COGNITIVE STUDIES | ÉTUDES COGNITIVES, 26
Warsaw 2026

Article No.: 3490
DOI: 10.11649/cs.3490

Citation: Adler, A. (2026). The construction and contestation of national identity via language as a statistical category in Germany. *Cognitive Studies | Études cognitives*, *2026*(26), Article 3490. https://doi.org/10.11649/cs.3490

Astrid Adler

Leibniz Institute for the German Language, Mannheim, Germany
adler@ids-mannheim.de

The Construction and Contestation of National Identity via Language as a Statistical Category in Germany

Abstract

This text examines German census statistics as a mechanism through which the state defines national belonging. It explores the interplay between the statistical categories used to assign collective identities and the underlying language ideologies that both shape and are influenced by these taxonomies. Historically and currently, official statistics have employed language not only to identify population affiliation but to actively construct it. Consequently, language serves as a 'proxy characteristic' – a statistical category standing in for national identity. The analysis traces how language-related questions have been formulated in German censuses from the foundation of the German Empire in 1871 to the present, with a particular focus on the contestation inherent in state-led categorisation.
Keywords: language ideologies; official statistics; German censuses; national identity; belonging

1 Language, Nation and Categorisation

What makes someone German – besides German citizenship? What constitutes German national identity, and how does it differ from Austrian, Swiss, or French identities? What are its defining characteristics? Who belongs to the German nation, and who does not? These questions were pivotal during the formation of nation-states in Europe; they were crucial in times of mass population movements, such as in the context of war, and they remain equally significant today as identities – be they personal, communal, or attributed to others – are negotiated in response to contemporary migration. Although defining national identity is inherently complex, language appears essential; specifically, a common language serves as a marker of shared identity and culture (cf. e.g. Smith, 1991, p. 23). In a recent survey by the Pew Research Center (Huang et al., 2024), the majority of respondents worldwide indicated that proficiency in their country's primary language is somewhat' or very' important for being considered a 'true' member of the national community. Approval ratings are particularly high in European countries, such as Germany, Poland, and France. These opinions do not necessarily reflect expert consensus but rather illustrate lay perspectives on the nexus between national identity and language. Language is crucial to the construction of individual and group identity, reflecting how national identity is imagined. These ideologies are fundamental to the attribution of identity and membership: on the one hand, how individuals define their own belonging, and on the other, how they perceive, categorise, and ascribe such characteristics to others. Such aspects of identity are inextricably linked to the processes of social categorisation and the construction of group boundaries.

Identity categories are an essential part of censuses. They identify, count and sort the population according to predefined units (categories). These categories are implemented by the state, or state institutions (state categories, cf. Bialas et al., 2024, pp. 698–699). Censuses are official and usually mandatory; they cover the entire population, and the information is collected for state purposes. Generally, the categorisation of objects, such as the people in a census, and the formation of groups, is a social construction. This construction represents a shared point of view and meaning that is attributed to this category (cf. Yanow,

2003, p. 14). The categorisation used in censuses entails implications about the structure of the categorised objects and classification systems in general. Categories are thought of as mutually exclusive and classification systems as exhaustive (cf. Bowker & Star, 2000, pp. 10–11). But "no real-world working classification system […] meets these 'simple' requirements" (Bowker & Star, 2000, p. 11). Eventually, combining sets of binary characteristics may not be adequate as classifications tend to be much fuzzier (alluding to prototype theory developed by Eleanor Rosch; Bowker & Star, 2000, p. 62). Objects that do not fit are problematic for the classification system; marginality, multiplicity, and heterogeneity are seen as exceptions, and sometimes put in a "residual" category (Bowker & Star, 2000, p. 300). Categorical classifications, as used in censuses, can only inevitably render the complexity of social reality in a reduced and simplified form. This also concerns identities and their typically multiple and dynamic nature. This cannot be reflected by sorting people according to a single expression of a category. However, multiple classifications are rarely provided for in census statistics. Thus, complexity and multidimensionality are usually levelled out. Furthermore, assigning an object to a category merely records a certain status, hence the inherent dynamism necessarily disappears ('inertia'; Bowker & Star, 2000, p. 14). These issues with categorisation in general and censuses in particular create significant potential for inadequacies in category attribution, thereby rendering the results inherently contested. Contestation relates to categorisation – not only in the very specific context of contested languages (i.e. regarding the status of a variety, cf. Tamburelli, 2021), but also more generally, because ascribing a category or identity goes hand in hand with the exclusion of alternatives. The core of categorisation therefore lies in determination, which is also the contestation of alternative options. The attribution of a certain category entails the questioning of all other options.

This contribution examines these aspects by focusing on the language category in official German statistics. It describes the interconnection between census-based categorisation, language ideologies, and the German language are connected and how they relate to contestation. Section 2 outlines language ideologies and how they relate to national identity and categorisation of people. Section 3 then examines the nature of categories and processes of categorisation within censuses. Section 4 analyses how how language ideologies and categorisations are intertwined in German censuses to define national identities. Finally, Section 5 provides concluding remarks.

2 Ideologies on Language, Nation and National Identity

The perception and evaluation of people are influenced by what and how they speak, thus by ideas on language and languages. Language ideologies then also relate to how people are categorised, which identities are ascribed to them, which groups they are ascribed to, and crucially, how national identities are attributed. The attribution of national identity involves the imagination, i.e. construction, of a nation as a social community (Anderson, 2006). This imagination of a nation is connected to the imagination of a national language. Both imaginations relate to the idea that a homogeneous language and a cultural and national identity belong together and form a unity: one language, one nation (Blommaert & Verschueren, 1998, pp. 194–195). Through this link, language is interpreted as a marker of national-ethno-cultural affiliation (Mecheril, 2003, p. 23). In many countries, especially in Europe, language is seen as essential to national identity, e.g. also in Germany (see the results of the Pew survey mentioned in section 1).

This very European idea of the unity of language and nation is rooted in the history of nation states. In the Enlightenment sense, language played a unifying, democratising and socially integrative role. The German language, for example, was understood as a bond that united the community and, in contrast to the languages of the elites, Latin and French, German was seen as the language of the people (cf. e.g. Gardt, 1999, p. 95). The democratising and emancipating power of the unifying common language is linked to the increasing literarisation of the population, the growing importance of written and spoken standard German in the 19th and early 20th centuries and to compulsory education and elementary school from 1919. In this one-nation, one-language ideology, the link between language and nation is perceived as natural, timeless and indisputable (Horner & Bradley, 2019, p. 300). The monolingual individual (i.e. monolingual in the national or standard language) is set as the norm and individuals speaking other languages and multilingual individuals can be constructed as deviating from the assumed standard and norm. In this sense, languages

other than the national language can be perceived as a potential threat to the nation's unity and therefore marginalised and stigmatised. The extent of this deviation and accordingly the evaluations of the speakers depends on which language is involved. Additionally, ideologies on languages are linked to those on national identities. As a result, people in Germany who do not speak German, i.e. people who speak other languages and multilinguals, sometimes relate to and get ascribed a non-German identity. Thus, speaking a language other than the national language leads to the contestation of a German national identity and the questioning of membership and belonging to the national–ethnic group. This connection between national identity and language is reinforced by categorisations in censuses.

3 Census, Categories, and Categorisation

At the same time as nation states, modern censuses developed in Europe. They are a state means of controlling the heterogeneous population, managing diversity, and creating unity. From the nation-state perspective, a census identified, demarcated and counted the people, identities and groups that were understood to be part of the nation. Through this, the imagined national community was defined. Thus, categorisation and counting contributed to the process of nation-building. This construction and imagination entails defining and identifying those who do not belong to the national group. Censuses are therefore not only bureaucratic acts that are intended to depict social reality, they – intentionally or incidentally – actually shape and construct it (Kertzer & Arel, 2002, p. 2). Thereby, they perpetuate "existing social hierarchies and power relations" (Bialas et al., 2024, p. 703). They also show the official state view of which classifications are considered relevant, which groups, names and designations are used and in focus.

Categories are historical artefacts (Bowker & Star, 2000, p. 287). The stating of a possible classification is at the same time the contesting of another possible one, which is then marginalised or silenced (Bowker & Star, 2000, p. 5). Multiple categorisations, ambiguous categorisations, transitions – any ambiguities cannot be rendered by the taxonomy and are thus put in residuals or made to disappear. Classification systems contain standardisations; their diffusion leads to normalisation and naturalisation (Starr, 1992, p. 294). Usually, these classification systems, the decisions and norms underpinning them, are not made explicit or are even hidden and become invisible (Bialas et al., 2024, p. 701; Yanow, 2003, p. 14). This reinforces the process of naturalisation and the categories' appearing "to be objective, natural, and self-evident" (Starr, 1992, p. 294). Once naturalised, questioning classification systems and making them visible is all the more difficult. The underlying cognitive organising principles are known by members of the community, however usually tacitly, and passed on to new members through socialisation and acculturation processes (Yanow, 2003, p. 14). Thus, categories are embedded in national cultural repertoires of evaluation – that is, language ideologies (see Section 2 above), i.e. shared representations of groups, their definitions, characteristics and boundaries (Elrick & Schwartzman, 2015, p. 1542). This touches upon a crucial aspect of categorisation: Categories and denominations are accompanied by evaluations, rankings and hierarchies.

The categorising and counting of a population according to certain identity categories creates collective identities. Also, this sorting makes minorities and majorities visible. People are divided into discrete groups based on individual characteristics, these groups usually are named, and on this basis, further characteristics are then attributed to the individuals that are part of this group. By ascribing people to categories, they are assigned membership in a group and also a socially relevant identity (Bialas et al., 2024, p. 703). Sometimes, people share and think along these identity categories and collective identities, also when they are the object of them (i.e. as a categorised person). Before the development of nation states and censuses, ethnic identities were rather fluid and non-exclusive; and after, they were seen as divisible into separate and neat categories (Kertzer & Arel, 2002, p. 6). Censuses thus depict a certain static view of a social reality and thereby contribute to a simplification of social complexity. Regarding language this is particularly difficult, as language repertoires can be very complex. However, this complexity is often discarded with language categories in censuses and their use for the creation of ethnic identity.

4 Constructing German National Identity in Official Statistics

4.1 Determining National Identity via German Language: Historical Development

Following the foundation of the German Empire in 1871, population censuses were conducted at regular intervals throughout the Empire using a standardised questionnaire. They asked about religion and citizenship of the population, and regularly, but not always, place or country of birth, and also language. Language was recorded as the mother tongue, but not in all censuses and not in all parts of the German Empire; in some censuses, an additional question was asked about knowledge of German. The emergence and design of language as a statistical and state category in German censuses is due to German statistician Richard Böckh and his ideas on the connection between language and national identity. He developed the so-called nationality statistics (see e.g. Labbé, 2009; Leuschner, 2004). One of the main interests in census counts at that time – of nation building – lay in determining the national identity of the population. According to Böckh, however, national identity could not be directly surveyed as a subjective criterion (Böckh's view was not a singular one at the time, rather it is well embedded in the contemporary sociopolitical and ideological context, see Gardt, 1999, and Section 2). Instead, language was the true bearer of national identity and the unmistakable bond that linked all members of a nation into a spiritual community (Böckh, 1866, pp. 264, 297–299). Accordingly, the mother tongue as a statistical category could be used for determining national identity, and to accordingly identify people who could be counted as belonging to the German nationality and those who did not. Hence, the statistical category of language was not intended to represent linguistic realities and diversities, but served as a proxy for another category. Thus begins the tradition of language as an official statistical category used to ascribe national identity.

Böckh's view was controversial and contested among contemporary European statisticians. The discussions about language and other census categories reveal their artificial character and their being based on ideologies, negotiation processes, and the different sociopolitical realities. European statisticians and census officials (from Prussia, Austria, England, and France, for example) met and exchanged at the International Statistical Congress, which took place at regular intervals between 1853 and 1876 (Labbé, 1997). Austrian statisticians (Karl Freiherr von Czoernig and Adolf Ficker) did not share Böckh's view of the connection between language and national identity; they advocated an ethnographic approach: national identity could not be investigated by asking questions, rather one had to analyse a bundle of objective observable features (cf. e.g. Labbé, 2009). There was disagreement not only on the conceptualisation of language and national identity, but also on the formulation of the question on language (e.g. whether to inquire about the respondents' mother tongues, family languages, or vernaculars). Furthermore, the perspective promoted by French statistician Alfred Legoyt understood nationality as a mere political, i.e. legal and administrative, concept, and language as a cultural factor should not be queried at all (cf. e.g. Labbé, 2009). The contemporary discussion among statisticians reveals the contestation that is part of the emergence and formation of statistical categories. However, in the end, this contestation and the different perspectives and arguments are no longer visible once the category is implemented and used; they ultimately become invisible and naturalised (cf. Section 3).

In the German Empire, the mother tongue was regularly surveyed in censuses as a proxy for national identity, i.e. in 1890, 1900, 1905 and 1910, 1925, 1933, 1939, 1946 and 1950 (see Hofmeister-Lemke, 1987).[1] The censuses of 1905, 1910 and 1925 included an additional question on the population's knowledge of German. The wording of the questions and answers shows, in addition to the ideological foundation provided by Böckh's nationality statistics, further insight into the prevailing ideas about language and languages at the time and, above all, those of the census officials. The question – which in fact was more of a list than a question – on mother tongue in the censuses was as follows (here using the example of 1905, Prussia): "Mother tongue: whether German, Dutch, Frisian, Danish, Walloon, Polish, Masurian, Kashubian, Wendish, Moravian, Czech, Lithuanian, or what other language?" This list was followed by a longer line, which could probably also be used to give further answers. The proposed answers are varieties of larger contemporary language groups in the German Empire. The naming of these languages shows the census

[1] It is unclear whether mother tongue was part of the list of questions in the 1871 census. In some years, the question on mother tongue was only asked in some regions, but not in the whole Empire (e.g. 1890 and 1910).

officials' understanding of languages and their status in the Empire. The named ones were apparently recognised as having a certain status by being included in this list with their own name. They were eligible as answers and were expected as such. Dialects, on the other hand, are not defined as possible mother tongues by those responsible for the census: "Dialects ('Mundarten'), e.g. Low German, do not count as mother tongues. The Masurian mother tongue is to be described as 'Masurian', not as Polish" (Statistisches Reichsamt, 1928, p. 11). Masurian – whose status as a dialect or language is contested (Wicherkiewicz, 2021, pp. 41–44) – is explicitly described as a language, which as such can be given as an answer. The question about the mother tongue is asked in the singular. This corresponds to the expected default answer, as shown by additional explanations. According to these the rule is that every person has only one mother tongue and multilingual constellations are presented as exceptions. Only one answer is expected, as there can probably only be one national identity, not several. These notices on how to answer were probably meant to lead to an increase in the number of respondents whose answer was counted as "German", whereas for other languages, speaker numbers were reduced, for example, Polish. Nevertheless, apparently sometimes more than one language was given as an answer.

In three censuses there was an additional question on German language skills. It was asked in 1925 throughout the entire Empire; in 1905 and 1910 only in parts of it.[2] The implementation of the additional question is related to the contemporary contestation about the mother tongue question and, above all, to tensions and conflicts of interest between local and national authorities (cf. Labbé, 2007). While from a national perspective, the interest in the mother tongue question was a proxy for national identity, the local authorities (e.g. Prussia) were interested in the everyday use of the German language by the population. This was to verify the success of assimilation and Germanisation policies. Ultimately, only the question of mother tongue remains in the censuses.

Further information on the ideological basis of the mother tongue question and its intentions can be found in the publications of the results. The tables and texts describing the results show where the focus of the analysis lay. In the course of the preparation process, the data did not remain uncontested and were partially changed (this concerns, for example, answers for Frisian in 1900, cf. Königliches Statistisches Bureau, 1902, pp. VI–VII). The documentation of the 1910 census contains a note regarding the instructions to the enumerators for the question on mother tongue: "The enumerator will correct obviously incorrect information without further ado; in the case of doubtful information about the mother tongue, he must make special investigations, if necessary with the assistance of the local police" (Kaiserliches Statistisches Amt, 1915, p. 10*). Indeed, a certain amount of intervention is to be expected: from the survey situation in the households to the evaluation by the statisticians. Tables and texts in the publications usually present only parts of the collected results. The selection of the data provides further indications of contemporary interests, intentions and ideologies. By choosing and selecting certain topics, these are presented, perceived and consolidated as relevant results. Some parts of the population are linked to selected characteristics, thereby constructing groups and ascribing identities; some groups are highlighted, others silenced. The main result of all these censuses is certainly that the vast majority of the population stated German as their mother tongue.

The censuses of 1925, 1933 and 1939 mark a change in the status of the mother tongue question as a proxy for ethnic affiliation. The 1925 census report, interestingly, makes a reference to the fact that language and ethnicity are not always identical. Even so, information on citizenship and mother tongue was used to analyse foreigners, stateless persons and persons without an assigned citizenship according to their language, thus taking language as the decoder of national identity. The vast majority of these people had German as their mother tongue. This argumentation underlined their identification as ethnically German – in particular if they came from regions that were ceded from Germany after World War I and who accordingly "unjustifiably" lost German affiliation (i.e. citizenship). In addition, the argument that language and nation are not always identical may also have been formulated in view of the next censuses and the introduction of questions on ethnic belonging ("Volkszugehörigkeit") and race.

[2] The wording of the question changed: "If not German, whether full command of the German language? (Yes or no!)" (1905), "If not German, whether command of German? (Yes or no!)" (1910), "If not German, does the person in question understand German?" (1925).

There were two census counts during the Third Reich: 1933 and 1939. The census of 1933 included questions on citizenship, mother tongue, religion and place of birth, and no questions on ethnicity or race (Statistisches Reichsamt, 1935). There were several published reports of the results. One of them featured the results on mother tongue. It focused on foreigners and their mother tongue in some regions (the overall results on mother tongues were not displayed). So, again data on mother tongue was used to illustrate foreigners rather than the native population. The census of 1939 inquired about mother tongue and religion, among other things. Two new questions were added for the first time – on ethnic belonging ("Volkszugehörigkeit") and on race ("Abstammung", focussing on Jews; this information was asked using a separate card). These latter categories showed a major break with the well-established practice of using mother tongue as a proxy and only denominator for national identity (Labbé, 1997). Language was thus merely one of several features (ethnic belonging, race) to determine or contest the national identity of the people.

The career of language as a statistical and state category used to determine national identity does not stop here. It is noteworthy that the first census after the Third Reich in 1946 included questions on mother tongue and on national descent ("nationale Abstammung"; cf. Ausschuss der Deutschen Statistiker für die Volks- und Berufszählung 1946, 1951), and the census in 1950 also featured a question on mother tongue.[3] The Allied Control Council that was responsible for the post-war census intended to bring order into a population that was heavily on the move (refugees, displaced persons etc.). To this end, a category on identity was needed and as citizenship was oftentimes unclear, language was chosen. The question on mother tongue in 1950, however, was constructed as a binary category with the distinction of German vs. non-German. This underlines that the detail of non-German languages was, apparently, of no interest, the focus rather lay, again, on determining German national identity and belonging versus its absence. This was the last time a question on language featured in a German census. Possibly, census officials saw no need to determine German language or national identity anymore.

The current German government claims to generally not collect population statistics and socio-economic data on an ethnic basis since the end of World War II.[4] As the analysis of German censuses shows, there is a long tradition of determining the national identity of the population, mostly using language. Thus, language has a strong tradition as a state category. This enduring tradition in more than 150 years of census taking has consolidated and naturalised it. Thus, contestation and questioning are very difficult (this parallels ideologies, which are rarely questioned the more they seem natural and given; cf. Section 2).

4.2 Determining Integration via German Language: Current Trends

The 1950 census was the last German census in which there was a question about the language of the population. After that, language was no longer surveyed in German censuses. This changed in 2017 when, for the first time, the German microcensus included a question directed at all people about their language (previously there had been one on foreigners' knowledge of German, namely in 1976, 1987, 1980, 1982; it resembled the additional question on German in the censuses of 1905, 1910 and 1925). The microcensus is an official census of the population in Germany that has been conducted annually since 1957 (resp. 1991). It is not a complete census. A representative sample of the total population is surveyed (in 2021, for example, there were 880,137 people, which corresponds to one percent of the total population). The results therefore apply to the entire population and can be extrapolated to it. Most of the questions are mandatory. As the microcensus regularly asks a large set of questions, it is a very important official statistic in Germany. It is also used as a source of information for preparing and making political decisions.

[3] The 1946 census was conducted across all occupation zones. Planning for the 1950 census was initiated by the Committee of German Statisticians for the 1946 Population and Occupational Census (*Ausschuss der Deutschen Statistiker für die Volks- und Berufszählung 1946, 1951*). Simultaneously, the division of Germany accelerated, leading to the establishment of distinct administrative bodies: the Federal Statistical Office in the West and the Central Statistical Office in the East. Consequently, data analysis was carried out independently by each institution.). The analysis of the data then took place separately.

[4] This is stated by the Bundesministerium des Innern und für Heimat (Federal Ministry of the Interior and Community) (Bundesministerium des Innern und für Heimat, n.d.).

The question on language introduced in 2017 is: "Which language is mainly spoken in your household?" It was to be answered by one person for the entire household. One answer had to be selected from a list of 9 resp. 10 individual languages (German, Arabic, English, French, Italian, Polish, since 2019 Romanian, Russian, Spanish, Turkish), three generic answer categories (another European language, another African language, another Asian language) and a residual one (another language). The answer "German" is proposed at the beginning and separately from the following answers by the subheading "Not German but". Minority or regional languages are not included. The wording and the lack of possibility to give several answers make it impossible to depict multilingual constellations. In addition, the list of possible answers is very limited. Furthermore, there is no option for an open answer. This means that the answer must be chosen from the provided ones, even if they do not fit. In conjunction with the obligation to answer, this can lead to a bias in the response (for detailed criticism, see e.g. Adler, 2019). The question was changed in 2021. The new design consists of a combination of two questions that allow for more detailed data collection ("Which language or which languages do you speak at home?" and "Which language do you predominantly speak at home?") and a more realistic account of multilingual language repertoires. Also, the list of possible answers increased considerably, consisting now of 26 languages. These are listed in alphabetical order, complemented by three generic answers and the residual one at the end: Albanian, Arabic, Bosnian, Bulgarian, Chinese, Croatian, Danish, Dutch, English, French, German, Greek, Hungarian, Italian, Kurdish, Macedonian, Pashto, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Spanish, Turkish, Vietnamese, another language spoken in Europe, another language spoken in Africa, another language spoken in Asia, another language. Yet, there is still no open answer, and for all respondents who speak several languages at home, but predominantly German, information on the other language(s) is not collected. Therefore, the results on languages other than German are still imprecise (for a more detailed critique, see Adler, 2023).

After nearly seventy years during which the German state had no apparent reason to inquire about the population's language, the situation changed prior to the introduction of a language question in 2017. The sociopolitical situation in Germany before 2017 was characterised by the so-called refugee or migration crisis, i.e. large numbers of people had fled to Europe and Germany. There are very few documents on the motives and development of the language question. The draft Microcensus Act of 2016 indicates that the question's intention is about the integration of migrants and the assessment of their cultural integration.[5] Another indication comes from the description of the microcensus by the German Federal Statistical Office (Statistisches Bundesamt). Therein, language is presented as one of several new characteristics since 2017 referring to the topic of migration.[6] This is reflected by the structure and order of the questions: the question on language features among questions on migration. Language is thus used as an ethno–cultural indicator, i.e. again, as a proxy measure. This continues the tradition of language as an official statistical category that developed in the 19th and 20th century censuses. The official statistical category of language is connected to migration and integration and this puts it next to another statistical category that has been used in official German statistics since 2005 to report on migration, namely the migration background.

The German Federal Statistical Office introduced the migration background in 2005. Before that, it already existed in education statistics and child and youth statistics (see Will, 2020, p. 4). The introduction of the new statistical category was justified by the fact that there was an interest in the integration of naturalised citizens and children of immigrants. The previously used differentiation by citizenship was deemed no longer sufficient, as many immigrants had German citizenship (Will, 2020, p. 4). A migration background is attributed to persons who themselves or at least one parent of whom does not have German citizenship by birth (Statistisches Bundesamt (Destatis), 2023, p. 5). To calculate the migration background, the answers to up to 19 questions are combined.[7]

[5] See Deutscher Bundestag (2016, p. 32). *Gesetzentwurf der Bundesregierung. Entwurf eines Gesetzes zur Neuregelung des Mikrozensus und zur Änderung weiterer Statistikgesetze*. Drucksache 18/9418.

[6] These are: detailed information on the acquisition of German citizenship, questions on the main reason for immigration, information on the country of birth of the respondents and their parents, and the language predominantly spoken in the household (Statistisches Bundesamt (Destatis), 2023, pp. 6–7).

[7] Refugees and displaced persons from the former German eastern territories are excluded from this census by definition because the cut-off date was defined in such a way that immigration before 1950 is not recorded (the federal

The close connection between both language and migration background can also be illustrated by how the latter is used in school statistics. These are carried out separately by the German federal states, which make use of different definitions of migration background (for a detailed analysis cf. e.g. Kemper, 2017). The three defining characteristics are citizenship, country of birth and language in the family or household. A pupil is attributed a migration background if one of these characteristics applies (i.e. non-German citizenship, country of birth not in Germany, another language than German in the family or household). Thus, language alone, i.e. non-German family language, can be used as sole characteristic to infer a migration background.

In recent years, contestation and criticism of the migration background has increased. The group it describes is large and heterogeneous, people included into the category share very different life experiences. The category, however, is not about migration experiences (by definition, people are included regardless of their own or their family's migration: migration before 1950 is excluded, cf. fn. 7, and e.g. someone having been adopted by a parent without German citizenship by birth is included without having any migration history, or, taking another example, someone born abroad from German parents, with all of them having German citizenship, but never having lived in Germany is excluded), but about emphasising non-German origins (see Will, 2020, e.g. p. 6; see also Heinsohn & Mäding, 2024, with an evaluation of the divergence between self-perception and external attribution of the category). In addition, the statistical category of migration background has increasingly been ethnified or racialised, it is reinterpreted and disseminated as an ethnic, racist and also as a social category (see Elrick & Schwartzman, 2015, p. 1540; Will, 2019, p. 9). The attribution of a migration background sometimes turns people into "foreigners" (Will, 2020, p. 6), the category may get conflated with that of "foreigner" (Will, 2019, p. 9); only, the majority of people to whom a migration background is attributed have German citizenship. However, because the migration background is ultimately about ancestry and not migration, the category is defined as ethnic (Elrick & Schwartzman, 2015; Will, 2019, p. 2; the former evaluate parliamentary debates in a qualitative content analysis). The collective category of migration background can function as an exclusion category and is oftentimes also interpreted as an indicator of problems and support needs (Will, 2019, p. 19). This tends to continue and emphasise the logic of foreigners and, above all, their being connected to "problems" instead of an inclusive component of migration normality (Will, 2019, p. 11). The migration background is transformed into a homogeneous social category that is defined by linguistic characteristics (language deficits, do not speak German at all or poorly), social characteristics (class; lack of education, low socio-economic status) and "exclusion from the imagined national community" (Elrick & Schwartzman, 2015, pp. 1549–1551). These ascribed characteristics are not inherent in the statistical category itself; they are derived from "principles of classification available in Germany's national cultural repertoire and the policy repertoire developed during the guest worker era" (Elrick & Schwartzman, 2015, p. 1551). Part of this national cultural repertoire consists of the nationality statistics and the use of the language question in previous statistics and censuses, their ideological foundation and effects on opinion formation. Both statistical categories – migration background as well as the current and historical language category – are part of ideological structures and at the same time produce and reinforce them. One function of both these categories is to constitute and consolidate the national self-group, the "ethnonational core of "Germans without a migration background" (Will, 2019, p. 4). This group is typically not defined; rather it is constructed via defining the other group (cf. Hall, 1989).

5 Constructing and Contesting Germanness and Belonging

In this text, it has been shown how language was and still is operationalised as a statistical category in German censuses and how the construction of national identities along this category is intertwined with ideologies on language and national identity. Historically, language as a statistical category was operationalised as a proxy for national identity – most recently in the 1950 census to determine national affiliation following the turmoil of war. Since 2017, the language category has been reintroduced in official censuses,

states deviate from this in their definition of migration background and use a different cut-off date, see Will, 2019, p. 6).

primarily to evaluate the cultural integration of non-native populations. Thus, alongside the category of migration background it serves as an ethnic category, i.e. as a proxy again. Ideologies on the relation between language and national identity are the basis for the construction of such statistical categories and official taxonomies and at the same time they are enforced by them. Also, the interpretation of statistical categories such as language and migration background as ethnic and/or social categories adds properties to these categories which are not inherently there. However, these categories are integral to how individuals perceive, discuss, and enact them (cf. Yanow, 2003, p. 5); they complement attitudes and beliefs regarding identity and, in this case, language. Ultimately, "social classifications take on a life of their own apart from the claims initially advanced with them" (Starr, 1992, p. 294).

Categorisation and naming in the context of censuses is mainly an external attribution. Although usually people give the required information by self-declaration, the questions and categories are chosen and designed by state agents – thus providing the frame for possible answers – as is the analysis and presentation of results. Self-categorisation, names, and concepts chosen to refer to oneself or oneself's group may differ from externally attributed ones (as is the case for the category of migration background, cf. Heinsohn & Mäding, 2024). Proposed responses in a questionnaire may therefore not be perceived as adequate for the representation of one's own social reality (for example, multilingual language realities). The felt inadequacy of externally attributed categories may be reinforced by the fact that their content and/or meaning is vague or unclear, they are used as proxies, and are imposed rather than deliberately chosen. There may also be differing opinions concerning the status or name of a variety. For example, whether the variety one speaks is regarded and treated as a dialect or a language. Thus, there is a lot of potential for inadequacy and thus contestation. Only, contesting official categorisation is not an easy task (acts of contestation, cf. e.g. Bialas et al., 2024, p. 701). First, taxonomies must be made visible so that their – at times hidden – narratives can be understood and, subsequently, actively contested. Categories and taxonomies are standardised – even the interpretation of statistical categories as ethnic and social categories, as is the case for language for example – to such an extent that they seem natural and therefore not questionable and contestable. There has been some criticism about the migration background category, but more or less none about the language category. Lately, there have been some community-based statistics to avoid categorisation and counting by others (for example in the context of data on discrimination, e.g. Aikins et al., 2021).

Ultimately, it is important to define the aim of language as a statistical category in such statistics. What can be the intention behind such data on language? Generally, data on the languages of the population are desirable. Demolinguistic information makes it possible to identify specific needs, for example to develop funding and promotion opportunities and other language policy measures (for specific language communities, e.g. minority languages, in the educational context, e.g. heritage language teaching, or in public services). In principle, of course, it makes sense to acquire the majority language, the national language. It is also clear that there is a link between language and national or ethnic identity, and therefore integration. However, it is unclear how exactly this connection is formed. It is also unclear what the predominant use of a non-German language at home, as asked for in the microcensus, is supposed to mean. It is completely unclear how exactly constellations of several languages and identities take shape. It is therefore questionable whether language is a suitable proxy for determining national identity and thus for contesting national belonging. Should national membership be denied or contested to someone who predominantly speaks a language other than German at home? Many different constellations are conceivable for such linguistic behaviour, which is only a section of an individual linguistic reality: Such a person could have German as their mother tongue, use it in many everyday contexts, but speak a different language at home with a partner. Or: The person has several mother tongues, including German, speaks both at home and actually the same amount, but has to choose one of them to answer the question in the microcensus. Or: The person has a mother tongue other than German, has spoken German since early childhood and uses both languages equally in everyday life, sometimes one and sometimes the other at home. Or: The person has several mother tongues but not German, speaks them only at home, learns German and speaks it at work. Or: The individual possesses a non-German mother tongue, which is spoken at home; while they may use this or other languages – including German – in public spheres, their overall proficiency in German remains limited. And so on, there are many other conceivable constellations. But what does the one language spoken predominantly at home say about ethnic affiliation – regardless of official affili-

ations stated by citizenship? And as ethnic affiliations are in focus without saying so – could it perhaps be more adequate to ask people directly about their ethnic identity? In the end, an important question is how ethnic identity should be conceived and defined in Germany – and also how to handle hybrid identities. For example, under current practice, possessing German citizenship, having only one German parent, and speaking a language other than German at home is insufficient to establish an uncontested identity. What, then, constitutes German ethnicity, and which criteria must be met for an individual's belonging to remain undisputed? These questions require urgent clarification, particularly in contexts where multiple languages and identities intersect.

References

Adler, A. (2019). Language discrimination in Germany: When evaluation influences objective counting. *Journal of Language and Discrimination*, *39*(3.2), 232–253. https://doi.org/10.1558/jld.39952

Adler, A. (2023). Die Erhebung von Sprachen im deutschen Mikrozensus. Amtliche Statistiken über Sprachen in Deutschland: Folge 1. *Sprachreport*, *2023*(4), 20–29. https://doi.org/10.14618/sr-4-2023_adl

Aikins, M. A., Bremberger, T., Aikins, J. K., Gyamerah, D., & Yıldırım-Caliman, D. (2021). *Afrozensus 2020: Perspektiven, Anti-Schwarze Rassismuserfahrungen und Engagement Schwarzer, afrikanischer und afrodiasporischer Menschen in Deutschland*. https://afrozensus.de/reports/2020/Afrozensus-2020.pdf

Anderson, B. (2006). *Imagined communities: Reflections on the origin and spread of nationalism*. Verso.

Ausschuss der Deutschen Statistiker für die Volks- und Berufszählung 1946. (1951). *Bibliographie der Volks- und Berufszählung vom 29. Oktober 1946*. Duncker & Humblot.

Bialas, U., Lukate, J. M., & Vertovec, S. (2024). Contested categories in the context of international migration: Introduction to the special issue. *Ethnic and Racial Studies*, *48*(4), 695–717. https://doi.org/10.1080/01419870.2024.2404493

Blommaert, J., & Verschueren, J. (1998). *Debating diversity: Analysing the discourse of tolerance*. Routledge. https://doi.org/10.1075/z.184.75blo

Böckh, R. (1866). Die statistische Bedeutung der Volksprache als Kennzeichen der Nationalität. *Zeitschrift für Völkerpsychologie und Sprachwissenschaft*, *4*, 259–402.

Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. MIT Press. https://doi.org/10.7551/mitpress/6352.001.0001

Bundesministerium des Innern und für Heimat. (n.d.). *Nationale Minderheiten in Deutschland*. Retrieved January 24, 2025, from https://www.bmi.bund.de/DE/themen/heimat-integration/gesellschaftlicher-zusammenhalt/minderheiten/minderheiten-in-deutschland/minderheiten-in-deutschland-node.html

Deutscher Bundestag. (2016). *Gesetzentwurf der Bundesregierung: Entwurf eines Gesetzes zur Neuregelung des Mikrozensus und zur Änderung weiterer Statistikgesetze* (Drucksache 18/9418). https://dserver.bundestag.de/btd/18/094/1809418.pdf

Elrick, J., & Schwartzman, L. F. (2015). From statistical category to social category: Organized politics and official categorizations of 'persons with a migration background' in Germany. *Ethnic and Racial Studies*, *38*(9), 1539–1556. https://doi.org/10.1080/01419870.2014.996240

Gardt, A. (1999). Sprachpatriotismus und Sprachnationalismus: Versuch einer historisch-systematischen Bestimmung am Beispiel des Deutschen. In A. Gardt, U. Haß-Zumkehr, & T. Roelcke (Eds.), *Sprachgeschichte als Kulturgeschichte* (pp. 89–113). De Gruyter. https://doi.org/10.1515/9783110807806.89

Hall, S. (1989). Ethnicity: Identity and difference. *Radical America*, *23*(4), 9–13.

Heinsohn, T., & Mäding, A. (2024). Die normative Kritik am Konzept des "Migrationshintergrunds" auf dem Prüfstand: Kommunalstatistische Ableitung vs. Selbstwahrnehmung der Befragten. *Stadtforschung und Statistik*, *2024*(2), 90–96.

Hofmeister-Lemke, K. (1987). Die Volkszählung 1987 in historischer Perspektive. *Berliner Statistik*, *1987*(7), 154–158.

Horner, K., & Bradley, A. F. (2019). Language ideology. In J. Darquennes, J. C. Salmons, & W. Vandenbussche (Eds.), *Handbücher zur Sprach- und Kommunikationswissenschaft* (pp. 296–307). De Gruyter. https://doi.org/10.1515/9783110435351-025

Huang, C., Clancy, L., & Austin, S. (2024, January 18). *Language and traditions are considered central to national identity*. Pew Research Center. https://www.pewresearch.org/global/2024/01/18/language-and-traditions-are-considered-central-to-national-identity/

Kaiserliches Statistisches Amt. (1915). *Statistik des Deutschen Reichs: Vol. 240. Die Volkszählung im Deutschen Reiche am 1. Dezember 1910*. Puttkammer & Mühlbrecht.

Kemper, T. (2017). Die schulstatistische Erfassung des Migrationshintergrundes in Deutschland. *Journal für Bildungsforschung Online / Journal for Educational Research Online*, *9*(1), 144–168. https://doi.org/10.25656/01:12972

Kertzer, D. I., & Arel, D. (2002). Censuses, identity formation, and the struggle for political power. In D. I. Kertzer & D. Arel (Eds.), *Census and identity: The politics of race, ethnicity, and language in national censuses* (pp. 1–42). Cambridge University Press. https://doi.org/10.1017/CBO9780511606045.002

Königliches Statistisches Bureau. (1902). *Preußische Statistik: Band 177. Die endgültigen Ergebnisse der Volkszählung vom 1. Dezember 1900*. Verlag des Königlichen Statistischen Bureaus.

Labbé, M. (1997). Le projet d'une statistique des nationalités discutés dans les sessions du Congrès International de Statistique. In F. Ronsin, H. Le Bras, & E. Zucker, *Démographie et Politique* (pp. 127–142). Editions Universitaires de Dijon.

Labbé, M. (2007). Institutionalizing the statistics of nationality in Prussia in the 19th century (from local bureaucracy to state-level census of population). *Centaurus*, *49*(4), 289–306. https://doi.org/10.1111/j.1600-0498.2007.00078.x

Labbé, M. (2009). Internationalisme statistique et recensement de la nationalité au 19ème siècle. *Courrier des statistiques*, *2009*(127), 39–45.

Leuschner, T. (2004). Richard Böckh (1824–1907): Sprachenstatistik zwischen Nationalitätsprinzip und Nationalstaat. *Historiographia Linguistica*, *31*(2–3), 389–421. https://doi.org/10.1075/hl.31.2.09leu

Mecheril, P. (2003). *Prekäre Verhältnisse: Über natio-ethno-kulturelle (Mehrfach-)Zugehörigkeit*. Waxmann.

Smith, A. D. (1991). *National identity*. Penguin Books.

Statistisches Bundesamt (Destatis). (2023). *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund – Ergebnisse des Mikrozensus 2021–2021 (Endergebnisse)*. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Publikationen/Downloads-Migration/migrationshintergrund-endergebnisse-2010220217004.pdf?__blob=publicationFile&v=4

Statistisches Reichsamt. (1928). *Statistik des Deutschen Reichs: Vol. 401. Volks-, Berufs- und Betriebszählung vom 16. Juni 1925*. Reimar Hobbing.

Statistisches Reichsamt. (1935). Volkszählung: Die Bevölkerung des Deutschen Reichs nach den Ergebnissen der Volkszählung 1933. In *Statistik des Deutschen Reichs: Vol. 451/1. Stand, Entwicklung und Siedlungsweise der Bevölkerung des Deutschen Reichs*.

Starr, P. (1992). Social categories and claims in the liberal state. *Social Research*, *59*(2), 263–295.

Tamburelli, M. (2021). Contested languages and the denial of linguistic rights in the 21st century. In M. Tamburelli & M. Tosco, *Contested languages: The hidden multilingualism of Europe* (pp. 21–39). John Benjamins Publishing Company. https://doi.org/10.1075/wlp.8.02tam?locatt=mode:legacy

Wicherkiewicz, T. (2021). Letters of freedom and captivity: Scriptal planning and language ideologies in Baltic Central-Eastern Europe. *Acta Slavica Iaponica*, *42*, 37–54. https://doi.org/10.14943/ASI.42.37

Will, A.-K. (2019). *Was ist eigentlich ein Migrationshintergrund? Was verbirgt sich dahinter?* Informations- und Dokumentationszentrum für Antirassismusarbeit.

Will, A.-K. (2020). *Migrationshintergrund – wieso, woher, wohin?* Bundeszentrale für politische Bildung. https://www.bpb.de/themen/migration-integration/regionalprofile/deutschland/304523/migrationshintergrund-wieso-woher-wohin/

Yanow, D. (2003). *Constructing "race" and "ethnicity" in America: Category-making in public policy and administration*. M. E. Sharpe. https://doi.org/10.4324/9781315705460

The publication was financed at the author's expense.

The author declares that she has no competing interests.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0), which permits redistribution, commercial and non-commercial, provided that the article is properly cited.

© The Author 2026

Publisher: Institute of Slavic Studies, Polish Academy of Sciences, Warsaw, Poland
Publishing history: Received 2024-08-01; Accepted 2025-07-26; Published 2026-02-01.