Policy Papers

December 2020

Counting languages: how to do it and what to avoid. A German perspective

by Astrid Adler

- **The annual microcensus provides Germany's most important official statistics. Unlike a census it does not cover the whole population, but a representative 1%-sample of it.**

- **In 2017, the German microcensus asked a question on the language of the population, i.e. 'Which language is mainly spoken in your household?'**

- **Unfortunately, the question, its design and its position within the whole microcensus' questionnaire feature several shortcomings. The main shortcoming is that multilingual repertoires cannot be captured by it.**

- **Recommendations for the improvement of the microcensus' language question: first and foremost the question (i.e. its wording, design, and answer options) should make it possible to count multilingual repertoires.**

1 The language question in the German microcensus is not a good question

In 2017, the German microcensus asked a question on the language of the population. This is noticeable insofar as it was the first time since 1950 that such a question had been included in a census of the German population. That means that for almost 70 years there had been no official data regarding the population's languages Significant changes in the distribution of languages in Germany have therefore not been documented (e.g. by immigration of work migrants from Italy, Poland or Turkey; and also regarding the status of minority and regional languages such as Danish and Low

German). Germany is not the only country that has recently integrated a question on the language of its population. There are other European countries that started asking a question or questions on language in their census, e.g. the UK in 2011, and Luxembourg in 2011.

A question on language in an official census is very welcome. For Germany, the absence of official counting of languages would come to an end. Official data on spoken language is necessary for different stakeholders. They are the basis for making informed decisions, e.g. for the education sector, public health, and other public services. However, the details of the question and its design proved this to be a false hope. In the 2017 questionnaire the question was worded as follows: 'Which language is mainly spoken in your household?' (in German: "Welche Sprache wird in Ihrem Haushalt vorwiegend gesprochen?"). The first possible answer is "Deutsch" ('German'), followed by all other options introduced by the title "Nicht deutsch, sondern …" ('Not German, but …'); the options subsumed under this heading are Arabic, English, French, Italian, Polish, Romanian (added in 2019), Russian, Spanish, Turkish, another European language, another African language, another Asian language, another language. This question was repeated with some small changes in the following years. There are several aspects of the question that have to be criticised:

- Its wording using the singular of 'language' and correspondingly the restriction to one sole answer
- The list of answers, i.e. the set of single languages that are explicitly named and the use of general categories, also the design of these answer options
- The lack of an open answer box
- The restriction to one respondent per household
- Placing the question within the set of questions concerned with citizenship and length of stay

These are such fundamental errors that one is surprised that they occur. On top of that, there is also a specific guide for census questions provided by the United Nations. According to the guidelines of the UN, it is recommended to ask more than one question on language, to allow for multiple answers, and – concerning the granularity of the answers – to allow for the naming of single languages and dialects explicitly and to include an open answer category. None of these options can be found in the language question in the German microcensus.

Some of the shortcomings of the language question become even more serious when one takes into consideration that all respondents of the microcensus are obligated by law to answer. Clearly, for the majority of the German population, this question is not problematic insofar as most are monolingual German speakers. For multilingual residents, however, answering is indeed difficult. Typically, linguistic realities of multilingual households are quite complex; usually there are several languages present and not all of them are being used equally by all members. By restricting the elicitation to only one language this complexity is reduced and the linguistic reality misrepresented. The quality of the produced data is not only subject to the decision of merely one representative of a household but also to social desirability and pure coincidence (e.g. when two languages are spoken equally: how meaningful is the named language that the representative decided to give as an answer? Just try to imagine having to pick one of your several children as the main child – this is how Thomas Bak visualises this challenge, see below for the reference of his text). Thus, the resulting statistics for languages other than German might be too small, whereas the results for German might be too high.

A major challenge for the adequate evaluation of the microcensus in general and the language question in particular is the lack of documentation on both of them. The little information one can find

for the German context are all legal texts constituting the legal basis for the microcensus (cf. Mikrozensusgesetz ('German Microcensus Law')). These texts give an idea about what the purpose of the language question might really be. It is not a census on language but rather the documentation of a language that then shall serve as a proxy to determine cultural integration. Interestingly, many years ago the language question in predecessors of the German microcensus already had a similar purpose (i.e. particularly at the beginning of the twentieth century, language then served as a proxy to supposedly determine people's nationality). In the end, the main issue with the language question in the German microcensus certainly is that the aim it has been generated for cannot be fulfilled because the question design does not allow for the elicitation of precisely the information in focus. That is, eliciting part of the language repertoire of multilinguals to be then used as a proxy for cultural integration. Thus, ecological validity of its produced results must be severely doubted.

In a census, questions on language are always connected to ideologies, language policies, and issues of national identity. This is evident for the German microcensus, but also in censuses of other countries. Also, shortcomings of the census' language questions are not a genuinely German peculiarity. Recently, there have been discussions about the language questions in England and Scotland for example. In England, the question on language in the 2011 census used the concept of main language and only allowed for one possible answer. Clearly, this does not allow for adequate representation of multilingualism and typically, languages other than English would be under-reported. Also the concept of 'main language' is not well-defined but rather ambiguous. In Scotland, for the first time the 2011 census asked about Scots and proficiency in English. Before that the census only contained a question on Scottish Gaelic (starting from 1881). Unfortunately, Scots is not a well-defined variety (i.e. is it a language or a dialect?). Furthermore, the other two language questions are equally problematic, as is the combination of the three (for details see the referenced paper below). So, the elicited numbers of speakers of Scots are certainly unreliable. However, a positive outcome of the Scottish census 2011 is that it had a positive impact on awareness of Scots. One has to keep in mind that this is a census; this means that virtually everyone in Scotland has seen this question.

All this indicates that the quality of the produced data of a census' language question is not a specific German problem. (Although the fundamental difference here is that the language question in Germany is part of the microcensus not the census. Therefore, not the whole population has seen and answered this question, but only 1% of it.) This and the diversity of questions on languages in censuses make cross-national comparisons of census data on populations' languages more or less impossible.

2 How it could be done: recommendations for the census' language question

In its own words, "the Federal Statistical Office is the institution to contact first for official data on the society, the economy, the environment, and the state." It provides objective data that bears the title of "official" statistics as a brand name and a seal of quality. Quality and independence as well as accuracy and reliability are core values of the German Federal Statistical Office (FSO). Furthermore, it acknowledges its position in reinforcing trust in official statistics. It is clear that an obviously biased question producing inevitably biased results does not align with the FSO's own aspirations. Thus, while it is fundamental that the

FSO acknowledges this inadequacy, relevant reforms are not an easy task. Due to the federal structure the German microcensus is a cooperative product between the FSO and statistical offices of the federal states. Nevertheless, this complex situation should not hinder an improvement of the question.

Therefore, the recommendations presented here on how to improve the question on language cover several different possibilities, starting with the best option from a linguistic point of view and ending with options that may not represent the best option but are easier to implement and still better than the question as it is right now. All these aspects are already implemented in language questions in censuses worldwide.

Documentation (general recommendation)

First of all, a general and very obvious desideratum is that the whole process should be documented thoroughly. Such a document should make transparent what the purpose of a question is, whether there are different possible designs and wordings of a question and what the reasons are to choose one of the possibilities over another. Questions used in a census should be reliable and valid instruments. Otherwise questions should be pretested. All these processes should be comprehensibly documented. There is for example extensive and openly available documentation on the development of census questions for the census in England and Wales provided by the Office for National Statistics. As for now, there is nothing of the kind for the German microcensus, which means that all actions are opaque and cannot be traced or understood.

Open question and multiple answers (ideal scenario part 1)

The most radical way to improve the language question is to replace it with an open question using the plural and allowing for the formulation of multiple answers (and discarding the restrictive adverb "vorwiegend" ('mainly') used so far). This seems feasible in general, as there are already questions in the microcensus allowing for open answers and multiple answers. This design would make a real census on languages possible. In the New Zealand census 2013 for example the respondents can give as many answers as they want (there is a short list with given answers followed by some free space to write in other languages).

Ask more than one question (ideal scenario part 2)

Moreover, adding a second question would permit a more thorough analysis of the complex situation: one could preferably ask about mother tongues first and then about languages that are spoken at home or usually used. In this context, it would certainly be worth discussing whether or not the concept of household language is the one to use for a single question. Also, the concept of mother tongue can be subject to discussion. One might prefer the label 'first language' or use specific wordings as the language(s) in which one could have a conversation about a lot of everyday things (cf. the New Zealand census 2013). In general, all

the members of a household should be asked the question instead of only one. In the Luxemburg census 2011 a set of three questions on language (similar to the Swiss census) was asked of each respondent (i.e. not per household).

This first set of recommendations concerning the question and its wording – especially having an open answer, and the option for multiple answers – may represent a major challenge to the FSO, generating more costs, e.g. printing costs for a longer questionnaire and more costs connected to the complex processing of the multidimensional data. Therefore, the next recommendations represent ways of improving the question that might be easier to implement.

Multiple answers and an extensive list of possible answers (alternative, compromise option)

Another possibility to improve the current question then is to opt for a design similar to the current question on citizenship in the German microcensus. This question consists of two parts: first, all members of a household are asked whether they have German citizenship (there are three answer options: yes, only German citizenship; yes, German citizenship and at least one other; no). Then, all those who do not hold only German citizenship are asked to name these foreign citizenships. Two answers are possible. The answer(s) can be chosen from a list consisting of 81 categories.

Accordingly, the question on language would start with a first individual question on the language (this works regardless of the language concept chosen, i.e. mother tongue, first language or language spoken in the household/at home) and offer three types of answers: 1. Only German, 2. German and a language/languages other than German, 3. Only a language/languages other than German. The vast majority would supposedly opt for the first answer. For these respondents the question on languages stops here. The other respondents, who pick the second or third answer, should then give specifics in an additional question on the language(s) in question (using an open answer box or a list of answers to choose from – preferably at least two). This option would allow for unidimensional processing of the data (i.e. avoid having respondents in more than one category as it would be the case when using a multiple answer question) while preserving the possibility to reflect on multilingual realities.

Improving the list of possible answers and its design (minimal suggested option)

Lastly, if it is not possible to include an open answer, then at least the list of given answers should be improved. A more exhaustive list of individual languages would be preferable to reduce the general categories to a minimum or even dismiss them entirely. The clearly two-folded design of the current list of answers should be dropped as it may cause effects of social desirability; also, there does not really seem to be a need for this kind of presentation.

Currently, the FSO has started to acknowledge our concern. We have engaged in discussions to find means how to satisfactorily improve the language question. It has to be carefully weighed which pieces of the ideal scenario can realistically be implemented. We are confident that in the near future the question will be a better one and therefore we will eventually have valuable official statistics on languages in Germany.

This article was amended on 18th December 2020 to reflect that a question on language last appeared in the German census of 1950, rather than the 1939 census, as previously stated. This is corroborated e.g. by the Census Act 1950. Documentation on this census and the question on language is sparse. It seems that results to this question have not been tabulated and displayed.

Resources

The German Federal Statistical Office (Statistisches Bundesamt, Destatis): https://www.destatis.de/EN/About-Us/_node.html

The Federal Office's 2020 Communication Strategy: https://www.destatis.de/EN/About-Us/Goals-Strategy/communication-strategy-download.pdf?__blob=publicationFile

Census 2021: An opportunity to acknowledge multilingualism, http://www.meits.org/news/item/census-2021-an-opportunity-to-acknowledge-multilingualism

Survey Guidelines by GESIS (Leibniz-Institute for the Social Sciences): https://www.gesis.org/en/gesis-survey-guidelines/instruments

United Nations Economic Commission for Europe. 2015. Conference of European Statisticians Recommendations for the 2020 Census of Population and Housing, https://www.unece.org/fileadmin/DAM/stats/publications/2015/ECECES41_EN.pdf

United Nations Economic Commission for Europe. 2006. Conference of European Statisticians Recommendations for the 2010 Census of Population and Housing. Prepared in cooperation with the Statistical Office of the European Communities (EUROSTAT), https://www.unece.org/fileadmin/DAM/stats/publications/CES_2010_Census_Recommendations_English.pdf

Further reading

Adler, Astrid. 2019. 'Language discrimination in Germany: when evaluation influences objective counting', Journal of Language and Discrimination, 3.2: 232 253, https://journals.equinoxpub.com/JLD/article/view/39952/pdf

Adler, Astrid. 2018. 'Working paper: Germany's micro census 2017: The return of the language question', https://ids-pub.bsz-

bw.de/frontdoor/deliver/index/docId/7857/file/Adler_Germanys_micro_census_2017_The_return_of_the_language_question_2018.pdf

Bak, Thomas. 2018. 'Linguistic Identity in the 2021 Census' https://www.parliament.scot/S5_European/Inquiries/CTEEA_CensusBill_BakThomas_CTEEA_S5_18_CB_05.pdf

Busch, Brigitta. 2016. 'Categorizing languages and speakers. Why linguists should mistrust census data and statistics', Working Papers in Urban Language & Literacies, 189, https://www.academia.edu/20770728/WP189_Busch_2016_Categorizing_languages_and_speakers_Why_linguists_should_mistrust_census_data_and_statistics

Humbert, Philippe, Renata Coray and Alexandre Duchêne. 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), http://www.institut-mehrsprachigkeit.ch/en/file/610/download?token=T8FzI_By

Sebba, Mark. 2019. 'Named into being? Language questions and the politics of Scots in the 2011 census in Scotland', Language Policy, 18.3: 339 362, https://eprints.lancs.ac.uk/id/eprint/126982/1/Named_into_being_postprint.pdf

Cite this article

Adler, Astrid. 2020. 'Counting languages: how to do it and what to avoid. A German perspective'. *Languages, Society and Policy.* https://doi.org/10.17863/CAM.62271

About the author

Astrid Adler is a postdoctoral researcher at the Leibniz Institute for German Language in Mannheim. Her research focuses on sociolinguistics, especially on multilingualism, language attitudes and language statistics.