Ethnic, Religious and Language Groups: Towards a Set of Rules for Data Collection and Statistical Analysis

Werner Haug

The establishment of reliable and authoritative figures on ethnic, religious, and language groups plays a prominent role in efforts to fight discrimination and foster individual human rights as well as minority group rights. However, there is no consensus among statisticians, demographers and social scientists at large concerning standards to be used in the identification of minority groups and in the processing and analysing generated data. This article identifies different data sources and measurement problems of ethnic, religious and language groups, building on the rich international experience to date. The conclusion presents a set of rules and guidelines for the improvement of data collection, especially in the context population censuses and sample surveys, as well as for data analysis.

Data sources and their problems

Traditionally, population censuses are by far the most extensive and regular data source relating to the size, growth and structure of religious, language and ethnic groups. At the end of the 19th century, many European countries started to include questions on language, religion, ethnic and/or national groups in their censuses. The most recent United Nations/Eurostat recommendations [1] include essentially these same variables among the census topics. In the 1990 round of censuses, a significant majority of European countries collected data on one or another cultural characteristics, with language data having a clear lead. [2]

The picture from censuses is far more complete than that from vital statistics and especially migration statistics. The number of European countries producing a full range of demographic data – census, vital statistics and migration – by ethnic or national group is very limited (Estonia, Latvia, Romania, Russia, Slovakia and Slovenia). Finland, with its integrated register based system, is the only country with a complete set of demographic data (fertility, mortality, migration) according to language groups. [3]

Although censuses are the most important data source, limitations in the scope and usage of census data are considerable. The main limitations are political in nature and they are due to the close relationship between a country's policy on ethnic, religious and language minorities and the availability of official statistical data, especially census data.

Where minorities have no official recognition, national statistical institutes usually follow exactly the same policy. In France, for instance, the INSEE does not collect any data on language, religious or ethnic groups, on the principle of the secular and unitary nature of the French Republic. [4] A recent sample survey by INED and INSEE on the spatial mobility and social integration of immigrants in France included questions on ethnic origins and religious practices of immigrants for the first time. [5] It provoked a violent debate among demographers and policy makers about the "political correctness" and validity of such questions. [6] Restrictions are also applied in other unitary States, such as Greece and Turkey, where the Statistical Institutes do not publish any statistical information which could run counter to the homogeneity proclaimed by the State, except for the minority groups which were recognised in 1923 in the Peace Treaty of Lausanne.

Similar but more limited restrictions concern the suppression of certain questions of relevance to minorities in censuses. For instance, most Socialist and Communist countries suppressed questions on religion and citizenship during the 1950s and 1960s, following the Soviet model. Under Franco, monolingualism was imposed in Spain, and until 1986 censuses did not reflect the plurilingual character of the country. Belgium suppressed questions on language in the census in 1961/62 after the territorial boundaries of the language communities were defined. All respective decisions were a matter of ideology and a sign of increasing political rigidity.

A more subtle but sometimes insidious way of denying recognition to minorities is through the design of the questionnaire (e.g. only pre-printed answers for the officially recognised minorities) and the classifications used. The history of statistics on minority groups is full of examples of politically motivated re-definitions and re-groupings to eliminate, inflate, recreate or subdivide language, religious or ethnic groups. Small groups – and above all non-territorial groups, e.g. Roma – were often completely suppressed or assigned to anonymous "catch-all" categories. [7]

In the former Soviet Union and Yugoslavia, ethnic and national categories were key components in the distribution of political power. The federal structure of both States was designed on ethnic and territorial lines. In the Soviet Union, each citizen had an ethnic – or rather national – affiliation which was recorded in his or her internal passport

and which was almost impossible to change. In 1989, some 70 national groups with varying degrees of territorial autonomy were officially recognised. The statistical classifications reflected the official minority policy and the statistical system was an important instrument for its implementation, especially under Stalin.

In the former Yugoslavia, Tito reproduced the fundamental distinction of the Communist policy regarding minorities between nations (Serbs, Croats, Slovenes, Montenegrins, Macedonians), with a "right" to statehood in their own Republic, and nationalities with the main "national bodies" living outside Yugoslavia, where they constituted a state of their own (Albanians, Hungarians, Italians, Turks). In addition, there were minority groups without any affinity to other States (e.g. Roma, Vlachs) and finally "Yugoslavs" who were counted as a special ethnic group in the censuses. The distinctions were political in nature, but they had direct impact on the statistical system and the results of the censuses according to language, religion and especially ethnic and national group.

Another problem area relates to the way in which censuses are conducted and the political climate during census operations. In countries with a high degree of illiteracy, censuses are usually conducted by enumerators who fill in the questionnaires based on the answers of the respondents. Communication problems, lack of confidence in enumerators and State bodies, [8] inadequate data protection or even direct attempts at pressure or manipulation during the conduct of field operations influenced census results in many places in the past. Populations regularly underestimated in census enumeration are Roma. The way in which census inquiries are posed and the official policy towards Roma made many of them change their ethnic identity. Similar factors appear to lie behind the underestimation of Hungarians in the 1950 census of Czechoslovakia, [9] for instance, or the size of the Turkish minority in the 1975 Bulgarian census. [10] In both cases significant numbers of people underwent a change of national designation from one census to the other as a result of the altered political conditions.

But oppression and conflict are not the only factors that can influence the quality of the data. The same can be true of policies that are supportive of minorities and allocate financial resources, governmental or parliamentary seats, education and employment opportunities or migration quota according to the size of minority groups or individual minority characteristics. These can create positive incentives to declare as a member of a minority group.

The measurement of linguistic, religious and ethnic affiliation

Although many reservations about the reliability of census data stem from the past, they clearly indicate challenges to collecting and interpreting the data. The difficulties are vividly illustrated by the studies on the global Jewish population, which are probably the only attempt to give a global overview of a religious and ethnic community living in very different cultural and political settings. [11] Cultural characteristics are generally less clear-cut than many demographic or economic characteristics. But the question of measurability certainly necessitates a qualified answer for linguistic, religious and ethnic affiliation.

Language groups

Membership of language groups can be determined most reliably. Language skills are present in and influence the daily actions and social communication of every person. Therefore, non-response plays virtually no role, in contrast to questions about religion or ethnicity.

Language skills and language use still can be defined in very different ways. The most frequent definitions include: mother tongue, i.e. the first language learned in early childhood; main language; language(s) spoken at work or in the family; knowledge of languages. When languages are grouped, attention has to be paid to what is often a blurred distinction between official language and dialects, or rather spoken and written languages. Particularly, in multicultural societies and among members of minority groups, multilingualism (first language, second language, etc.) is widespread and should be covered by empirical research and official statistics.

Religious groups

In general, what is investigated is membership of a particular church, community or confession. Investigations of greater depth also go into questions of individual beliefs, creed or the frequency and intensity of religious practice. In contrast to language groups or ethnic groups, multiple affiliations play only a subordinate role in the light of the exclusive nature and clear demarcation of most systems of belief.

Ethnic and national groups

Investigation of membership is hardest of all and most controversial in the case of ethnic groups (or national groups, a term more frequently used in Central and Eastern Europe). Ethnicity is multidimensional, incorporating linguistic, religious, historical and territorial aspects, and has necessarily a subjective dimension. In addition, individual members of minorities may have lost their cultural, language or religious practices, making them indistinguishable

from the majority on the basis of these criteria, yet who still identify with their ethnic or national group. This complexity and the lack of a convincing, generally recognised definition make it hard to make the concept operational for surveys.

Censuses and other statistical surveys seldom attempt to predefine ethnic or national groups on the questionnaires. Instead, they base themselves (and this is typically the case in Central and Eastern European countries) on the perceptions and image the interviewees have of themselves. This presupposes that a population actually defines itself in terms of ethnic or national group criteria and that ethnic demarcations are regarded as meaningful and informative. The more heterogeneous a population from a historical, cultural and racial viewpoint, and the more society and politics highlight such differences, the more true this will be.

Traditional immigration societies (USA, Canada, and Australia, as well as the United Kingdom) also establish data about ethnic ties. [12] But they are usually interested only in selective aspects of ethnicity. Censuses in Canada, for instance, ask questions about ethnic background at the time of immigration, or about ethnic ancestry (for the first time in 1971). The data provide no information about current ethnic group affiliation but are more historically oriented and link geographic information (e.g. on countries and regions of origin like Germany, Holland but also England, Wales or Scotland) with racial features (black, white, mixed) and concepts relating to ethnic groups in the narrow sense (Inuit, North American Indian). Since the adoption in Canada of equality legislation in 1986, statistics also cover membership of visible minorities. These are non-white by definition, and classification combines racial with geographic criteria. [13]

The approach adopted in the UK (and the Netherlands) has been strongly influenced by the North American tradition of race relations and ethnic relations, with the aim of combating racial discrimination and ethnic inequality. Since the 1980s, political and statistical attention has focused on ethnic minorities, a term reserved for non-white immigrants and their descendants, particularly those who immigrated from the Commonwealth to the United Kingdom after the Second World War. A blend of racial (i.e. physical) attributes white, black as well as geographical origin criteria (Caribbean, India, Pakistan, Bangladesh, Africa, etc.) are used to characterise ethnic minorities. The 1991 Census (with the exception of Northern Ireland) was the first to put questions about memberships of racial/ethnic groups defined in this way. Nowadays, most official statistical surveys in the UK make a distinction based on ethnic groups whose integration is the key focus of interest. [14]

Towards a set of rules for data collection and statistical analysis

After this overview of political, conceptual and measurement problems, we would like to propose by way of conclusions a few general rules for data collection and data –analysis, which can contribute to the establishment of more reliable and authoritative information on ethnic, religious and language groups as well as minority-majority relations.

- Censuses, surveys and, as far as possible, vital and migration statistics also should secure the proper enumeration of language, religious and ethnic groups whenever feasible and relevant. Questionnaires should always provide open answer fields to leave room for self-declaration, including the declaration of not belonging to any ethnic or religious group. The path from self-declaration to statistical classification should be documented and should be accessible for scientific confirmation. Definitions and classifications should be based on recognised scientific concepts and be published along with the statistical findings.

- The active involvement of minority representatives in the preparation, conduct and analysis of statistical surveys, the printing of questionnaires and explanatory notes in minority languages, the recruitment of enumerators from the minority community are some of the obvious means of ensuring trust and participation. Statistical surveys should not be linked directly to minority-policy measures and investigations because these can lead answers or even result in the refusal to answer.

- An important safeguard against abuse and misinterpretation of data on minorities is the independence of national statistical institutes and research institutions. These bodies should be free to chose the appropriate statistical methods and scientific concepts without

political interference. Measures to secure confidentiality and guarantee statistical secrecy are essential for inspiring trust in the statistical institutes. Individual data on group membership should never be used for other than statistical purposes.

- Scope for multiple answers should be provided in surveys (insofar as these are objectively meaningful) and multiple attributions should be evident from the data evaluation, in order to reflect overlaps and multiple identities.
- Data interpretation should be based on combining the replies to various questions, not only on language, religion and ethnic or national affiliation but also on citizenship, place of birth (including parent's place of birth), place of origin in the case of migrants, etc.
- Religious, ethnic and language categories should be interpreted in a wider demographic, educational and socio-economic context in order to prevent over-interpretation and cultural artefacts. Multivariate analysis should isolate ethnic, religious and language variables and check for demographic, educational, socio-economic and other variables (e.g. urban-rural) when comparing different groups and looking for causal relationships.

Werner Haug was the chairman of the international group of experts for the 1994 population census in the Republic of Macedonia and the chairman of the Council of Europe's group of specialists on the demographic situation of national minorities. The views expressed in the present article are the sole responsibility of the author.

Footnotes

[1] United Nations/Eurostat, *Recommendations for the 2000 Round of Censuses*. Geneva: United Nations, 1997.

[2] Y. Courbage, "Survey of the statistical sources on religion, language(s), national and ethnic group in Europe," in Haug, Courbage, Compton (eds.), *The demographic characteristics of national minorities in certain European States*, Vol. 1, Strasbourg: Council of Europe Editions, 1998; for the Balkans see B. Kotzamanis, A. Parant, J.-P. Sardon (eds.), *La démographie des Balkans*. Sources et données, Paris: AIDELF, 1998.

[3] See M. Nieminen, "The demographic characteristics of the Swedish speaking population in Finland" in W. Haug, Courbage, Compton, (eds.), *The demographic characteristics of national minorities in certain European States*, Vol. II. Strasbourg: Council of Europe Editions, 1999.

[4] J.-L. Richard, "A French Point of View: Statistics, Integration and Universalism," in The Siena Group Seminar on Social Statistics, *On the Way to a Multicultural Society?* Neuchâtel: Swiss Federal Statistical Office, 1997; P. Simon, "La représentation statistique de l'immigration: peut on comptabiliser l'ethnicité?" in Rallu, Courbage, Piché (eds.), *Old and new minorities*, Paris: Editions John Libbey, 1997.

[5] M. Tribalat, "L'enquête mobilité géographique et insertion sociale: remise en cause des habitudes statistiques françaises," in *Espace, Populations, Sociétés*, Paris, 1996, pp. 2-3; also, M. Tribalat, *De l'immigration à l'assimilation*. *Enquête sur les populations d'origine étrangère en France*, Paris, 1996; P. Simon, "Nationalité et origine dans la statistique française: les catégories ambigues," in *Population*, Paris: INED, 1998.

[6] P. Bernard, N. Weil, "Une virulente polémique sur les données "ethniques" divise les démographes," in *Le Monde*, 6 November 1998; A. Blum, "Comment décrire les immigrés? A propos de quelques recherches sur l'immigration" in *Population*, Paris: INED, 1998, p. 3.

[7] See for instance A. Liebich, "Minorities in Eastern Europe: Obstacles to a reliable count," in RFE/RL *Research Report, Vol. 1*, 1992, p. 20.; G. A. Bondarskaya, "Nationality in Population Statistics in the USSR," in Goldmann, McKenney (eds.), *Challenges of Measuring an Ethnic World, Politics and Reality*, Washington D. C., 1993.

[8] An extreme case is the call for non-cooperation with the authorities. This was the case for the Catholics in Northern Ireland, who were actively encouraged not to state their religion in 1971 and 1981, and for Albanians in Kosovo and the Republic of Macedonia who were called on by their leaders to boycott the Yugoslav census of 1991.

[9] Y. Courbage, "Demographic characteristics of national minorities in Hungary, Romania and Slovakia," in Haug, Courbage, Compton (eds.), *The demographic characteristics of national minorities in certain European States*, Vol. 1, Strasbourg: Council of Europe Editions, 1998.

[10] D. Philipov, P. Naidenova, A. Totev, "The demographic situation of the main ethnic groups in Bulgaria," in Haug, Courbage, Compton (eds.), *The demographic characteristics of national minorities in certain European States*, Vol. II, Strasbourg: Council of Europe Editions, 1999.

[11] S. Della Pergola, "World Jewish Population 1997," in D. Singer, R. Seldin (eds.), *American Jewish Year Book*, Vol. 99, New York, 1999.

[12] J. Bühlmann, P. Röthlisberger, B. Schmid (eds.), *Monitoring Multicultural Societies*, in The Siena Group Report, Neuchâtel: Swiss Federal Statistical Office, 1998.

[13] G. Goldmann, "La mesure de l'ethnicité au Canada: évolution et débats actuels," in Rallu, Courbage, Piché (eds.), *Old and new minorities*, Paris: Editions John Libbey, 1997; D. A. Norris, "Measuring Ethnicity. The Canadian Experience," in The Siena Group Seminar on Social Statistics, *On the Way to a Multicultural Society?* Neuchâtel: Swiss Federal Statistical Office, 1997.

[14] Swiss Federal Statistical Office, *On the Way to a Multicultural Society? The Siena Group Seminar on Social Statistics*, Neuchâtel: Swiss Federal Statistical Office, 1997; D. Lassalle, "La généralisation progressive du recueil de statistiques ethniques au Royaume Uni," in *Population*, Paris: INED, 1998.