches Bundesamt, destatis*) frequently publishes statistics listing the number of foreign nationals living in Germany based on the Central Register of Foreigners (*Ausländerzentralregister*). Table 1 shows the distribution of different groups of foreigners by country and region in 2016. Foreigners are defined here as individuals who do not hold German citizenship. Having a migrant experience could be defined in broader terms, however: Second generation Germans with a migrant background (*Personen mit Migrationshintergrund*) or Germans who hold dual citizenship are not included in this statistic (cf. Statistisches Bundesamt 2016:6). For this reason, many of those with a native or heritage language other than German remain invisible in the statistics. Extrapolating languages from a limited statistic such as in Table 1 is unsatisfactory and problematic not only for purposes of linguistic research but also in terms of migration planning and policy making. It is precisely for these reasons why it would be desirable to collect reliable data and learn more about the actual languages spoken in Germany today. Yet, in spite of its weakness, Table 1 can serve to control how many migrant languages with an important role in Germany today are actually covered by the language question in the micro census.

Turkey	1,492,580
<i>Territory of former Yugoslavia</i> (Bosnia and Herzegovina, Kosovo, Croatia, Macedonia, Montenegro, Serbia, Slovenia)	1,135,960
Poland	783,085
<i>Territory of the former Soviet Union</i> (Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Moldova, Russian)	654,390



Federation, Tajikistan, Turkmenistan, Ukraine, Uzbekistan)	
Syria	637,845
Italy	611,450
Romania	533,660
<i>Territory of former Serbia and Montenegro</i>	475,305
Greece	348,475
Croatia	332,605
Bulgaria	263,320
Afghanistan	253,485
Russian Federation	245,380
Iraq	227,195
Serbia (with and without Kosovo)	223,100
Kosovo	202,905
Hungary	192,340
Austria	183,625
Bosnia and Herzegovina	172,560
Spain	163,560
the Netherlands	149,160
Ukraine	136,340
Portugal	136,080
France	130,915

**Table 1** Foreign population of Germany on December 31<sup>st</sup> 2016, absolute frequencies ordered by size, multiple listings due to areal categorizations (e.g. former Yugoslavia); extracted from Table 3b in *Statistisches Bundesamt* (2016)

The largest seven groups of foreigners living in Germany with more than 500,000 members come from Turkey, former Yugoslavia, Poland, the former Soviet Union, Syria, Italy and Romania. Using Ethnologue,<sup>7</sup> a resource that provides estimates of national linguistic distributions, we can generate a list of the most important indigenous languages in these seven countries and regions:

- Turkey: Turkish, Northern Kurdish, Zazaki, North Levantine Arabic, Kabardian
- Former Yugoslavia: Albanian, Aromanian, Bosnian, Croatian, Hungarian, Macedonian, Montenegrin, Romani, Serbian, Slovene
- Poland: Polish, German
- Former Soviet Union: Armenian, Azerbaijani (North, South), Bashkort, Belarusian, Chechen, Chuvash, Estonian, German, Georgian, Kazakh, Kyrgyz, Latvian, Lithuanian, Romanian, Russian, Tajiki, Tatar, Turkmen, Ukrainian, Uzbek
- Syria: North Levantine Arabic, Standard Arabic, Northern Kurdish, Najdi Arabic
- Italy: Italian, Napoletano-Calabrese, Sicilian, Venetian, Lombard
- Romania: Romanian, German, Hungarian, Romani

The list with 42 language labels immediately shows that there is no direct correspondence of national origin and the family or household language. Of course, the labels should be interpreted with caution. Language labels such as “Bosnian”, “Croatian” and “Serbian” or “Sicilian” and “Venetian” which could also be characterized as dialects of Italian result from Ethnologue’s policy to scrupulously document speakers’ political sensitivities when categorizing languages and dialects.

<sup>7</sup> Cf. <https://www.ethnologue.com/>. It should be noted though, that the languages and their respective number appearing in this compendium can certainly be discussed rather severely.

However, even if all dichotomies between languages and dialects are interpreted as liberal as possible, this extensive list contains more than thirty languages that are mutually incomprehensible. Of these thirty languages only six (German, Arabic, Italian, Polish, Russian and Turkish) are listed as direct options in the German micro-census. Many important Northern, Eastern and South-Eastern European national languages such as Estonian, Latvian, Lithuanian, Romanian, Hungarian and the languages of former Yugoslavia simply do not appear, although a considerable amount of citizens of these countries live in Germany today – many of them from EU-member states. Important minority languages such as Kurdish (in Turkey), Romani (in South-Eastern Europe) or Chechen (in Russia) are missing. Regional variants of Arabic that may be mutually incomprehensible also do not find consideration.<sup>8</sup>

While many languages that would be expected are missing in the language question of the German micro census, some are listed that would not be expected. French and English are not indigenously spoken in any of the prominent countries and regions of migrant origin listed in Table 1. French citizenship stands at the end of the list of the foreigner statistic with approximately 131,000 citizens – even less than the foreign nationals from the Netherlands and Portugal living in Germany. Although there are other countries with French as an official language (e.g. Belgium and many African states), it is completely speculative whether French eclipses the use of Dutch or Portuguese as a household language in Germany. The same can be said for English, which is not spoken indigenously in any of the countries in Table 1. So why does the micro census list French and English as possible answers? A possible reason is that in Germany these languages represent traditional foreign languages in school, and are therefore considered to be languages of education and prestige (cf. Gärtig/Plewnia/Rothe 2010). Other reasons may be that they are perceived as neighbouring languages or languages of status due to their official use in the European Union. These reasons probably made French and English prominent options for a list of languages that are somehow “important” in Germany. However, these answer options are folkloristic, at best, and biased, at worst. In any case, they undermine the solid inquiry into the linguistic habits of the German population that is direly necessary. In summary, a comparison with the available statistics on the foreign population shows that the languages listed as possible answer options in the micro census neither correlate with the most frequently spoken migrant languages in Germany, nor do they contain the most reasonable candidates for “household languages”.

The answer options representing individual languages are followed by three more general categories, namely “another European language”, “another African language” or “another Asian language” which are no longer ordered by alphabet but by according to their geographical distance. The first problem with these general categories is quite obvious: they eliminate all detail that could be recorded in an open answer category. None of the information behind the answer option can be recovered for analysis. The second problem is that the categories themselves are highly ambiguous: For example, should a respondent want to name Kurdish as his or her household language he would have to decide whether Kurdish is “another European language” or “another Asian language”. Perhaps the option “another language” would then be most appropriate – which is the final category provided in the list, again as a closed option that cannot be further specified. However, regardless of the answer choice the response “Kurdish” would get lost and is irretrievable.

---

<sup>8</sup> Another project at the Institut für Deutsche Sprache, “Deutsch im Beruf” (German at Work), documents language repertoires of refugees and finds that most Arabic speakers will not specify which dialect of Arabic they speak unless the issue is directly addressed.

But the problems are not restricted to languages that are new to Germany. Several indigenous languages would have similar problems. Low German is an officially recognized regional language in Germany and it is protected by the European Charter for Regional- and Minority Languages. If a respondent wants to indicate that this language is his or her predominant household language<sup>9</sup> then this person would have to check one of the general categories, most likely “another European language”. The specific answer “Low German” would therefore fall into an aggregate category with languages such as Portuguese or Czech, and get lost for analysis in spite of its particular status in Germany.

German Sign Language (Deutsche Gebärdensprache, DGS) presents a similar case. The language was legally recognized in 2002 by the Act on Equal Opportunities for Persons with Disabilities (*Behindertengleichstellungsgesetz*, BGG) and is being used by approximately 80,000 to 200,000 Germans. The number differs depending on the source, and an official statistic is not available. In spite of its status, the language is not listed as an option in the language question. If DGS is predominantly used in the household, then depending on the respondent's personal definition the response might fall into the general category “another language” or into the category “German” (cf. the description of DGS in the BGG as a form of communication of German). In both cases the detailed answer gets lost.

All previous examples show that the missing granularity of the response options is a major issue for the language question in its present form. Altogether, the survey construct is unable to quantify minority languages in Germany – both indigenous and new to the country. This has to be taken into account when results are assessed and interpreted. An appropriate classification of the general categories' significance, particularly the last one, will be very difficult to provide, especially if many answers fall into these categories. If the portion of general answers eventually exceeds expectations, and they cannot be interpreted merely as the “remainder” of answers breaking down the answers will be desirable. However, the current closed-category-design of the general answer categories does not permit a more fine-grained description. In contrast, an open answer format would not have adversely affected the analysis. If detailed answers were principally available then general categories could still be generated, rearranged and decomposed, a posteriori. A possible reason why general categories were used for data collection may have been to avoid an exhaustive list of possible languages and especially the complex task of analysing an open answer category (cf. below 3.3).

The response options and the classification of some of them as single-language-answers and others as general categories also indicate which languages apparently shall or may be counted. While selected national languages are found in the single-language-list, minority languages of all types are omitted. Even officially recognized minority and regional languages with long standing roots in Germany are ignored. The language question in its present form exacerbates the disadvantage of minority languages vs. official national languages. Because people freely commit to minority languages, counting their speakers is difficult. Usually, a lack of reliable figures perpetuates their lack of recognition. On the other hand, it is precisely this type of recognition that may be unwanted (cf. section 2). From this vantage point it would also be no accident that the current phrasing of the language question completely ignores the regional varieties of German, i. e. the dialects. Currently, there is no reliable data on the use of German dialects at home or in German households respectively (cf. representative data on subjective dialect competence in Gärtig/Plewnia/Rothe 2010

---

<sup>9</sup> For current figures on numbers of speakers cf. Adler et al. (2016).

or data in Schmidt 2017). Possibly, dialects have not been included in the micro census question to avoid complexity and because all those who predominantly speak a dialect at home are thought to select the first answer option, i. e. "German". But in contrast to the census of 1925 which provided specific instructions on how to deal with dialect use (namely, to ignore them and not to treat them as native languages<sup>10</sup>) the new language question no longer even mentions the existence of regional variants of German.

### 3.3 The missing open answer option

The micro census does not provide an open answer in the language question, i.e. an option for all those who cannot find their household language represented by the existing answer suggestions, and want to offer a distinct response, e.g. Kurdish, German Sign Language or Low German. Since there is no possibility to provide no answer at all and respondents are legally obligated to reply, they can only select the least ill-fitting category. Often this will be the last category ("another language"). The specific single answer that respondents originally wanted to provide cannot be explicitly counted, analysed or listed in the results. By including an open answer this foreseeable loss of information could be avoided, and the problems listed in 3.2 could be mitigated. An open answer category would best replace the current remainder category "another language".

The effort and foreseeable expenses connected to the analysis of an open answer, are usually the main reasons for avoiding open answers. The costs for handling an open answer within the language question are not that high, however, in light of the following considerations: According to the German Federal Statistical Office (*Statistisches Bundesamt, destatis*) the micro census is handed to 830,000 people each year (see section 1, above). Dealing with 830,000 open answers would certainly amount to a lot of work, but it can be ruled out that all respondents will provide individual answers in the open answer box. First, the number of open answers would be smaller because the basis for the respective open answers is not the whole population but only the number of households, i.e. only one person per household provides the answer to this special question (which in itself is not desirable, cf. below 3.5). Second, the majority will probably choose one of the answer options provided, so that the number of open answers would probably be relatively small and manageable in size. How many respondents make use of the open answer option certainly also depends on the list of closed answers (which should be altered and expanded as explained in 3.2.).

Furthermore, regardless of its exact size, only the first sample of open answers will require an extensive analysis. All following repetitions of the language question will expand or adapt the respective closed-answer options based on the knowledge gained from the previous open answers. The most frequent answers could enter the language list as closed answer options in the following year, as in the example of the Canadian census, given above. The described procedure would result in a longer and possibly even exhaustive list of closed answer options, thereby keeping the share of open answers small in future versions of the survey.

It should be mentioned that, in principle, long and detailed lists of answer options can be integrated in the micro census. Other considerably more detailed lists in the micro census 2017 show this. For instance, the question on the state of birth for respondents who were not born in Germany offers a relatively long and apparently exhaustive list.

---

<sup>10</sup>"Dialekte (Mundarten), z.B. Plattdeutsch, gelten nicht als Muttersprache" ('dialects (patois), e.g. Low German do not count as native languages'),

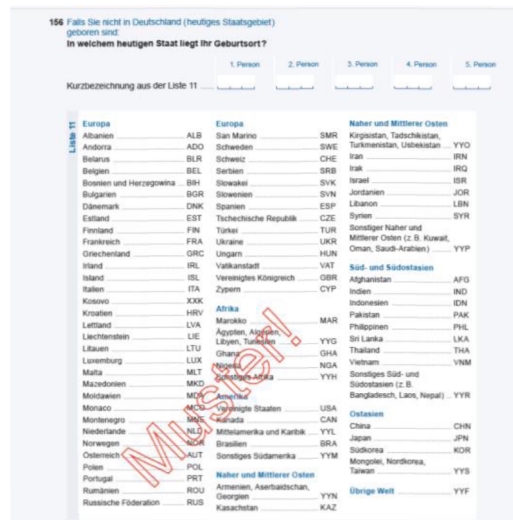


Figure 2 Question on state of birth with respective list of possible answers, German micro census 2017

It is also interesting that the set of answers for state of birth forgoes any general categories. In light of such a detailed list, the truncated list of possible answers indicating the household language appears the more questionable. Its composition points to a rather neglectful handling of the language question or a layman approach to linguistic reality. Another possibility is that the desired answers are covered by the proposed answers so that other options do not appear necessary.

### 3.4 The lack of multiple answers

It is also noteworthy, that the question on language does not allow for multiple answers. The lack of such an option corresponds to the wording of the question, i.e. the singular ('welche Sprache\_' 'which language'). In principle, the question's wording does not exclude the possibility of several spoken languages per household. However, only one predominant language is elicited. Other languages in the household that are potentially spoken with equal frequency cannot be checked. This is also true for any other languages in the household that are not predominant. Accordingly, multilingual households cannot be represented by the results.

The exclusion of multiple answers seems to be a common pattern of the micro census: generally speaking, the sample of the micro census questionnaire avoids multiple answers. Even questions that should clearly offer multiple answers (e.g. the list of reasons why one works part time or why one has a fixed-term contract) are provided with an additional note to "choose the main reason" if several reasons apply ("Wenn mehrere Gründe zutreffen, geben Sie bitte den Hauptgrund an."). It stands to reason that the presentation of multiple answer options leads to a more complex analysis because it deals with cases, i.e. individuals, that fall into several categories of a variable. As a result cross tables and tabular visualisations will be relatively complex. Subsequently, such an analysis also generates higher costs. From a statistical and economic point of view multiple answers are therefore not preferable. However, the design of a questionnaire also requires considerations about how well a question reflects the social reality it intends to elicit. If a question is only a good question with

potentially good results when it allows for multiple answers, then this argument should outweigh any economic considerations that make a simpler statistical analysis more appealing.<sup>11</sup>

It is also important to remember that the language question returned to the German micro census after a long absence. Basically there is no good data on the reality this question should reflect. Under these circumstances it is the more important to opt for a well-designed question that leads to an adequate and detailed picture of the current linguistic situation. This would imply not only having an open answer option but also the possibility of multiple answers. However, this suggestion assumes that the intention of the survey is to accurately portray the linguistic reality in Germany. Unfortunately, this intention has to be called into question (cf. 3.9).

### 3.5 The restriction on respondents

The shortcomings of the language question described so far compound when one realises who gets to answer the question. Although all respondents are confronted with the language question, it is listed in a set of questions related to migration. However, in contrast to the other questions in the set there is only one column available for the answer which reads "1. Person" ('first person') meaning that only one person per household can answer this question (probably the head of the household). When the language question was repeated in the micro census of 2018 (as the micro census law, i.e. *Mikrozensusgesetz* stipulates), an additional note was added explaining that "with regard to question 155, it is sufficient if one person representative of everyone living in the same household answers the question. All other persons of the household continue with question 156." („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 Frage 156 weiter.“, see Figure 3 below).

---

<sup>11</sup> To offer a question including an option of multiple answers is not per se a guaranty for an adequate analysis. An example for an inadequate handling of multiple answers can be found in the Austrian census of 2011 (cf. Busch 2010, 2016). In this census the question on the common language (*Umgangssprache*) allowed for multiple answers. However, in the analysis, the multitude of the responses were reduced: "Mehrfachangaben wurden nur ausgewertet, wenn eine nichtdeutsche Sprache und Deutsch markiert waren." ('Multiple answers have only been analysed when they included a non-German language and German', cf. Statistik Austria no year: 5, cited in Busch 2010, p. 11-12), i.e. when respondents named several non-German languages then only one of these was analyzed: "Bei zwei oder mehreren nichtdeutschen Sprachen wurde nur eine Angabe ausgewertet (hier kam in der Aufarbeitung eine Rangordnungstabelle zum Einsatz)." ('With two or more non-German languages only one answer has been evaluated (for this process a rank-order-table has been applied)', cf. Statistik Austria (no year specified): 5, cited in Busch 2010: 13). This rank-order-table can at least be found in a user manual (cf. Statistik Austria 2007c: 209, cited in Busch 2010:14). That means that the analysis included only those languages that were ranked higher in the table. According to Busch (2010:14) the rank order of this table goes from languages nearby like German to more foreign languages to mostly unknown languages. The rank order and the respectively named languages clearly reveal the underlying language ideology. The list uses labels such as "westafrikanische Eingeborenensprache" ('westafrican indigenous languages') or "Indianersprache" ('Indian language' with 'Indian' probably designating indigenous/Indian people; cf. idem).

155 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 Frage 156 weiter.  
**Welche Sprache wird in Ihrem Haushalt vorwiegend gesprochen?**

Deutsch	01	<input type="checkbox"/>
Nicht deutsch, sondern ...		
... arabisch	02	<input type="checkbox"/>
... englisch	03	<input type="checkbox"/>
... französisch	04	<input type="checkbox"/>
... italienisch	05	<input type="checkbox"/>
... polnisch	06	<input type="checkbox"/>
... russisch	07	<input type="checkbox"/>
... spanisch	08	<input type="checkbox"/>
... türkisch	09	<input type="checkbox"/>
... eine sonstige europäische Sprache	10	<input type="checkbox"/>
... eine sonstige afrikanische Sprache	11	<input type="checkbox"/>
... eine sonstige asiatische Sprache	12	<input type="checkbox"/>
... eine sonstige Sprache	13	<input type="checkbox"/>

Figure 3 Language question in the German micro census 2018<sup>12</sup>

In other words, the information that was implicit in the design of the 2017 question was now made explicit in the version from 2018. This procedure for indicating the household language is unsatisfactory at several levels. When a household consists of only one monolingual individual the term "household language" may be infelicitously chosen due to the lack of interlocutors at home, and yet the information may very well be the closest representation of an actual linguistic situation that the micro census can capture. Multilingual individuals living alone cannot enter all relevant information and are therefore insufficiently addressed. In the case of families or households with multiple members, the answer of a single individual on behalf of the whole household will very likely lead to an oversimplified view of the social realities in many homes. Because of the restriction on multiple answer options, the data will be unable to cover the linguistic situation of families using multiple languages or dialects. Even families with more than one predominant language, e. g. when languages are spoken by different family members at different levels of frequency and proficiency or when languages are used by everyone with perceived equal frequency, cannot enter their data the way the question demands. Information on all these scenarios is lost due to the limitation to a single responding individual and the restriction to a single answer option.

### 3.6 The exclusion of multilingualism

A major issue for the current language question appears to be the underlying understanding of multilingualism. The designers seem to adopt a stringent monolingual perspective dubbed as "monolingual bias" in the linguistic literature.<sup>13</sup> That the monolingual bias would occur in Germany is not surprising, given that the country is conceptually monolingual; indeed, monolingualism is the default assumption in most political and social affairs and in official communication, and it is a perceived reality in the everyday life of many Germans (cf. Adler/Beyer 2018). However, the monolingual bias creates contradictions and has harmful consequences when it is applied to the language question of a census: the language question itself (as apparent from the micro census law) suggests the existence of other languages in the country, and therefore acknowledges the existence of possible multilingual scenarios. The readmission of the language question into the census already

<sup>12</sup> Source: [https://www.statistik.sachsen.de/download/Erhebungsboegen/2A\\_MZ\\_EHB.pdf](https://www.statistik.sachsen.de/download/Erhebungsboegen/2A_MZ_EHB.pdf), last accessed on 18.1.2018.

<sup>13</sup> It is not rare that such a perspective is adopted by census designers; similarities can be found in context of the language question in the census of England in 2011 (cf. Sebba 2017).

shows that there is an increasing awareness of the multilingual situation in Germany. The issues with the question itself are mainly professional and technical: first, it appears that the designers were not familiar with the present constellations of multilingualism in Germany. Specialists in the field and the professional literature were apparently not consulted, leading to the format of the current language list and the imbalances therein. Second, multilingual scenarios are being technically treated as if they were monolingual: the question's wording, the set of answers and the restriction to a single answer make it virtually impossible to express constellations of multilingualism thereby levelling the results to apparent monolingualism. For these reasons, the current implementation of the language question leads to a skewed perception of linguistic and social reality. Because the census results are also intended to serve as a measure of cultural integration (cf. section 3.9 below) a distorted view of linguistic practices can lead to misjudgements and have severe political implications.

In reality, when multiple languages are spoken in a household, the possible constellations are manifold. There might be two or more languages or varieties being spoken one of which can be German. With several persons in a household it is possible that the language proficiencies differ, i.e. the languages at hand are not being spoken equally well or perhaps even not at all by some members of the household. The age and generation of the respective speaker might also play a role. For instance, it is possible that the language between siblings is not the same as the one spoken with or among the parents. In such a complex linguistic environment it is not necessarily clear which language is shared by all household members and which language predominates.<sup>14</sup> However, the decision of which language is eventually named in the answer is only being taken by one sole member of the household. This blurs the accuracy of the decision even more. Besides, the current socio-political situation might also play a role and influence the response by means of social desirability.

### 3.7 Reactions of the media

Considering that the micro census did not feature a language question for almost eighty years, and that the reintroduction of such a question could have raised some suspicion in Germany for historic reasons, there have been relatively few media reactions so far. Possibly this is the case because there are no results available yet. To our knowledge, there also is no documentation describing the development process of the question (which is different from the situation in England and Scotland, for instance; cf. section 2). The intent to integrate a language question into the census received sporadic attention in the media after a press release by AFP (*Agence France Press*). The headlines and links of three reports are listed below:

- (1) „Für Mikrozensus wird künftig Sprache im Haushalt erfragt“ (‘Language of household to be elicited in micro census’, ZD, Zeitschrift für Datenschutz, Oktober 2016, <http://rsw.beck.de/cms/?toc=ZD.ARC.201610&docid=382262>)
- (2) „Regierung fragt Migranten künftig nach der zu Hause gesprochenen Sprache“. (‘Government will ask migrants about language spoken at home’, Zeit Online, 26.8.2016,

---

<sup>14</sup> See Sebba's comment (2017:275f.) on the similar shortcomings of the language question in the English census: "However, this set of categories is a very blunt instrument with which to try to dissect the complex linguistic organism of a multilingual household. The fact is that such households are more than the sum of their mono- or multilingual parts. Different members of the household may draw on similar repertoires but use the languages concerned to a different extent depending on age, generation, gender, occupation or status. The categories used in the census table are not sensitive to these differences."



<http://www.zeit.de/news/2016-08/26/deutschland-regierung-fragt-migranten-kuenftig-nach-der-zu-hause-gesprochenen-sprache-26010404>)

- (3) „Staat will mehr über Migranten wissen. Sprechen Sie zu Hause Deutsch?“ (‘State wants to know more about migrants: Do you speak German at home?’, *Bild*, 25.8.2016, <http://www.bild.de/politik/inland/deutschkurse/sprechen-sie-zuhause-deutsch-47502368.bild.html>)

As the headlines show, the reports emphasize different aspects of the matter. Especially the latter two reports interpret the (re-)introduction of a language question only in one direction, i.e. as a measure to control the linguistic integration of migrants. Indeed, this strikes at the core of the language question's apparent intention (cf. 3.9), though strictly speaking all three headlines are actually wrong. The caption by the *Zeitschrift für Datenschutz* is perhaps the most fitting one, even if the caveats and problems discussed above are kept in mind. That the government is asking migrants what language they speak at home, as the second example drawn from *Zeit online* suggests, is erroneous: first, not only migrants are providing answers to this question. Second, the question only asks for the language that is predominantly spoken in any given household and not which languages are used at home. Third, it is solely the head of the household who gets to answer the question. The first part of the headline from *Bild*, an infamous tabloid newspaper in Germany, may be spelling out the actual intention behind the language question in the micro census. It is an open secret that the German state wants to know more about its migrant population. However, the second part of the headline "Do you speak German at home?" misses the point of the language question. Nobody is asked whether or not they speak German at home. Only the predominant use of seven languages as perceived by a single member of any given household is being explicitly recorded.

The three headlines show how the language question and the intentions behind its reintroduction may be interpreted and enter the current political discourse. Future reactions to the results of the question will most likely be influenced by the fact that the language question touches upon the sensitive topic of integration which is currently being debated even more controversially due to the refugee crisis and its consequences. It is already clear, for instance, that the socio-political climate in Germany in 2017, the year of the micro census, has been rather charged. For the language question, in particular, this means that social desirabilities, i.e. which responses are socially desirable and which are not, might have influenced the responses.

### 3.8 Comparison with United Nations recommendations

It is not easy to produce a lucid formulation and design of the language question that is easy to answer for respondents and conducive to an analysis that eventually leads to results resembling linguistic reality. This generally holds for the design of language questions in all surveys. However, there are different ways to select the best or the most appropriate question from a set of possible questions. For example, one can ask experts (as e.g. in England, cf. Sebba 2017, or in Prussia, where Theodor Siebs was asked to define Frisian, cf. Königlich Statistisches Bureau 1902), conduct a pre-test, or consult other language questions in other countries or in earlier surveys for a comparison. Another possibility is to consult existing recommendations. For example, the United Nations provides a set of recommendations for population censuses of 2010 and 2020. The recommendations of 2006 for the censuses of 2010 were produced in cooperation with Eurostat, the statistical office of the European Union (cf. United Nations Economic Commission for Europe 2006). In 2015 an updated and partly modified set of recommendations was published for the censuses in 2020 (cf. United Nations

Economic Commission for Europe 2015). The recommendations also deal with the language question and suggest a question on the language of the household. According to the recommendations, however, a census should at least ask another, second question on language, e.g. about general language proficiencies (speaking and writing).

Most of the shortcomings of the language question in the German micro census described above find appropriate solutions in the recommendations of the United Nations: multiple responses are preferred, an open answer category should be provided and it should be possible to break down the answers in detail. According to the recommendations the latter is relevant to guarantee a differentiated analysis of individual languages, including languages with small speaker numbers such as dialects and sign languages. For the most part, the question on language of the German micro census does not meet the UN-recommendations for language questions.

### 3.9 Genesis of the language question

It is not clear – at least there is no corresponding documentation available – whether different versions of the language question were trialled in preparation of the micro census. This is the usual procedure by which the quality of the question and the answers of respondents can be analysed and evaluated (cf. the trials of the language question in Scotland and England mentioned by Sebba (2016, 2017)). According to those pretests, the question can be adjusted, changed and optimised. Perhaps such a procedure would have helped improve the current language question and turned it into a more adequate instrument for describing the linguistic situation in Germany's households. However, there are strong indications that the reintroduction of the language question followed an entirely different rationale. An empirical approximation to the actual linguistic situation was apparently never intended.

The only available documents providing the basis for the micro census 2017 point towards the main reason for reintroducing a language question: the increase of migration to Germany in 2015 and 2016. The micro census has a corresponding legal basis which was drafted in summer 2016 and passed as the Micro Census Law (*Mikrozensusgesetz, MZG*)<sup>15</sup> on December 7<sup>th</sup> 2016, taking effect on January 1<sup>st</sup> 2017.<sup>16</sup> The legal text specifies the new regulations of the micro census and the changes of other laws related to population statistics. According to this text, the laws main objective is to obtain a representative statistic of Germany's population including adequate information on participation in the labor market as well as the housing situation. The text also describes why the language question is part of the set of questions on nationality and migration background. Other questions of this set are e.g. the state of birth, the state of birth of the parents as well as their nationality (cf. above 3.1). The position of the language question approximately corresponds to the thematic organisation of the UN-recommendations, which lists language as a subitem of ethnocultural characteristics of the population along with ethnicity and religion (cf. UN 2006, 2015). Thus, the question on language is not part of the demographic information in a narrow sense which would include items such as sex, year of birth and marital status. The thematic environment of the language question already reveals what it is about: although every household is to answer the question, it does not focus on German speaking residents of Germany, i.e. the alleged majority of the

---

<sup>15</sup> The exact title is *Gesetz zur Durchführung einer Repräsentativstatistik über die Bevölkerung und die Arbeitsmarktbeteiligung sowie die Wohnsituation der Haushalte* („Law on the processing of a representative statistic on the population and the participation at the labour market as well as the household's housing“).

<sup>16</sup> Bundesgesetzblatt Jahrgang 2016 Teil I Nr. 59, ausgegeben zu Bonn am 13. Dezember 2016.

resident population in Germany, and their (complex) language repertoires. It rather focuses on people with a so-called migration background and treats language as an ethnocultural indicator. This is most explicitly spelled out in the comments of the corresponding legal text (18/9418, p. 32):

'Migration from foreign countries and the integration of migrants are important political subjects. The micro census data are a crucial part of the report on integration. Information on naturalization facilitates conclusions about migrants' formal willingness to integrate. [...] The recording of the language that is predominantly spoken in the household complements the information on migration background and is significant for the assessment of different dimensions of integration. It allows for differentiated analyses on the status of integration. Cultural integration, in particular, is closely related to the language being spoken in the household. The inclusion of this token will allow for comparisons to other statistics, such as the School Statistic or the Child and Youth Welfare Statistic.'" (emphasis A.A.)<sup>17</sup>

So the actual intention behind the reintroduction of a language question is to survey elements for a report on integration, i.e. the formal willingness to integrate and the so called cultural integration of migrants, which is exclusively measured by means of the language that is predominantly spoken at home. Similar to the language question in the 19<sup>th</sup> century the question serves as a proxy to measure another phenomenon, i.e. integration (cf. section 2). This very likely explains why the language question does not allow for a detailed survey of the linguistic reality in multilingual environments: such detail seems to be beyond the aim of the question. The only piece of information seen as relevant for "cultural integration" is the constructed criterion of the predominant language spoken in the household. However, the claim that "cultural integration" stands in a close relation to this criterion is open to debate. What the previous sections have made sufficiently clear, is that the current survey tool does not measure linguistic realities. In other words, should "cultural integration" be connected to linguistic reality, the language question would fail to measure it.

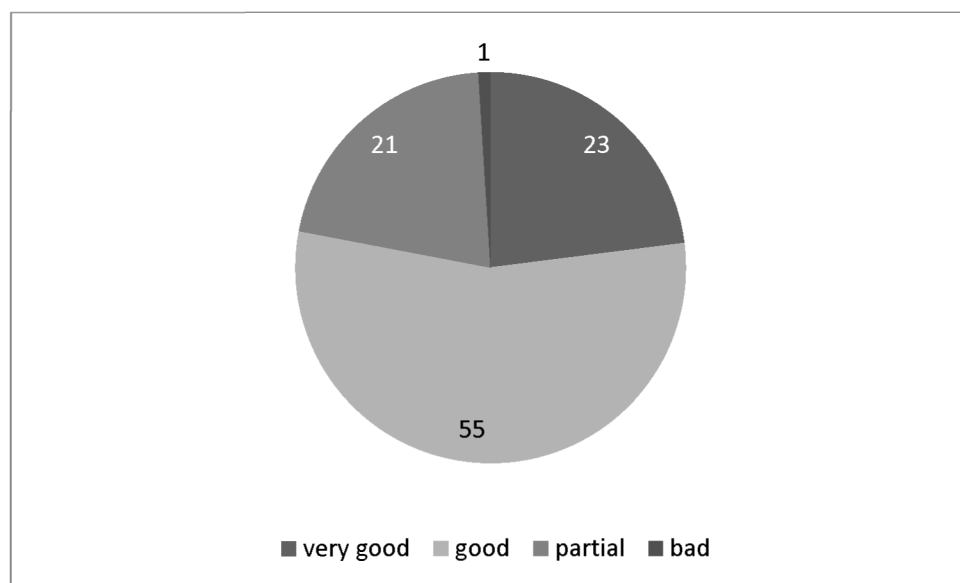
#### 4. Language use and proficiency of German of non-native German-speakers in Germany

Given the legal texts, it is obvious that the intent of the language question is to describe the linguistic behavior of migrants, i.e. multilinguals. At the same time, the question's wording and structure prevent answers that describe multilingualism of any kind. Because only one answer is possible, multilinguals have to decide which of their languages to indicate (if it is available in the list). Otherwise the information is lost. Keeping the questions objective in mind, namely to describe at multilinguals' "cultural integration" it seems highly problematic that they are coerced into to reducing their linguistic repertoire to a single language. It is highly questionable why such a deviation

<sup>17</sup> The original German text is the following: „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 Integrationsberichterstattung. 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.“ (emphasis A.A.)

from linguistic reality should serve as a reasonable indicator of one's "cultural integration" in Germany. Several results of a nationwide survey corroborate this objection.

In 2008, the Institute for the German Language in Mannheim and the Chair of Social Psychology of the University of Mannheim conducted a nationwide telephone poll with over 2000 participants (cf. Gärtig/Plewnia/Rothe 2010). The survey was realised with help of the German polling institute (*Forschungsgruppe Wahlen*), and consisted of questions on language and language proficiencies. Its results are representative for Germany's resident population.<sup>18</sup> 169 respondents (i.e. 8.4 %) of the total of 2004 interviewed indicated another language than German as their native language. The subsample of other native languages than German provided supplementary information on the speakers' proficiency of German and their language use in specific contexts. This information is of special interest for the assessment of the micro census' language question. Figure 4 illustrates the answers relating to the proficiency of German in this subsample:

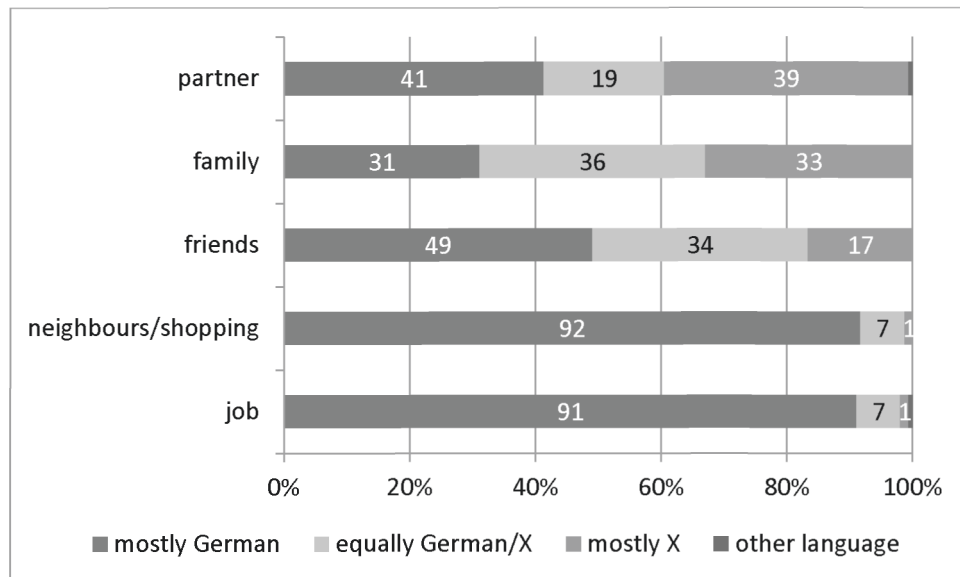


**Figure 4** Proficiency of German in percent, subsample of speakers with another native language than German (N=169, the representative Germany-Survey 2008, cf. Gärtig/Plewnia/Rothe 2010)

Almost eighty percent of the subsample with another native language than German say they have good or very good proficiency in German (78.6 %), 20 % are uncertain about their answer and give themselves a partly good rating (20.8 %). There are almost no respondents in the subsample that rate their command of German as "poor" which may be connected to the manner the survey was conducted (by telephone) so that a rather small amount of respondents with poor proficiency of German was addressed. The assumption would therefore be that the portion of people who do not have German as their native language and have a rather poor proficiency of German is larger than in this subsample. However, apart from this caveat, the data shows clearly that having another native language than German does not necessarily imply a poor command of German. This is acutely relevant for the question in the micro census which aims to elicit the predominant household language. The results should be treated with extreme caution: it would be false to equate the predominant language with the native language, and there also can be no conclusions about the

<sup>18</sup> The representativity for respondents with other language than German is limited: their portion is representative and this subsample's data can be evaluated. However the data of smaller subsets of this subsample may not be meaningful once they are too small and therefore not representative anymore.

proficiency in German. For instance, if the communication in the home is indicated to be mainly in Russian, this has no implications for the proficiency in Russian or German. It merely represents the perceived reality of a single household member that Russian is being used more often than German in the household. A low competence in German can never be inferred from the results of the current language question in the micro census. This is a substantial constraint on the interpretation of results, particularly with regard to the intention of using language as a measure of “cultural integration”. Further findings of the representative survey of 2008 confirm this.



**Figure 5** Language use of respondents with another native language than German (N=169, the representative Germany survey 2008, cf. Gärtig/Plewnia/Rothe 2010)

Figure 5 shows the language use of respondents with another native language than German with interlocutors such as the partner, the family and friends but also in domains such as in shops or at work. Notably, people with another native language than German indicate to almost exclusively speak German in public domains such as the neighbourhood, while shopping and in professional contexts. Only the rather private domains of partnership, family and friendship vary more in language use. At least a third of the subsample usually speaks German in all these private domains, in the context of friendship even half of the respondents communicate mainly in German. A third speaks German and the other language equally often with family and friends. In detail, the results read as follows: 41 % of the subsample usually speak German with their partner, while 19 % speak German and the other language equally often. 39 % mostly use their mother tongue with their partner and only 0.9 % almost always communicate with their partner in their native language other than German. With regard to the family, 31 % usually speak German, 36 % speak German and the other language equally often, and 33 % mostly use the other language. Almost half of the subsample (49 %) usually speaks German with their friends, 34 % speak German and the other language equally often, and 17 % mostly use the other language with their friends. The reason for the variation might be that the partner or family members do not necessarily share the same native language, e.g. if the partner's native language is German or another language, communication will often take place in German. Another explanation for the variation may be the well documented process of language shift in migrant families from one generation to another (cf. Thomason & Kaufmann 1984). In these cases young speakers will tend to use German more often (also among each other) than older speakers.

In summary, respondents with another native language than German are not only proficient in German but also regularly speak it depending on the linguistic domain.<sup>19</sup> This result highlights that the predominant language in the household cannot predict which other languages are used in other contexts. To return to the example of the Russian speaking family: based on these data we can assume that the family members speaking Russian at home, very likely speak mostly German in other contexts of their lives. German and Russian may also be used equally often in many contexts, even in private contexts. Around one third of those with another native language than German might exhibit these bilingual patterns. In the micro census these people are forced to make a statement that misses the point of their linguistic experience, and does not reflect social reality. Altogether, the results of language question in the current micro census will have an oversimplifying effect that is bound to lead to misunderstandings and false interpretations of the results. In light of its intended use as a measure of integration and in the agitated atmosphere of today's debates on integration it can have serious consequences.

## 5. Conclusions: language statistics in Germany

The fact that a language question re-entered the micro census in 2017 gave rise to hopes that there eventually might be reliable data on the number of languages and their speakers in Germany. Unfortunately, the analysis of the question has proven this hope to be false. There are too many shortcomings, beginning with the formulation of the question, the set of possible responses, their type, their design as well as the lack of an open answer and the restriction on multiple responses. Some of these shortcomings can be explained by way of the intention of the language question. Apparently, the question's aim is not to establish a detailed language statistics for Germany but rather to build a survey instrument to measure "cultural integration" as a complement the formal reports on integration. This is what the question about the predominant language spoken in each household, i.e. in the private domain, seems to be designated for.

Even though it is certainly legitimate not to want to establish a detailed language statistic of Germany but to focus on the country's multilingual population, especially the non-German speaking one, the current language question's design does not serve this purpose. The coerced reduction of all languages spoken in a given household to a single language will produce misleading results that are remote from social reality. The available data on language proficiencies and language use of German residents with other native languages highlights this inadequacy, because an individual's proficiency in German is not necessarily linked to the use of the language at home. Moreover it is unclear how exactly "cultural integration" should be deduced from information on the language that is predominantly spoken in a household. The relationship between languages and cultures is by no means a simplistic one, and it is rarely of a causal nature. Rather, the empirical study of linguistic groups in a country should go hand in hand with studies of their social situation and cultural practices. This can lead to a more comprehensive picture about their societal integration, and it could lead to an empirically grounded migration policy. A systematic acquisition of empirical data on the languages in Germany would be an important step towards this goal.

---

<sup>19</sup> The information will be slightly skewed again since the distributions of language use among people with rather poor proficiency of German might differ and therefore the portions may well slightly shift. However, this does not affect the validity of the principal conclusions.

## 6. Literature

- Adler, Astrid (in Vorbereitung): Über die Notwendigkeit und die Tücken von Sprachstatistiken. Die Sprachfrage im deutschen Mikrozensus.
- Adler, Astrid; Beyer, Rahel (2018): Languages and language policies in Germany/Sprachen und Sprachenpolitik in Deutschland. In: Stickel, Gerhard (Hrsg.): National language institutions and national languages. Contributions to the EFNIL Conference 2017 in Mannheim.
- Adler, Astrid; Ehlers, Christiane; Goltz, Reinhard; Kleene, Andrea; Plewnia, Albrecht (2016): Status und Gebrauch des Niederdeutschen 2016. Erste Ergebnisse einer repräsentativen Erhebung. Mannheim: Institut für Deutsche Sprache. Online verfügbar unter <http://www.ins-bremen.de/de/umfrage/ergebnisse/download-broschuere.html>.
- Arel, Dominique (2002): Language questions in censuses: backward or forward-looking? In: David I. Kertzer und Dominique Arel (Hg.): Census and identity. The politics of race ethnicity and language in national censuses. Cambridge: Cambridge Univ. Press (New perspectives on anthropological and social demography, [1]), S. 92–120.
- Bourdieu, Pierre (1991): Language and symbolic power. Hg. v. John B. Thompson. 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 (Working Papers in Urban Language & Literacies, 189). Online verfügbar unter [http://www.academia.edu/20770728/WP189\\_Busch\\_2016.\\_Categorizing\\_languages\\_and\\_speakers\\_Why\\_linguists\\_should\\_mistrust\\_census\\_data\\_and\\_statistics](http://www.academia.edu/20770728/WP189_Busch_2016._Categorizing_languages_and_speakers_Why_linguists_should_mistrust_census_data_and_statistics), zuletzt geprüft am 01.03.2018.
- Gärtig, Anne-Kathrin; Plewnia, Albrecht; Rothe, Astrid (2010): Wie Menschen in Deutschland über Sprache denken. Ergebnisse einer bundesweiten Repräsentativerhebung zu aktuellen Spracheinstellungen. 1. Aufl. Mannheim: Institut für Deutsche Sprache (Arbeitspapiere und Materialien zur deutschen Sprache, 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.
- 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.
- International Journal of the Sociology of Language, Special Issue: Surveying speakers and the politics of census, Issue Editors: Alexandre Duchêne and Philippe N. Humbert, 2018, Number 252 (Jul 2018)
- Kertzer, David I.; Arel, Dominique (2002): Censuses, identity formation, and the struggle for political power. In: David I. Kertzer und Dominique Arel (Hg.): Census and identity. The politics of race ethnicity and language in national censuses. Cambridge: Cambridge Univ. Press (New perspectives on anthropological and social demography, [1]), S. 1–42.
- Königliches statistisches Bureau (1893): Die endgültigen Ergebnisse der Volkszählung im preußischen Staate vom 1. Dezember 1890, nebst einem aktenmäßigen Berichte über die Ausführung dieser Zählung. Berlin: Verlag des Königlichen Statistischen Bureaus (Preußische Statistik, 121).
- Königliches statistisches Bureau (1902): Die endgültigen Ergebnisse der Volkszählung vom 1. Dezember 1900 im preußischen Staate sowie in den Fürstentümern Waldeck und Pyrmont. III. Teil: Die Muttersprache. Berlin: Verlag des Königlichen Statistischen Bureaus (Preußische Statistik, 177).
- Königliches statistisches Bureau (1908): Die endgültigen Ergebnisse der Volkszählung vom 1. Dezember 1905 im preußischen Staate sowie in den Fürstentümern Waldeck und Pyrmont nebst einem aktenmäßigen Berichte über die Ausführung dieser Zählung. Berlin: Verlag des Königlichen Statistischen Bureaus (Preußische Statistik, 206).

- Königliches statistisches Bureau (1913): Die endgültigen Ergebnisse der Volkszählung vom 1. Dezember 1910 im preußischen Staate sowie in den Fürstentümern Waldeck und Pyrmont. Berlin: Verlag des Königlichen Statistischen Bureaus (Preußische Statistik, 234).
- 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), S. 39–61.
- Labbé, Morgane (2009): Internationalisme statistique et recensement de la nationalité au 19ème siècle. In: *Courrier des Statistiques* (127), S. 39–45.
- Leuschner, Torsten (2004): Richard Böckh (1824–1907): Sprachenstatistik zwischen Nationalitätsprinzip und Nationalstaat. In: *Historiographia Linguistica* 31 (2), S. 389–421.
- Lévy, Paul M. G. (1960): La querelle du recensement. Bruxelles: Institut belge de science politique.
- Lieberson, Stanley (1966): Language Questions in Censuses. In: *Sociological Inquiry* 36 (2), S. 262–279. DOI: 10.1111/j.1475-682X.1966.tb00629.x.
- Plewnia, Albrecht; Rothe, Astrid (2012): Sprache – Einstellungen – Regionalität. In: Ludwig M. Eichinger, Albrecht Plewnia, Christiane Schoel, Dagmar Stahlberg und Gerhard Stickel (Hg.): Sprache und Einstellungen. Spracheinstellungen aus sprachwissenschaftlicher und sozialpsychologischer Perspektive. Mit einer Sprachstandserhebung zum Deutschen. Tübingen: Narr (Studien zur deutschen Sprache, 61), S. 4–134.
- Schmidt, Jürgen Erich (2017): Vom traditionellen Dialekt zu den modernen deutschen Regionalsprachen. In: Wolfgang Klein (Hg.): Vielfalt und Einheit der deutschen Sprache. Zweiter Bericht zur Lage der deutschen Sprache. Unter Mitarbeit von Heinrich Detering und Hanns Hatt. Tübingen: Stauffenburg Verlag, S. 105–143.
- 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. DOI: 10.1075/jlp.14026.seb.
- Statistisches Bundesamt: Der Mikrozensus stellt sich vor. Online verfügbar unter <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Mikrozensus.html>, zuletzt geprüft am 12.02.2018.
- Statistisches Bundesamt (2016): Bevölkerung und Erwerbstätigkeit. Ausländische Bevölkerung: Ergebnisse des Ausländerzentralregisters (Fachserie 1 Reihe 2). Online verfügbar unter [https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/MigrationIntegration/AuslaendBevoelkerung2010200167004.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/MigrationIntegration/AuslaendBevoelkerung2010200167004.pdf?__blob=publicationFile), zuletzt geprüft am 01.03.2018.
- Thomason Kaufmann 1984
- United Nations Economic Commission for Europe (2015): Conference of European Statisticians Recommendations for the 2010 Census of Population and Housing. New York, Geneva: United Nations.
- United Nations Economic Commission for Europe (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, Geneva: United Nations.
- Urla, Jacqueline (1993): Cultural politics in an age of statistics: Numbers, nations, and the making of Basque identity. *American Ethnologist* 20(4), 818–843.
- Urla, Jacqueline; Burdick, Christa (2018): Counting matters: quantifying the vitality and value of Basque. *International Journal of the Sociology of Language*, Special Issue: Surveying speakers and the politics of census, Issue Editors: Alexandre Duchêne and Philippe N. Humbert, Number 252, 73–96.
- Vries, John de (2006): Language Censuses/Sprachenzensus. In: Ulrich Ammon (Hg.): Sociolinguistics. An international handbook of the science of language and society. 2nd completely rev. and extended ed. Berlin, New York: Walter de Gruyter (Handbooks of linguistics and communication science, 3.3), 1104–1116.



This text is a working paper. There is also German version available ([https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/7318/file/Adler\\_Die\\_Sprachfrage\\_im\\_deutschen\\_Mikrozensus\\_2018.pdf](https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/7318/file/Adler_Die_Sprachfrage_im_deutschen_Mikrozensus_2018.pdf)). They both represent shorter versions of a longer text that will present language question in census in general, worldwide and especially in Germany – also taking a historic perspective (i.e. in Prussia or during the Third Reich; cf. Adler in Vorbereitung, „Über die Notwendigkeit und die Tücken von Sprachstatistiken. Die Sprachfrage im deutschen Mikrozensus“).

Contact: [adler@ids-mannheim.de](mailto:adler@ids-mannheim.de)