

# How Many Welsh, Sámi, Sorbs ... Are There? Numeric Data on Ethnicity and Language Speakers as Controversial Phenomenon

LEOŠ ŠATAVA\*

Kolik je Velšanů, Sámů, Lužických Srbů...? Početní údaje o etnicitě  
a mluvčích jazyka jako diskutabilní fenomén

**Abstract:** In connection with professional texts, statistics, encyclopedic entries or other information about ethnic communities, data on their abundance are also commonly mentioned. However, an apparently straightforward fact as *ethnicity/ethnic identity*, seemingly easy to measure exactly, is highly debatable, as it hides a number of pitfalls and difficult-to-answer questions; it is also easy to manipulate and misuse politically. The same applies to the no less vague and difficult-to-measure phenomena of *knowledge/use of language or mother tongue* and their registration. The main emphasis of the text is on the current specific problems of ethnic and linguistic records, which are documented and analyzed: 1/ at a general level; 2/ on several current examples of European minority ethnic groups (Welsh, Sámi, Sorbs, Kalmyks, Rusyns). The tradition of ethnic/linguistic censuses is strong especially in Central and Eastern Europe. The author documents this fact on the Czech example, which is still based on the Austro-Hungarian model. It follows from the above that the standard publication of results of ethnic and linguistic censuses – despite their widespread and apparent exactness – must be received, due to the controversial dimension present, with considerable caution.

**Keywords:** Ethnic censuses; language censuses; ethnolinguistic minorities in Europe; Czech example; case studies

DOI: 10.14712/23363525.2023.4

## Introduction<sup>1</sup>

In the context of technical texts, statistics, encyclopedia entries or other types of information about ethnic and linguistic communities, data on their numbers are usually also given. Nevertheless, the phenomenon of *ethnicity/ethnic identity*, seemingly easy to quantify with accuracy and hardly any problems, is highly questionable as it involves many pitfalls and questions difficult to answer; it is also easy to manipulate and misuse politically.

Although the given issue is not among the most frequent in the field of social science texts, it does not mean that attention is never paid to it. It applies both to the solely scientific production over the past two decades [Kertzer – Arel 2001; Laversuch 2007; Sebba 2018] and to, i.a., materials of (inter)national organizations and institutions [UNECE 2015].<sup>2</sup> In

\* Prof. PhDr. Leoš Šatava, CSc., Ústav etnologie FF UK, nám. Jana Palacha 2, 116 38 Praha 1; Katedra etnologie a mimoeurópskych štúdií FF UCM, Námestie J. Herdu 2, 917 01 Trnava. E-mail: Leos.Satava@ff.cuni.cz; leos.satava@ucm.sk

<sup>1</sup> The immediate impulse to drafting the paper was the text [Ela 2010].

<sup>2</sup> The mentioned text explicitly states: “Ethno-cultural characteristics generally have a subjective dimension as there is often no common understanding as to what ‘characteristic’ or ‘concept’ is really being measured in any

the given context, one of the significant conclusions is the claim that “ethnicity is multi-dimensional and is considered to be more of a process than a static concept, and so ethnic classification should be treated as dynamic with movable boundaries. This may mean that classifications of ethnic categories will change between censuses which, while mirroring society at any one time, may lead to a degree of non-comparability between one census and another” [UNECE 2015: 149].

Still, a number of questions in this field remain hinted at rather than answered. Study of ‘the classics’ of the theory of ethnicity and nationalism – e.g. texts by E. Gellner [Gellner 1983; Gellner 1997], the concept of imagined communities by B. Anderson [Anderson 1983] or the work of M. Hroch [Hroch 2015] – can be used only indirectly in analyzing the validity of ethnic/linguistic censuses. However, in the study and interpretation of ethno-linguistic identities (whether primordialist or ‘imaginary’), presented in censuses, it is necessary to look for one’s own prisms and procedures, compare, differentiate, avoid political influences and look carefully into the past as well.

Unlike the early centralised Western European countries, in which the form of “political nation” prevailed, the idea of an ethnically defined modern nation on a linguistic and cultural basis prevailed in Central and Eastern Europe<sup>3</sup> during the 19th century. In that age of the onset of nationalism and the ongoing *national revivals*, the idea of an empirically comprehensible and relatively easy to detect phenomenon of *ethnicity* (i.e., of ethnic belonging) on and of entire collectivities was also being formed. Although these objectivist (essentialist) approaches, which predominated until the middle of the 20th century, persist in many ways to this day, it is more than obvious that this category is difficult to define, and, above all, variable. In many cases, it combines biological attributes with often only very vague and almost always problematically compatible ethnic, linguistic, cultural, or even emotional levels.

Within the context of the ethnic community understood as an ‘agelong’ and essential category, identification, based, in the spirit of primordialism, on a seemingly *objective* phenomenon of the (biological-cultural) origin, is one outmost point of efforts at defining the concept of ethnic identity (frequently affected by political concerns and conflicts). The results of (ethnic) censuses, which can be even taken as *technologies of truth production* [Urla 1993: 819], should be its exact expression. Despite the long dominance of this approach, however, its weaknesses have manifested themselves over time in comparison with modernist views – among other things, the unrealistic static nature of the primordialist model ignoring constructivist, situationalist or instrumentalist views on the issue – developed by the social sciences community since the 1960s.<sup>4</sup> The new prisms emphasized, among other things, the possibility of a pluralistic or fluctuating ethnic identity, and admitted multiple identities; at the same time, however, the boundaries of ethnicity (and thus the possibility of a trouble-free registration of members of the ethnic community) grew

---

particular census. Moreover, different countries will adopt different concepts and definitions. The ethno-cultural characteristics may also be politically sensitive and may apply to very small, yet identifiable, populations. Members of certain minority groups may be particularly vulnerable to discrimination on the grounds of ethnicity, religion, or language. However, the free and open declaration of respondents should be of essential importance” [UNECE 2015: 148].

<sup>3</sup> In the text, the countries in which the concept of “nation” on a linguistic and cultural basis gained ground in the 19th century are viewed as “Central- and Eastern Europe”.

<sup>4</sup> The paper [Barth 1969] is usually considered an important cornerstone.

loose and blurred further. Therefore, it is obvious that the nomothetic approach displays its considerable limits in this case.

Related to the changes in perceiving the phenomenon of ethnicity and its forms, a number of changes have been under way since about the turn of the 21st century. In the academic field, a marked shift has occurred away from the rather generalising theses of the “founding fathers” of modernist, or in other words, ethnicist approaches [Barth 1969; Anderson 1983; Gellner 1983; Hobsbawm – Ranger 1983; Smith 1986; Gellner 1997]. Innovative, “postmodern” views and schemes come into existence responding to the changes in forma and ethnic perceptions over the time [Brubaker 2004; Vertovec 2007] or educated comparative eclectic popularisation of the topic of ethnicity and nationalism [Eriksen 1993/2002]. However, in spite of the aforementioned development, a (subliminal) primordialist or perennialist perception of the phenomenon of ethnicity in the spirit of “primordial sentiments” [Geertz 1973] or “banal nationalism” [Billig 1995] still persists in everyday popular practice of European populations

Due to problems associated with the concept and definition of ethnicity defined by *consciousness*, the no less fuzzy and difficult-to-measure phenomenon of **language** has prevailed – but defined differently as: *mother tongue*,<sup>5</sup> *native*, *first*, *home*, *spoken*, *most frequently used* or *communal language*, in the form of *knowledge*, etc. [Moore et al. 2010]. However, even these data are almost always subjective, because within the census the claimed (not defined exactly) *knowledge* of the language, or its *use*, is not verified (with exceptions)<sup>6</sup> by any hypothetical independent and objective body, but its data are provided by respondents. The phenomena of knowledge or use of speech are thus exceedingly difficult to evaluate (and then add up the number of users) – this fact is even more complicated in cases of non-state ethnic groups with often double or multiplied identities or in linguistically defined populations without nation-building tendencies and often (almost) without their own education (e.g., Occitans or Aromanians). A completely special category is also formed by *new speakers*, i.e., persons who have acquired knowledge of a given language in a different form than as part of socialization in a family environment [Smith-Christmas et al. 2018; Šatava 2019]. Not only can the phenomenon of *ethnic identity* be quite debatable and controversial, but records of the number of *language speakers* can be open to discussion as well.

**Linguistic criterion:** ‘Spoken language’ (*langue parlée*), i.e., the language spoken by a person in their immediate community, was called the decisive criterion by the International Statistical Congress in Brussels in 1853. At the Congress in St. Petersburg in 1872, the previous provision was changed in favour of ‘the mother tongue’. However, the Austro-Hungarian agenda continued to adhere to the notion of ‘spoken language’; it, however, brought about the problem of distortion (especially for members of minorities living outside the homeland ethnic area). In interwar Czechoslovakia, since 1921, ethnicity was defined by the mother tongue [Šamanová 2005].

Although the effective identification of ethnicity and language seems an instance of ‘adding apples and oranges’, the two given aspects (both in theory and in practice) still mingle or are often confused even today; there are also discussions about relevance, exactness,

<sup>5</sup> Cf. six different possibilities of conception of the term *mother tongue* [Šatava 2009b: 116].

<sup>6</sup> Exceptions are, in the given context, e.g., language tests conducted in connection with the naturalization of immigrants. A specific example is the state language examination which took place after the restoration of the Baltic countries’ independence in 1991 and is necessary for acquiring their citizenship.

mutual relationship, benefits, or disadvantages of both prisms [Arel 2001]. In the relativistic discourse, they are supplemented by other (cultural, social, emotional, etc.) factors that complete the network of identification ties. However, it turns out that the modernist loosening of the boundaries of ethnicity, especially the constructivist resignation to a clear definition of *nation* by a fixed, static list of attributes<sup>7</sup> (which significantly liberalized the issue), did not lead to unambiguous and generally accepted models and conclusions. The unprecedented boom in history in the field of ethnicity and its relationship to language, which has occurred in recent decades, thus rather obscures this fact.

### General Problems of Ethnic and Linguistic Records

In the Central and Eastern European social science and general awareness, the long-standing tradition of ethnic records within the censuses has given rise to the idea that this is a phenomenon, in its imaginary exactness, comparable e.g., with the data on the total population of a given state or another clearly and unambiguously identifiable group (e.g., the number of foreign nationals in a country); it is, moreover, taken as something obvious, usual, and 'normal'. In the pan-European and global contexts, however, it is rather the opposite. Among the (still) active 'census takers' of ethnicity, we find almost exclusively Central and Eastern European (mostly former socialist) countries; some of them (e.g., Romania or Hungary) also have a long tradition of parallel ethnic and linguistic records.<sup>8</sup> Only recently, or currently, the urgency of this fact is beginning to decline in this area as well.<sup>9</sup> On the contrary, Western European countries (in whose conception of a *nation* ethnic and civic elements/principles, leading to the designation of a nationality, to a greater or lesser degree, tend to merge) practically have not observed the ethnicity of their inhabitants.<sup>10</sup> Thus, providing there are any records, in Western Europe, to some extent associated (at least from the point of view of the Eastern European concept) with ethnicity, these are almost exclusively linguistic censuses. These statistics can be divided into two main types: 1/ in states, the very origin and existence of which are strongly linked to multilingualism (e.g., Switzerland, Belgium, Ireland, Finland<sup>11</sup>); 2/ in areas inhabited by ethnic or linguistic minorities (Wales, Scotland, Basque Country, West Friesland ...).<sup>12</sup>

**Welsh case:** Especially in Western Europe, language is not a necessary *sine qua non* when indicating one's ethnicity. E.g., it is undoubtedly possible to be a 'Welsh person' even without the knowledge of Welsh. If we remain at the linguistic level, it is necessary to look 'under the surface of things'. As of 2011, the official Welsh-speaking number in Wales is 562,000. This number was

<sup>7</sup> Cf. in the Soviet Union and other socialist countries, once a cult and binding Stalin's definition of the *nation*. Debate on it (the *Sovetskaya Etnografiya* journal in the early 1960s) was a bold and almost revolutionary act at the time.

<sup>8</sup> Comparison of results on both levels (ethnic and linguistic), often quite differing, can lead to interesting findings 'between the lines'.

<sup>9</sup> To illustrate this in the Czech Republic is the removal the column *ethnicity* from ID cards or official forms, as well as the 'optional' of this datum in the current censuses.

<sup>10</sup> One of the few exceptions is e.g., Austria recording both ethnic and linguistic self-identifications of its citizens.

<sup>11</sup> In the given context it is relevant that e.g., Swedish speaking citizens of Finland are not, in this country, considered 'Swedish ethnic minority' but 'Swedish speaking Finns'.

<sup>12</sup> Viewed from the East European standpoint it is unusual that it is possible to be a Welsh person, or a Basque based on the regional identity (i.e., without knowing the traditional local language). On the other hand, knowing and using the Gaelic language does not imply the existence of the 'Gaelic nation'.

based on the subjective answers of the census respondents, though. However, the mentioned data does not in any way distinguish between 'native speakers', or active (enthusiastic) users/recipients of the language, and those who have learned Welsh, e.g., within their school attendance (common in recent decades for many children from English-speaking families) and do not practically use Welsh or even take negative attitudes towards it. The number of 'genuine' Welsh users is thus significantly lower than the number indicates. By contrast, many Welsh-speakers living in the UK outside Wales (estimated at up to 150,000 people) drop from the statistics.

Outside Europe, we can find some ethnicity-registering states (this is especially true of countries influenced by the Soviet political model: China, Vietnam ...); overall, however, the phenomenon mentioned here tends to be absent. Exceptions (e.g., New Zealand) are, on the one hand, influenced by the traditional concept of European nationalism of the 19th century as a conceptual import into the non-European cultural environment, on the other hand by relatively young waves of prism, resp. policies of cultural pluralism and (including ethnically-linguistically formulated) human rights, resulting in numerous cases in the so-called *ethni(c)ization*.

Although the data on the number of members of specific ethnic groups, given both in the results of censuses and in encyclopedias or professional works, often seem factual, exact, and unproblematic at first glance, their informative value is often vague and complicated.<sup>13</sup>

In cases such as ethnically conceived Czechs (the existence of a socially relatively accepted definition of language, origin, territory, and consciousness, strengthened until recently by a small degree of exogamy and interstate mobility of the population), these figures can be considered an almost adequate reflection of such construed *reality*. However, we do not find such a clear situation in many other European ethnic groups. The situation is completely different for members of ethnic minorities, in whom the phenomenon of dual (or even multiplied) identity and practical (everyday) bilingualism/multilingualism is almost always present to a certain (often considerable) extent. Other specifics are characterized by ethnically not fully consolidated or otherwise 'non-standard' populations, such as Romanies.

When evaluating numerical data, it is necessary to consider mainly the method of origin of the given information. These can be:

- 1) 'Objective' information (e.g., based on official language censuses or in connection with – extremely vague – a category of declared and in some way verifiable *origin* or cultural *traditions*). This primordialist approach has been criticized by the social sciences since the 1980s and significantly 'deconstructed'; however, in common statistical practice (but also in the minds of the vast majority of the census respondents), it still persists or dominates due to the lack of other clues. It is largely due to the fact that its success, however surprising it may appear, relies on *deep, horizontal comradeship of the nation* [Anderson 1983: 16] – namely, the egalitarian form of 'extended kinship', as well as on Geertz' *primordial sentiments* [Geertz 1973: 255–310].

<sup>13</sup> It applies to many seemingly natural, or real components used in the context of the census as the basis in ethnic discourse, e.g., origin, biological background, sense of solidarity, territory, culture, or common history [Roosens 1989: 160].

Many specific situations can arise in cases where the data of (official) censuses differ (either intentionally for political or other reasons or for other grounds) from data declared by minorities themselves [Pan – Pfeil 2000: xix–xx], or when the reported number of members of a certain ethnic group situationistically decreases/increases depending on political and social variables.<sup>14</sup>

The setting of the added categories can be controversial. In the 19th century, Czech emigrants to the USA were often included in the ‘Austrians’ section. After 1900 (along with the growing awareness of the existence of the Czech nation), on the contrary, in American censuses we find (sub)categories such as ‘Moravians’, etc. [Šatava 1996: 168]. When evaluating such *objective* data, it is necessary to consider numerous political, social, and other contexts in the background.

- 2) At the opposite end of the imaginary scale of forms of ethnic self-identification is modernist (constructivist) approach – i.e., the subjectivist, volitional, or emotional attachment to a certain ethnicity (often based on practical and pragmatic reasons) without the need to prove the fact. This liberal, ‘democratic’ concept prevailed during the 20th century in contrast to the previously dominant objectivist approach; however, despite its undeniable advantages over a binding dogmatic list of characters, it also offers many pitfalls that lead to practical absurdity. Formally, although *nationality* (in its ethnic meaning) can be changed as often as desired, real life psychological practice is different ... “If it is easy to resign from the group, it is not truly an ethnic group” [Hughes – Hughes 1952: 156]. The possibility of completely ‘free’ registration of ethnicity also allows for the possible misuse of this phenomenon as a tool for self-assertion or political pressure from ‘professional ethnic activists’ [Roosens 1989: 14]; on another, less socially important level, it can become e.g., an object of comic constructs.<sup>15</sup>
- 3) As a compromise between the previous two concepts can be seen the perennialist (ethnicist) level, which does not look at ethnicity as an ‘eternal attribute’; it does not, however, consider it a random phenomenon, since it should arise from historical development.

**‘Daily plebiscite?’** The French historian Ernest Renan, or rather his frequently paraphrased statement from 1882: “*A nation’s existence is [...] a daily plebiscite*” [Renan 1996: 53].<sup>16</sup> Renan, however, was far from a genuine subjectivist – on the contrary, he strongly emphasized the need for traditions and especially the will of the community for to make common sacrifices. It is the level of the latter aspect that is significantly declining in the area of Europe in today’s ‘post-duty’ period in the intentions of Gilles Lipovetsky [Petrusek 2006: 8].

Obviously, the oscillation between these poles of approach to ethnicity (roughly in the intentions of objectivism versus constructivism) and many existing intermediate levels of this scale, significantly affect the informative value of the numbers given. This is also

<sup>14</sup> A good example of this phenomenon is the official data on the number of Hungarian ethnic group in southern Slovakia. The identification with Hungarian ethnicity fell sharply after World War II; since the 1960s, the number of people declaring themselves as Hungarians has risen again.

<sup>15</sup> In the Czech Republic, an attempt was made in 2001 to organize a practical joke in entering a certain number of people for the ‘Inuit’ ethnicity.

<sup>16</sup> Using the given witticism, Renan, in fact, defended the concrete political message – the right of the recently lost German-speaking Alsace to be part of France [Azurmendi 2014].

supplemented (locally and temporally) by a variably and ambiguously answered question of the extent to which language knowledge, religion, certain cultural and social attributes, and other aspects are a, in addition to *ethnic consciousness*, a sine qua non of full membership within a certain ethnic group.

**Sámi case:** The Sámi people are a good example in this field. Their number in Norway is usually given at 40,000; but, only about 20,000 of them speak Sámi. However, if we considered only the nationally conscious Sámi with a Sámi people family tradition and active language skills, we would come to an even significantly lower number. Specific (and somewhat controversial) is also the form of conditions for the possibility of voting for an ethnically defined, elected representative body – the Norwegian Sámi Parliament (established in 1989) with its seat in the municipality of Karasjok (Kárásjohka in Sámi). Electoral rolls are based both on self-identification and objective, linguistic criterion. Only adults who declares themselves as Sámi and with, at least, one great-grandparent whose home tongue was Sámi have the right to vote [Bergh et al. 2018: 266]. Between 1989 and 2021, the number of voters was gradually increasing up to almost quadruple the initial number.<sup>17</sup>

**Sorbian case:** Similarly, in (Upper and Lower) Lusatia in eastern Germany, the question of the link between language skills, ethnic origin or current ethnic identity is not clearly answered for the local Slavic population. Officially, the number of Sorbs is currently most often stated to be between 40,000 and 60,000. However, only less than 20,000 people are able to speak Sorbian; as the language of socialization, or at least the partial code of everyday communication, this language (divided, moreover, into Upper and Lower Sorbian varieties) is being used by a maximum of only 10,000 [Šatava 2020: 81].

The very basic problem of the relativity and vagueness of the results of ethnic and linguistic statistics is the terrain that is difficult to understand and interpret clearly. In this key context, several issues can be postulated, in particular:

Do census data result from efforts to objectify the existence of census populations, or is their purpose rather declarative? Why not trust the numbers presented by the censuses? Are or are not such freely presented data an expression of the respondents' identity (and thus of the real situation)? Aren't critical corrective comments on census results more of an expression of *ethnic engineering*? What are the reasons for the statistical 'increase' or 'decrease' of the observed populations? How do the targeted policy intentions and pressures, or even uncovered falsifications of results show? Are certain groups, for some reason, afraid to declare their identity or do they hide it for another reason?

The answer to each of the above questions would require a separate study. It is also practically impossible to propose a kind of hypothetical universal 'solution' to the problems outlined. It is impossible to unequivocally suggest: 'how to do it differently' or 'how to do it in a better way'. However, it is necessary to embark on a tedious journey – to assess each monitored case strictly individually, not to accept the presented data uncritically, but on the contrary with targeted analytical meticulous care and 'reading between the lines'. Only in this way is it possible to reach maximum objectifying findings and conclusions. Of course, even these will never be 'exact', but they can still prove significantly more informative than an uncritical acceptance of the statistical data.

<sup>17</sup> In 1989 5,505 people could be registered, in 1997 about 7,500, in 2009 13,890, and in 2017 16,958; most recently, in 2021, then, 20,541 voters [2021 Norwegian Sámi parliamentary election; Number of persons ... 2021].

The phenomenon of *ethnicity* can also, in the cases of (often only formal) territorial separation or autonomy defined based on ethnicity, coincide with the set of the entire population of the territory or with its statutory ethnic group (especially in cases, when a different phenotype supports the group definition).

**Kalmyk case:** In the former USSR, a strong tradition of ethnic identification was created with a certain territorial-administrative unit, resp. his statutory ethnic group. At present, such a considerable proportion of Kalmyks living in the Republic of Kalmykia in southern Russia automatically census 'Kalmyk' nationality in censuses, without the need-to-know Kalmyk language for this self-identification. Here, as in other non-Slavic populations of Russia, the fact seems to be strengthened by visible physiognomic (phenotypic) differences [Šatava 2009a: 268].

A difficult problem is the interpretation of data on ethnicity in ethnically mixed marriages, respectively partnerships. In many areas, the proportion of endogamous marriages has fallen sharply in recent decades in the context of globalization trends and increasing population mobility; children from mixed marriages/partnerships are often assimilated in favor of the majority ethnic group.<sup>18</sup>

A separate category in this context is the double or multiple identity, common in many minority ethnic populations. Its increasing frequency in recent decades relativizes and changes the until recently 'clear' and sharp boundaries between ethnic communities; yet it also changes e.g., the traditional 'Eastern European' perception of ethnicity as a cultural rather than a political phenomenon, based rather on the concept of 'either-or' than on a pluralistic basis.<sup>19</sup>

Many areas also include examples of indifferent *non-ethnic populations* whose identities were most often formed on a local/regional rather than on an ethnic basis. The specific solutions to such situations varied. In many cases, the population was *ethnicised* from the outside and in connection with the lines of state borders (in the Czech context e.g., the Hlučín region, in the European context e.g., Masurians, inhabitants of Istria, and others). Sometimes (as in the case of the Macedonians) at least part of the community eventually established itself as a standard ethnic group/nation, but these groups did not often transcend the level of insufficiently defined – apart from the quixotic efforts of small groups of enthusiasts – identity (e.g., efforts concerning the construction of the Lachian nation within Czechoslovakia; Polesie on the Belarusian–Ukrainian border, etc.). Obviously, in such cases, data on the numbers of members of individual ethnic groups or language/dialect users need to be treated with extreme caution.

**Rusyn (Ruthenian) case:** In the context of former Czechoslovakia, cf. the official decision on the Ukrainian ethnicity of the Rusyn population in eastern Slovakia in 1952 and the subsequent extremely vague and questionable data of official censuses on the numbers of Ukrainians in the country. The virtuality and fluidity of these data did not change fundamentally even after 1991,

<sup>18</sup> Assimilation imbalance can manifest itself in other ways. While the assimilation trend in favor of the majority is often seen as something natural, opposing changes, on the other hand, arouse embarrassment and uncertainty. The (rather exceptional) inclination of some Germans to Sorbian ethnicity is perceived more as a kind of hobby or eccentric model of behavior, rather than as a true assimilation.

<sup>19</sup> E.g., in (Upper and Lower) Lusatia it is common today to encounter a double Sorbian–German identity; the degree of intensity and interconnectedness of both components is then highly individual.

when it was possible to claim both Ukrainian and Rusyn ethnicities in the census. In addition, many inhabitants of the Rusyn origin and language give Slovak ethnicity in censuses [*Szlovákia etnikai térképe ... 2001*].

A special category consists of data on the *ethnic origin* of certain individuals and entire populations, often even based on the exact core of the results of censuses conducted in the new homelands of immigrants. While for the immigrants themselves (and to some extent for their children born in their new homeland), the mentioned data on *ethnicity* or *origin* can be accepted to a certain extent, for other generations their objectivity and validity are already significantly reduced. Usually, it is also not specified how and what the registered origin is delineated and defined – whether based on consciousness, language, (actively maintained) cultural traditions, country of origin, or other criteria ...? It is common to read e.g., about two million people of Slovak origin in the USA,<sup>20</sup> without specifying how (and to which generation of ancestors) this *origin* is defined. Thus, for example, if at least one grandparent were sufficient to recognize the origin, then in the case of their different ethnicity the grandchildren could (simultaneously) have up to four different ethnic origins ...!

### A Central European Prism: Austria-Hungary – Czechoslovakia – Czech Republic

The Central and Eastern European social science perspective has been rooted in the ‘Herderian’ concept of the nation as a community of cultural properties – language, culture, and manners –, that create the *Volksgeist* (‘spirit of the people’). Gradually the idea of a population census conducted on the ethnic-linguistic basis as a self-evident, exact, and relatively smooth process started to emerge and settle also in the general awareness since the end of the 19th century. Ethnic statistics was gradually developed, especially in the first decades of the 20th century, into a partial scientific sub-field with a relatively rich journal and book production – especially in German [*Winkler 1926*].

The adoption of the mentioned concept of ethnic/linguistic statistics can be well documented e.g., on the example of the Czech lands. There is an obvious connection with the official Austro-Hungarian censuses, on the example of which it is possible to present the further development of this approach. Linguistic-based statistics (first realized in 1880) were highly needed in the age of national self-definition and efforts in multiethnic Austria-Hungary, as practical demands for bilingual official communication, education, etc. were derived from it. Pragmatic ‘heritage’ of this ethnic record was also taken over by multiethnic interwar Czechoslovakia, in which professional production focused on ethnic issues was very numerous (the issue was also devoted to a separate journal *Národnostní obzor* (Ethnic Horizon), published since 1930). At the academic level, obviously, more sophisticated views of the phenomenon of *ethnicity* were admitted and developed, too; these were also used after the Second World War for pragmatic reasons (displacement of

<sup>20</sup> For instance, Ľ. Bartalská gives (as of 1990) 1,882,915 persons of Slovak origin in the USA [*Bartalská 2001: 253*].

Sudeten Germans, etc.) [*Hůrský 1947: 46–47*].<sup>21</sup> However, they did not have any significant effect on the traditional ‘everyday’ primordialist conception of the ethnic phenomenon.<sup>22</sup>

Neither did socialist Czechoslovakia deviate from the mentioned line of ethnic records – after all, its ‘big model’, the Soviet Union, was also strongly focused on (extremely static and often formal) records of ethnicity, which began to be differentiated from the linguistic aspect of the matter.<sup>23</sup> After the political change in 1989 and the disintegration of Czechoslovakia (1991), this tradition continued: all four censuses held between 1991 and 2021 included the questions of ethnicity, or that of mother tongue. However, since 1991 there have been some innovations, especially regarding the option of stating (previously unrecognized) Roma, or Moravian, and Silesian nationalities.<sup>24</sup> Also, the form of the latest Census of Population, Houses and Dwellings from March 2021 the query on *ethnicity* (‘národnost’) still contains (with the additional remarks: “Entering the data is not mandatory – Two ethnicities can also be given”).<sup>25</sup> In the explanations to the person’s census sheet, the ethnic and linguistic level of the observed phenomenon is clearly distinguished.<sup>26</sup> Demographers evaluate the frequencies of agreements of individual answers to the question of nationality and mother tongue as an expression of the degree of ethnic stability [*Sekera 1976*].

Official findings on the ethnicity of the population of the Czech Republic are not conceived as mere statistical material without further use – especially in the context of the official status of ethnic minorities and the financing of their cultural institutions and activities are used as important data.<sup>27</sup> The results of these findings are e.g., decisive element in the installation of bilingual signs in municipalities with more than a ten percent share of members of a national minority (it applies to the Polish minority in the Těšín region).<sup>28</sup>

<sup>21</sup> On the interwar evolution of approaches ranging between the objectivist view of the phenomenon of ethnicity and the volitional (situationalist) one, see: [*Bubeník – Křestian 1995*].

<sup>22</sup> For more information on the methodological basis for adding the *ethnicity* (‘národnost’) in Czechoslovakia and the Czech Republic, see: [*Národnostní struktura obyvatel 2014*].

<sup>23</sup> In 1950 the subject of the census in Czechoslovakia started to be declared as “being part of a nation with the cultural and labour community of which the census is internally connected” [*Šamanová 2005: 5–6*].

<sup>24</sup> Unlike Yugoslavia, where the ethnic category ‘Yugoslav’ existed, it was not possible in Czechoslovakia to officially declare the ‘Czechoslovak’ ethnicity. It was, ironically, only possible in the Czech Republic as late as 2011.

<sup>25</sup> The duty to give one’s ethnicity was first abolished in 2001 (at that time it was 173,000 people). In the 2011 Census, this option was used by 2,64 million people already (i.e. a quarter of the total population of the Czech Republic!). Currently, therefore, it is not possible to compare the data about the ethnicity of the population of the country with census results arrived at in censuses from the 20th century.

<sup>26</sup> In the explanatory notes to the census page of the person in Census of Population, Housing and Dwellings of March 2021, for the Czech term *národnost* (ethnicity) it is specified in more detail that: “Each person will enter the ethnicity data according to their own decision. Entering the data is not mandatory. Ethnicity means a membership of a nation, national or ethnic minority. The mother tongue or the predominantly spoken language need not be decisive for determining one’s ethnicity. Two ethnicities can also be given.” [In the Czech original: “Údaj o národnosti uvede každý podle svého rozhodnutí. Uvedení údaje není povinné. Národnosti se rozumí příslušnost k národu, národnostní nebo etnické menšině. Pro určení národnosti nemusí být rozhodující mateřský ani převážně používaný jazyk. Lze uvést i dvě národnosti.”] The mother tongue is defined as *the first language the respondent learnt to speak*. It is also possible to give two languages here.

<sup>27</sup> This can be seen e.g., in the context of the Roma ethnic group (according to the 2021 census, 4,500 people, another 17,200 were of dual ethnicity (Czech and Roma)). In contrast, unofficial estimates speak of about 250,000 Roma (*sic!*).

<sup>28</sup> Before March 2011, the first campaign in history started persuading the Poles in the Czech Republic to declare their Polish ethnicity.

The results of the census also formed the basis for the change (2013) of the position of the Vietnamese minority in the Czech Republic from the level of a *non-traditional and short-term* settled population to the position of an autochthonous minority. Such dependence of the linguistic protection of minorities on census data is rather an anachronism in the Western European context; even in the Czech Republic, according to some statements by state authorities, it should probably be based more on regional and historical contexts in the future.<sup>29</sup>

## Summary

It follows from the above that the standardly published results of ethnic censuses – despite their widespread incidence and apparent exactness – need to be accepted, due to the controversial present dimension, with considerable reserve and caution. Not only can the resulting figures of these statistics change in a situationalist fashion depending on political, social, economic, etc. circumstances, but the main weakness of such records lies in the very lack of definition and fluidity of the phenomenon of *ethnicity*. Moreover, the contours of this currently overused concept are becoming more and more loose and empty in the context of contemporary social science discourse – paradoxically, despite the overproduction on this topic. At the level of ethnic censuses, no fundamental changes in the ethnic composition of Europe in terms of “transnationalism” [Szaló 2008] or “superdiversity” [Vertovec 2007], occurring since the middle of the 20th century, have been reflected; an exception being the possibility of signing up for a dual ethnic identity.

Also, a seemingly less problematic language category can provide a slightly more accurate finding. Absence of completely accurate and obvious distinctions between native and non-native speakers, active and passive users of the language, or only a vague possibility to measure the level of language proficiency, even the results of language censuses rather provide only auxiliary data in monitoring specific (socio)linguistic situations.

However, it would be an exaggeration not to admit the legitimacy and importance of ethnic, or language statistics in the normal existence of states, regions, and their populations. All results, even those, most distant to imaginary *reality* and most questionable ones, have in the context of a given specific situation their informative values, which can be analyzed and interpreted. At the same time, however, it must be borne in mind that it is practically impossible in this context to achieve a kind of objective, exactly detectable and measurable *truth* by statistical methods.

---

## Bibliography

- Anderson, Benedict [1983]. *Imagined Communities. Reflection on the Origins and Spread of Nationalism*. London: Verso.
- Arel, Dominique [2001]. Language categories in censuses: backward- or forward-looking? In: Kertzer, David – Arel, Dominique (eds.). *Census and Identity: The Politics of Race, Ethnicity, and Language in National Censuses*. New York: Cambridge University Press, pp. 92–120.

---

<sup>29</sup> The statement made by O. Klípa, Head of the Secretariat of the Government Council for National Minorities. [Klípa 2011; cf. Motýl 2011].

- Azurmendi, Joxe [2014]. *Historia, arraza, nazioa. Renan eta nazionalismoaren inguruko topiko batzuk* [History, race, nation. Renan and some clichés about nationalism]. Donostia: Elkar.
- Bartalská, Lúbia (ed.) [2001]. *Sprievodca slovenským zahraničím* [The guide to Slovaks abroad]. Bratislava: Dom zahraničných Slovákov.
- Barth, Fredrik [1969]. Introduction. In: Barth, Fredrik (ed.). *Ethnic Groups and Boundaries. The Social Organization of Culture Difference*. Bergen – Oslo: Universitetsforlaget; London: George Allen & Unwin, pp. 9–38.
- Bergh, Johannes – Dahlberg, Stefan – Mörkenstam, Ulf – Saglie, Jo [2018]. Participation in Indigenous Democracy: Voter Turnout in Sámi Parliamentary Elections in Norway and Sweden. *Scandinavian Political Studies* 41 (4): 263–287.
- Billig, Michael [1995]. *Banal Nationalism*. London: Sage Publikations.
- Brubaker, Rogers [2004]. *Ethnicity without Groups*. Cambridge (MA): Harvard University Press.
- Bubeník, Jaroslav – Křesťan, Jiří [1995]. Národnost a sčítání lidu: k historickým souvislostem polemiky mezi Antonínem Boháčem a Emanuelem Rádlem [Ethnicity and censuses: On historical connections of the polemic between Antonín Boháč and Emanuel Rádl]. *Historická demografie* 19: 119–133.
- Ela, Ludwig [2010]. Kelko je Serbow? Abo – dyrbja so mjeňšiny poprawom ličić? [How many Sorbs are there? Or – do minorities actually have to be counted?]. *Rozhlad* 60 (5): 17–19; (6): 16–18.
- Eriksen, Thomas Hylland [1993/2002]. *Ethnicity and Nationalism. Anthropological Perspectives*. London: Pluto Press.
- Geertz, Clifford [1973]. *The Interpretation of Cultures. Selected Essays*. New York: Basic Books.
- Gellner, Ernest [1983]. *Nations and Nationalism*. Ithaca – Oxford: Cornell University Press – Blackwell Publishers.
- Gellner, Ernest [1997]. *Nationalism*. London: Weidenfeld & Nicolson.
- Hobsbawm, Eric J. – Ranger, Terence (eds.) [1983]. *Invention of Tradition*. Cambridge – New York: Cambridge University Press.
- Hroch, Miroslav [2015]. *European Nations: Explaining Their Formation*. New York: Verso.
- Hughes, Everett C. – Hughes, Helen M. [1952]. *Where Peoples Meet: Racial and Ethnic Frontiers*. Glencoe (Ill.): Free Press.
- Hürský, Josef [1947]. *Zjišťování národnosti. (Úvod do problému odnárodnění)* [Detecting one's ethnicity. (Introduction to problem of denationalization)]. Praha: Československý ústav zahraniční.
- Kertzer, David – Arel, Dominique (eds.) [2001]. *Census and Identity: The Politics of Race, Ethnicity, and Language in National Censuses*. New York: Cambridge University Press.
- Klípa, Ondřej [2011]. K čemu sčítat národnost? [Why to count ethnicity?] *Lidové noviny – Orientace* 34 (29. 3. 2011): 25.
- Laversuch, Iman Makeba [2007]. The Politics of Naming Race and Ethnicity: Language Planning and Policies Regulating the Selection of Racial Ethnonyms Used by the US Census 1990–2010. *Current Issues in Language Planning* 8 (3): 365–382.
- Moore, Robert E. – Pietikäinen, Sari – Blommaert, Jan [2010]. Counting the losses: numbers as the language of language endangerment. *Sociolinguistic studies* 4 (1): 1–26.
- Motýl, Ivan [2011]. Sčítání lidu bude i bitvou o polské cedule [Census will also be a battle for Polish signs]. *Týden* (online), 7. 2. 2011. Available at: <[http://www.tyden.cz/rubriky/domaci/scitani-lidu-bude-i-bitvou-o-cesko-polske-cedule\\_193179.html](http://www.tyden.cz/rubriky/domaci/scitani-lidu-bude-i-bitvou-o-cesko-polske-cedule_193179.html)>.
- Národnostní struktura obyvatel* [2014]. Český statistický úřad. Available at: <<https://www.czso.cz/csu/czso/narodnostni-struktura-obyvatel-2011-aqkd3cosup>>.
- Number of persons entitled to vote in the Sami parliamentary elections in Norway from 2005 to 2021* [2021]. Available at: <<https://www.statista.com/statistics/741443/number-of-persons-entitled-to-vote-in-the-sami-parliamentary-elections-in-norway>>.
- Pan, Christoph – Pfeil, Beate Sybille [2000]. *Die Volksgruppen in Europa. Ein Handbuch* [Ethnic Groups in Europe. An Encyclopedia]. Wien: Braumüller.
- Petrusek, Miloslav [2006]. Národní identita v globalizujícím se světě. Rizika a výzvy naivního multikulturalismu a militantního nacionalismu [Ethnic identity in a globalizing world. The risks and challenges of naive multiculturalism and militant nationalism]. In: *Výchova k občanství v rámci školního vzdělávacího programu se zaměřením na potírání rasové a národnostní nesnášenlivosti*. Sborník příspěvků z 13. ročníku Letní školy pro učitele občanské výchovy a základů společenských věd konané ve dnech 20.–24. srpna

- 2006 na Pedagogické fakultě Univerzity Palackého v Olomouci. Olomouc: Univerzita Palackého 2006. Available at: <[http://petrusek.cz/wp/wp-content/uploads/2017/10/Narodni\\_identita.pdf](http://petrusek.cz/wp/wp-content/uploads/2017/10/Narodni_identita.pdf)>.
- Renan, Ernest [1996]. "What Is a Nation?" In: Eley, Geoff – Suny, Ronald Grigor (eds.). *Becoming National: A Reader*. New York – Oxford: Oxford University Press, pp. 41–55.
- Roosens, Eugene E. [1989]. *Creating Ethnicity. The Process of Ethnogenesis*. Newbury Park – London – New Delhi: Sage Publications.
- Szaló, Csaba [2008]. *Transnacionální migrace. Proměny identit, hranic a vědění o nich* [Transnational Migrations. Changes in Identities, Borders and Understanding them]. Brno: Centrum pro studium demokracie a kultury.
- Šamanová, Gabriela [2005]. Národnost ve sčítání lidu v českých zemích [Ethnicity in demographic censuses in the Czech Lands]. *Naše společnost* 2005 (1). Available at: <[https://cvvm.soc.cas.cz/media/com\\_form2content/documents/c3/a138/f28/100023s\\_Samanova-narodnost.pdf](https://cvvm.soc.cas.cz/media/com_form2content/documents/c3/a138/f28/100023s_Samanova-narodnost.pdf)>.
- Šatava, Leoš [1996]. Vystěhovalectví do USA [Emigration to the USA]. In: *Češi v cizině* 9 (ed. Brouček, Stanislav). Praha: ÚEF AV ČR, pp. 155–171.
- Šatava, Leoš [2009a]. Etnická identita a jazykové postoje mládeže v Kalmycké republice (Ruská federace) [Ethnic identity and language attitudes of young people in the Kalmyk Republic (Russian Federation)]. *Český lid* 96 (3): 263–277.
- Šatava, Leoš [2009b]. *Jazyk a identita etnických menšin. Možnosti zachování a revitalizace* [Language and identity of minority language groups. Possible ways of their maintenance and revitalization]. Praha: Sociologické nakladatelství.
- Šatava, Leoš [2019]. New Speakers in the Context of the Minority Languages in Europe and the Revitalisation Efforts. *Treatises and Documents / Razprave in gradivo* 82: 131–151.
- Šatava, Leoš [2020]. *Lužičtí Srbové na přelomu 20./21. století. Etnicita – jazyk – historie – kultura* [Sorbs at the turn of the 20th/21st centuries. Ethnicity – language – history – culture]. Praha: Epoque.
- Sebba, Mark [2018]. Awkward questions: language issues in the 2011 census in England. *Journal of Multilingual and Multicultural Development* 39 (2): 181–193.
- Sekera, Václav [1976]. Národnost a mateřský jazyk [Ethnicity and the mother tongue]. *Demografie, revue pro výzkum populačního vývoje* 18 (3): 215–225; (4): 299–317.
- Smith, Anthony David [1986]. *The Ethnic Origin of Nations*. Oxford – New York: Blackwell.
- Smith-Christmas, Cassie – Ó Murchadha, Noel P. – Hornsby, Michael – Moriarty, Máiréad (eds.) [2018]. *New Speakers of Minority Languages. Linguistic Ideologies and Practices*. London: Palgrave Macmillan.
- Szlovákia etnikai térképe / Národnostná mapa Slovenska / Ethnic Map of Slovakia 2001 [2003]. Dunajská Streda: Lilium Aurum Könyvkiadó.
- UNECE: *Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing* [2015]. 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.
- Vertovec, Steven [2007]. Super-diversity and its implications. *Ethnic and Racial Studies* 29 (6): 1024–1054.
- Winkler, Wilhelm [1926]. *Die Bedeutung der Statistik für den nationale Minderheiten* [The importance of statistics for ethnic minorities]. Leipzig – Wien: Universität Wien.
- 2021 Norwegian Sámi parliamentary election [2021]. Available at: <[https://en.wikipedia.org/wiki/2021\\_Norwegian\\_S%C3%A1mi\\_parliamentary\\_election](https://en.wikipedia.org/wiki/2021_Norwegian_S%C3%A1mi_parliamentary_election)>.

*Leoš Šatava (\*1954) works at the Czech and Slovak academic institutions. In 1996–2001 he was a research fellow of the Sorbian Institute in Bautzen/Budyšin (Germany). Focusing on problems of ethnolinguistic minorities and their revitalization he carried out a number of field research projects. His book Lužičtí Srbové na přelomu 20./21. století (2020) is the collection of author's most important texts concerning the Sorbian topic. In his latest work Evropská etnika bez státu a regionální jazyky v Evropě (2022), he presents ethnolinguistic populations without the state of their own on the European continent.*