

Tilburg University

Mapping linguistic diversity in multicultural contexts

Extra, G.

Published in:
Handbook of language and ethnic identity

Publication date:
2010

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Extra, G. (2010). Mapping linguistic diversity in multicultural contexts: Demolinguistic perspectives. In J. Fishman, & O. Garcia (Eds.), *Handbook of language and ethnic identity* (pp. 107-122). Oxford University Press.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

MAPPING LINGUISTIC DIVERSITY IN MULTICULTURAL CONTEXTS: DEMOLINGUISTIC PERSPECTIVES

Guus Extra

Methodological issues: concepts, paradigms and databases

The focus of this chapter is on mapping linguistic diversity in multicultural contexts (Barni and Extra 2008). We will explore well-known and established approaches, now commonly referred to by the coinages *demolinguistics* (De Vries 1990) or *language demography* (Clyne 2003: 20–69) and *geolinguistics* (Van der Merwe 1989). The term *demolinguistics* originated among Quebec demographers, probably during the 1970s. Over the last few decades, the field has become an international crossing for demography and linguistics; the same holds for *geolinguistics* as the crossing for geography and linguistics. More recent approaches have been explored in terms of linguistic landscaping. Whereas *geolinguistic* and *demolinguistic* studies tend to focus on the spatial and temporal distribution and vitality of languages in the private domain of the home, *linguistic landscaping* has as its focus the public domain in the most literal sense, i.e., in terms of the visibility and distribution of languages on the streets. In this sense, the outcomes of *linguistic landscaping* research should be read with care: they do not intend to present a faithful mapping of the linguistic make-up of the population in a given place.

We will offer cross-national *demolinguistic* perspectives on languages other than “national” languages. Depending on particular contexts or perspectives, such languages are often referred to as minority languages or dominated languages. Numerical classifications do not necessarily coincide with social classifications. According to the 2001 census outcomes in South Africa, (isi)Zulu is the most widely spoken home language there and English functions commonly as *lingua franca* with all its power and prestige (Van der Merwe & Van der Merwe 2006). Whereas English is a minority language in the homes of South Africans, it is the dominant language in society. In Western Europe, Turkish belongs to the major immigrant languages, and it is spoken in the homes of far more people than any of the national languages of the three Baltic States for instance.

Demolinguistic and *geolinguistic* research can be characterized as empirical approaches with a strong fascination for large data sampling and for the visual representation of the resulting outcomes in tabulated figures and language maps. This is not to say that qualitative small-scale data, common in ethnographic research, should be under-estimated. In particular in the domain of multilingualism in multicultural contexts, there is a need for multidisciplinary and complementarity of data collection methods. Table 1 gives an outline of complementary approaches or paradigms in ethnographic *versus* *demo/geolinguistic* research.

Table 1. Complementary approaches or paradigms in ethnographic *versus* demo/geolinguistic research (Barni & Extra 2008: 5)

Research paradigms	Ethnographic research	Demo/geolinguistic research
Research methods	<ul style="list-style-type: none"> • Inductive / Heuristic • (Participating) observation • “Qualitative” 	<ul style="list-style-type: none"> • Deductive • Distance between researcher and informants • “Quantitative”
Usual data	<ul style="list-style-type: none"> • Observed data in multiple contexts • Open-ended and in-depth interviews 	<ul style="list-style-type: none"> • Reported data in single contexts • Selective set of questions in pre-designed questionnaires
Informants	<ul style="list-style-type: none"> • (Multiple) case studies • Single/few informants 	<ul style="list-style-type: none"> • Large-scale studies • Many informants

Particular validity issues arise in each of these two approaches – in ethnographic research in terms of making generalizations, in demo/geolinguistic research in terms of a possible (mis)match between observed and reported data.

A prominent concept in geolinguistic research is the spatial confinement of language groups to a particular geographical area. One should be aware that some language groups show a stronger degree of spatial or territorial confinement than others. Spatial confinement holds in particular for regional (minority) languages, spatial dispersion for immigrant (minority) languages. Henceforward, these two types of languages will be referred to as RM and IM languages, respectively (Extra & Gorter 2008). Taken from a dynamic perspective, regional languages may become (im)migrant languages within or across the borders of nation-states. Take the case of (isi)Xhosa as spoken in South Africa: it has its regional base (“center of gravity”) in the Eastern Cape but in the post-Apartheid era has started to spread to the Western Cape. As a result, Xhosa is beginning to be firmly established next to Afrikaans and English in Cape Town (Van der Merwe & Van der Merwe 2006). A similar awareness should take hold for the concept of “language groups” itself. Although there are many reasons for its popularity, one should be aware that this concept is problematic in any multicultural context. The language repertoire of people in such a context often consists of more than one language. In the European public and political discourse, this has led to the popularity of the reference to plurilingual people in multilingual societies. Taken from this perspective, plurilingualism refers to the ability of individuals to communicate in more than one language, whereas multilingualism refers to a key marker of societies at large.

Demolinguistics helps in analyzing questions of language and ethnicity in various ways. First of all, large-scale language data can offer key information on both the distribution and vitality of languages as core values of ethnic identity for different groups in multicultural societies (see Extra & Yağmur 2004: 100–132 for methodological considerations about the concepts of distribution and vitality of languages). Comparison of large-scale language data collected at different intervals (in particular census data) has led to studies of shift from minority/dominated languages to majority/dominant languages. De Vries (1990) and Veltman (1991) provide early accounts of Canadian census data, in particular with respect to the measurement of the concepts of mother tongue, home language use, and language speaking ability. There is a huge body of knowledge in English-dominant immigration countries on patterns of intergenerational shift from languages other than English (LOTE) to English in different groups. Clyne (2003: 23–46) offers analyses of Australian census data on language shift in first (overseas-born) and second (Australian-born) generations. Language data can be cross-tabulated with other types of data on individual and group factors collected at the same

time, e.g., data on exogamous *vs.* endogamous marriages, age, period of residence, gender, English proficiency, education, group size and group spread.

In this chapter, we want to compare the European state of knowledge on mapping linguistic diversity in multicultural contexts with initiatives taken in other parts of the world. Over the last century, Europe has shifted from being a continent of emigration to being a continent of immigration. Multicultural self-definitions have been created by former European immigrants in such non-European English-dominant immigration countries as Australia, Canada, the USA and South Africa (cf. the concept of “rainbow nation”). A similar multicultural self-definition also holds for Europe at large: its identity is commonly described in terms of “celebrating cultural and linguistic diversity”. A paradoxical phenomenon in the European public and political discourse is the absence of this celebration in the case of non-European immigrant groups and their languages. From a European perspective, much can be learnt from the experiences abroad in dealing with multilingualism and multiculturalism, both in terms of public and political discourse and in terms of data provision and data analysis.

Both in Europe and beyond, there is huge variation in the types of databases for the definition and identification of population groups in multicultural societies. Some of these databases may include language data, derived from a variety of single or multiple language questions. In the European context, Poulain (2008) makes a distinction between nationwide censuses, administrative registers and statistical surveys. Censuses take place at fixed intervals (commonly 5 or 10 years) and result in nationwide databases. Administrative registers are commonly built up at both the municipal and central level, and they are commonly updated every year. Statistical surveys may be carried out at regular intervals among particular subsets of population groups. All three types of data collection may finally take place in various combinations.

Census data are common phenomena across the world. In this chapter our focus is on census data in European Union (henceforward EU) countries, including a case study on the United Kingdom census in 2001 and 2011, and in Australia, Canada, South Africa and the USA. To give an example: the *Linguistic Atlas of South Africa* (Van der Merwe & Van der Merwe 2006) is based on comparative analyses of census data across space and time, collected at the national, provincial and metropolitan level (Cape Town) at intervals of 10 years (i.e., in 1991 and 2001). Register data at the national and municipal level commonly contain data on nationality and birth country (of municipal inhabitants and/or their parents in a given year). Sweden, Denmark and the Netherlands are examples of EU countries in which such information is collected and updated yearly through registers instead of censuses. European examples of survey data on home language use of school population groups have been analyzed and published in the context of the *Multilingual Cities Project* or MCP (Extra & Yağmur 2004) and in the context of *Local Educational Authorities* (LEA) projects in the United Kingdom (Baker & Eversley 2000). The MCP offers cross-national data on language distribution and language vitality derived from multiple language questions that have been collected among school populations in six multicultural European cities. LEA data on pupils’ home language use will be discussed in this chapter.

In the context of our reference to nation-states, we will refer to nationality rather than citizenship. Even if the two concepts are often used as synonyms nowadays, we should be aware of their historical and contextual difference in denotation (Guiguet 1998). Nationals belong to a nation-state but they may not have all the rights linked with citizenship (e.g., voting rights); in this sense, citizenship is a more inclusive concept than nationality.

Criteria for the definition and identification of population groups in multicultural contexts

Collecting reliable information on the diversity of population groups in multicultural contexts is no easy enterprise. What is, however, more interesting than numbers or estimates of the size of particular groups, is what the criteria are for determining such numbers or estimates. Comparative information on population figures in EU member-states can be obtained from the Statistical Office of the EU in Luxembourg (EuroStat). An overall decrease of the indigenous population has been observed in most EU countries over the last decade; at the same time, there has been an increase in the IM figures. For a variety of reasons, however, reliable and comparable demographic information on IM groups in EU countries is difficult to obtain. Seemingly simple questions like *How many Turkish residents live in Germany compared to France?* cannot easily be answered. For some groups or countries, no updated information is available or no such data have ever been collected. Moreover, official statistics only reflect IM groups with legal resident status. Another source of disparity is the different data collection systems being used, ranging from census data to administrative registers or statistical surveys. In addition, most residents from former colonies already have the nationality of their country of immigration. Most importantly, however, the widely used criteria for IM status – nationality and/or country of birth – have become less valid over time because of an increasing trend towards naturalization and births within the countries of residence.

For a discussion of the role of censuses in identifying population groups in a variety of multicultural nation-states, we refer to Kertzer & Arel (2002). Alterman (1969) offers a fascinating account of the history of counting people from the earliest known records on Babylonian clay tablets in 3800 BC to the USA census in 1970. In addition to the methods of counting, Alterman discusses at length who were counted and how, and who were not counted and why. The issue of mapping identities through nationwide periodical censuses by state institutions is commonly coupled with a vigorous debate between proponents and opponents about the following “ethnic dilemma”: how can you combat discrimination if you do not measure diversity? (Kertzer & Arel 2002: 23–25). Among minority groups and academic groups, both proponents and opponents of mapping diversity can be found: proponents argue in terms of the social or scientific need for population data bases on diversity as prerequisites for affirmative action by the government in such domains as labor, housing, health care, education or media policies; opponents argue in terms of the social or scientific risks of public or political misuse of such data bases for stereotyping, stigmatization, discrimination or even removal of the “unwanted other”. Kertzer & Arel (2002: 2) argue that the census does much more than simply reflect social reality; rather than merely reflecting it, it plays a key role in the construction of that reality and in the creation of collective identities. At the same time, it should be acknowledged that the census is a crucial area for the politics of representation. Census data can make people aware of under-representation. Minority groups often make language rights one of their key demands on the basis of (home) language databases.

Decennial censuses became common practice in Europe and the New World colonized by Europeans in the first part of the 19th century. The USA became the first newly established nation-state with a decennial census since 1790. The first countries to include a language question in their census, however, were Belgium in 1846 and Switzerland in the 1850s, both being European countries with more than one official state language. At present, in many EU countries, only population data on nationality and/or birth country (of person and/or parents) are available on IM groups. In 1982, the Australian Institute of Multicultural Affairs recognized the above-mentioned identification problems for inhabitants of Australia and

proposed including questions in the Australian census on birth country (of person and parents), ethnic origin (based on self-categorization in terms of which ethnic group a person considers him/herself to belong to), and home language use. In Table 2, the four criteria mentioned are discussed in terms of their major (dis)advantages.

Table 2. Criteria for the definition and identification of population groups in a multicultural society (P/F/M = person/father/mother) (Extra & Gorter 2001: 9)

Criterion	Advantages	Disadvantages
Nationality (NAT) (P/F/M)	<ul style="list-style-type: none"> • objective • relatively easy to establish 	<ul style="list-style-type: none"> • (intergenerational) erosion through naturalization or double NAT • NAT not always indicative of ethnicity/identity • some (e.g., ex-colonial) groups have NAT of immigration country
Birth country (BC) (P/F/M)	<ul style="list-style-type: none"> • objective • relatively easy to establish 	<ul style="list-style-type: none"> • intergenerational erosion through births in immigration country • BC not always indicative of ethnicity/identity • invariable/deterministic: does not take into account boundary changes in society (in contrast to all other criteria)
Self-categorization / ethnicity (SC)	<ul style="list-style-type: none"> • touches the heart of the matter • emancipatory: SC takes into account person's own conception of ethnicity/identity 	<ul style="list-style-type: none"> • subjective by definition: also determined by the language/ethnicity of interviewer and by the spirit of the times • multiple SC possible • historically charged, especially by World War II experiences
Home language (HL)	<ul style="list-style-type: none"> • HL is significant criterion of ethnicity in communication processes • HL data are prerequisite for government policy in areas such as public information or education 	<ul style="list-style-type: none"> • complex criterion: who speaks what language to whom and when? • language is not always a core value of ethnicity/identity • useless in one-person households

First of all, Table 2 reveals that there is no simple road to solving the identification problem. Moreover, inspection of the criteria for multicultural population groups is as important as the actual figures themselves. Seen from a European perspective, there is a top-down development over time in the utility and utilization of different types of criteria, inevitably going from nationality and birth-country criteria in present statistics to self-categorization and home language in the future. The latter two criteria are generally conceived of as being complementary criteria. Self-categorization and home language references need not coincide, as languages may be conceived to variable degrees as core values of ethnocultural identity in minority or migration contexts.

Mapping diversity in non-European English-dominant immigration countries

Various types of criteria for identifying population groups in multicultural societies have been suggested and used outside Europe in countries with a longer immigration history, and as a result of this, with a longstanding history of collecting census data on multicultural population groups (Kertzer & Arel 2002). This holds in particular for non-European immigration

countries in which English is the dominant language, like Australia, Canada, South Africa, and the USA. To identify the multicultural composition of their populations, these countries employ a variety of questions in their periodical censuses. In Table 3, an overview of (clusters of) questions is provided; for each country, the given census is taken as the norm.

Table 3. Overview of (clusters of) census questions in four multicultural countries (Extra & Yağmur 2004: 67)

Questions in the census	Australia 2001	Canada 2001	SA 2001	USA 2000	Coverage
1 Nationality of respondent	+	+	+	+	4
2 Birth country of respondent	+	+	+	+	4
3 Birth country of parents	+	+	-	-	2
4 Ethnicity	-	+	-	+	2
5 Ancestry	+	+	-	+	3
6 Race	-	+	+	+	3
7 Mother tongue	-	+	-	-	1
8 Language used at home	+	+	+	+	4
9 Language used at work	-	+	-	-	1
10 Proficiency in English	+	+	-	+	3
11 Religious denomination	+	+	+	-	3
Total of dimensions	7	11	5	7	30

Both the type and the number of questions are different for each of these countries. Canada takes up a prime position with the highest number of questions. There are only three questions that are asked in all countries while two questions are asked in only one country. There are four different questions asked about language. The operationalization of questions also shows interesting differences, both between and within countries over time (see Clyne 1991 for a discussion of methodological problems in comparing the answers to differently phrased questions in Australian censuses from a longitudinal perspective).

Questions about ethnicity, ancestry and/or race have proven to be problematic in all of the countries under consideration. In some countries, ancestry and ethnicity have been conceived of as equivalent, cf. USA census question 10 in 2000: *What is this person's ancestry or ethnic origin?* Or, take Canadian census question 17 in 2001: *To which ethnic or cultural group(s) did this person's ancestors belong?* Australian census question 18 in 2001 only involved ancestry and not ethnicity, cf. *What is the person's ancestry?* with the following comments for respondents: *Consider and mark the ancestries with which you most closely identify. Count your ancestry as far as three generations, including grandparents and great-grandparents. As far as ethnicity and ancestry have been distinguished in census questions, the former concept related most commonly to present self-categorization of the respondent and the latter to former generations. The diverse ways in which respondents themselves may interpret both concepts, however, remains a problem that cannot easily be solved.*

According to Table 3, South Africa remains as the only country where a racial question is asked instead of a question on ethnicity and/or ancestry. The paradox in South Africa is that questions on ethnicity are often considered to be racist, while the racial question (in terms of *Black/White/Colored/Indian*) from the earlier Apartheid era has survived. Although the validity of questions about ethnicity, ancestry and/or race is problematic, at least one question from this cluster is needed to compare its outcomes with those of questions on language. Language is not always a core value of ethnicity/identity and multiculturalism may become under-estimated if it is reduced to multilingualism. For this reason, one or more questions

derived from cluster 4–6 in Table 3 are necessary complements of one or more questions derived from cluster 7–10.

Although, according to Table 3, “ethnicity” is mentioned in recent censuses of only two countries, all four language-related questions are asked only in Canada. Over time, a “mother tongue” question has been replaced by a “home language” question in three out of four countries. Canada has retained the mother tongue question in addition to the home language question, which allows for comparative analyses of predictably different outcomes. The mother tongue question (7) in Canada is defined for respondents as *the language first learnt at home in childhood and still understood*, whereas questions 8 and 9 are related to the language *most often* used at home/work. Table 3 shows the added value of language-related census questions for the definition and identification of multicultural populations, in particular the added value of the question on home language use compared to questions on the more opaque concepts of mother tongue and ethnicity. Although the language-related census questions in the four countries under consideration differ in their precise formulation and commentary, the outcomes of these questions are generally conceived as cornerstones for educational policies with respect to the teaching of English as a first or second language and the teaching of languages other than English.

Table 3 also shows the importance of comparing different groups using equal criteria. Unfortunately, this is often not the case in the public or political discourse. Examples of such unequal treatment are references to *Poles vs. Jews*, *Israelis vs. Arabs*, *Serbs and Croatians vs. Muslims*, *Dutchmen vs. Turks* (for Dutch nationals with Turkish ethnicity), *Dutchmen vs. Muslims*, or *Islam vs. the West* (where does the West end when the world is a globe?). Equal treatment presupposes reference to equal dimensions in terms of Table 3.

Mapping diversity in European Union countries

The data presented here has been derived from the analysis of two comprehensive documents published by the European Commission and EuroStat (2004; 2005). In 23 out of 27 EU countries nationwide censuses held at variable intervals are still in use. Scandinavian countries and the Netherlands rely on yearly updated administrative (municipal) registers in combination with periodical sample surveys. Other countries combine nationwide census data with administrative register data and/or sample survey data (Austria, Belgium, France, Latvia, Slovenia). The following parameters are used in all or many EU countries for the definition and identification of population groups:

- *(dual) citizenship or nationality*: the category of (dual) citizenship or nationality is used in all EU countries; in North-Western European countries these two concepts are commonly used as synonyms nowadays; in Southern, Middle and Eastern European countries, however, these two categories are commonly used distinctively: in these countries citizenship refers to what is termed nationality or citizenship in North-Western Europe, whereas (ethnic) nationality refers to what is termed ethnicity in North-Western Europe; in the Czech Republic, for instance, the nationality question asks for an indication of *what nationality you consider yourself to be*, which is different from the citizenship question;
- *birth country/place*: this is a common category in all EU countries; in some countries, this question refers explicitly to the country/place of (permanent) residence of *your mother when you were born*;

- *ethnicity*: ethnicity or ethnic nationality is asked for in 13 EU countries, 3 of which consider this question to be voluntary/optional;
- *language*: one or more language questions are asked in 17 EU countries, 2 of which consider this/these question(s) to be voluntary/optional;
- *religious denomination*: religious denomination questions are asked in 15 EU countries, 6 of which consider this question to be voluntary/optional.

Table 4 gives an overview of the *status quo* with respect to the latter three parameters across EU countries. It should be noted that data collection on some or all of these questions in some EU countries is considered to be in conflict with privacy legislation and/or illegal, while in other countries such questions are taken to generate crucial information.

Table 4. Identification of ethnicity, language and religion in 27 EU countries (Extra & Gorter 2008: 19)

EU countries	Ethnicity/ ethnic nationality	Language	Religion
Austria	–	+	+
Belgium	–	–	–
Bulgaria	*	*	*
Cyprus	+	+	+
Czech Republic	+	+	+
Denmark	–	–	–
Estonia	+	+	*
Finland	–	+	+
France	–	–	–
Germany	–	–	+
Greece	–	–	–
Hungary	*	*	*
Ireland	+	+	+
Italy	–	–	–
Latvia	+	+	–
Lithuania	+	+	+
Luxembourg	–	–	–
Malta	–	+	–
Netherlands	–	–	–
Poland	+	+	–
Portugal	–	–	*
Romania	+	+	+
Slovakia	+	+	+
Slovenia	*	+	*
Spain	–	+	–
Sweden	–	–	–
United Kingdom	+	+	*

* = voluntary/optional question

Table 4 shows that there is strong variability across countries in the utilization of each of these three parameters. This holds also for the operationalization of questions asked. Detailed ethnicity questions are asked in the UK and Ireland. In the UK, 5 categories are distinguished, i.e., *White*, *Mixed*, *Asian (British)*, *Black (British)*, *Chinese/Other*, in all cases with subcategories. Similar questions are asked in Ireland. Hungary lists 14 categories (plus *other*) and asks *which of these nationalities' cultural values and traditions do you feel affinity with?*

Estonia lists 6 (plus *other*) *ethnic nationalities*, and Cyprus lists *Greek-Cypriot*, *Armenian*, *Maronite*, *Latin* and *Turkish-Cypriot*. Questions on religion are asked in terms of *belief*, *church*, *faith*, *religion* and/or *religious affiliation/community/confession/denomination*, and in terms of *religion/religious denomination you were brought up in*. The latter – additional – question is only asked in Scotland, not in the UK at large. Table 5 shows an overview of the operationalization of the language questions asked for in 17 out of 27 EU countries.

Table 5. Operationalization of language questions in 17 EU countries (Extra & Gorter 2008: 20)

EU countries	Mother tongue	(Other) language(s) spoken (frequently)	Language(s) (most frequently) spoken at home	Language(s) spoken with family or friends	Speak well/ average/ a little	Understand/ Speak/ Read/ Write
Austria	–	–	+	–	–	–
Bulgaria	+	–	–	–	–	–
Cyprus	–	+	–	–	–	–
Czech Republic	(1)	–	–	–	–	–
Estonia	+	+	–	–	–	–
Finland	+	–	–	–	–	–
Hungary	+	+	–	+	–	–
Ireland	–	(2)	–	–	–	–
Latvia	+	+	–	–	–	–
Lithuania	+	+	–	–	–	–
Malta	–	–	+	–	+	–
Poland	–	–	+	–	–	–
Romania	+	–	–	–	–	–
Slovakia	+	–	–	–	–	–
Slovenia	+	–	+	–	–	–
Spain	(3)	–	(3)	–	–	(4)
United Kingdom	–	–	–	–	–	(5)

- (1) Indicate the language spoken by your mother or guardian when you were a child
- (2) Only Irish; if yes, daily within/outside the educational system/weekly/less often/never
- (3) Both language questions in the Basque County, Navarre and Galicia, for Basque/ Galician
- (4) In Catalonia, Valencia and the Balearic Islands for Catalan
- (5) Only in Wales and Scotland, for Welsh and Gaelic respectively

Three main conclusions emerge from Table 5. First of all, European census questions on non-national languages focus on RM languages, not on IM languages. Secondly, the three most commonly asked questions on language use relate to mother tongue (11 countries), (other) language(s) spoken (frequently) (6 countries) and language(s) (most frequently) spoken at home (5 countries). Thirdly, Hungary makes the most investments in finding out about language use. For a complete picture, it should be mentioned that in some countries collecting home language data is in fact in conflict with present language legislation. This holds in particular for Belgium, where no census data on language use have been collected since 1947 and traditional language borders between Dutch, French and German have been allocated and fixed in the law.

Case study on census questions in the United Kingdom

Although the UK has been taking censuses since 1801 at intervals of 10 years, there have as yet been no nationwide language questions. However, home language data have been collected in periodical surveys held among school children, in particular at the initiative of Local Educational Authorities (LEA). Baker & Eversley (2000) carried out a re-analysis of LEA home language data from the 1998/1999 school year involving more than 850,000 primary school children in London, and were able to identify more than 350 different home languages. In their report, they not only offered a comprehensive description of these languages and their distribution in the different LEA districts in London, but also dealt with a number of methodological issues in analyzing home language data.

The first time a question on ethnicity was included in the UK census was in 1991. In the 2001 census, the question was phrased as *What is your ethnic background?* but the options given (*White/Mixed/Asian/Black or Black British/Chinese/Other ethnic group*) suggest race rather than ethnicity. In explaining why such a question was necessary, the word “racial” rather than “ethnic” was used. Responses to the question were to provide baseline figures against which the Government could monitor possible racial disadvantage and measure changes over time. In the latest census of 2001, a language question was asked only in Wales on Welsh language skills: *Can you understand, speak, read, or write Welsh?* Similar questions were asked in Scotland on Scottish Gaelic. The 2001 census was the first that included a question on religious affiliation, based on subjective subscription. The question did not intend to determine religious belief or practice. Moreover, it was voluntary and intended to complement the data obtained from the questions on ethnic group and country of birth.

The Office for National Statistics (ONS) in London is responsible for the planning of the next UK Census in 2011, which for the first time will include a nationwide question on languages other than English. A special ONS team is in charge of preparing the census questions on “ethnicity, identity, language and religion”, referred to as EILR. On its website, the ONS offers detailed information on the development of the EILR questions for the 2011 Census derived from consultations with data users and providers and from qualitative and quantitative pilot question testing. Forsyth et al. (2008) present a conceptual basis for measuring EILR, which gives rise to the following comments given the considerations presented in this chapter:

- From a conceptual point of view, the I in EILR is rather odd, as identity is generally conceived of as a cover concept for dealing with ethnicity, language and religion as its constituent dimensions. There was no question on “national identity” in the previous census and the ONS appears to be determined to include it in the 2011 census, in spite of all objections. The rationale for it is based on rather traditional British reasoning and leads to questions that immigrant groups would consider to be confusing, biased, ethnocentric or condescending.
- Multiplicity is potentially a property of almost all the dimensions discussed. It obviously holds for ethnicity, nationality and language rather than for religion. Reference to “mixed” ethnic groups should be substituted by reference to multiple ethnic groups. The same holds *a fortiori* for language groups.
- For religious affiliation, subjective subscription holds as well and, mirrored to the proposed ethnicity question, would result in *To which religious group do you consider yourself to belong?* Such a question would also value the concepts of “group” and “belonging”.

- There is a clear ranking in the recommended questions reported by respondents to the 2006/2007 Census Stakeholders Consultation, i.e., 58% are in favor of a nationality question, more than 80% are in favor of both an ethnic group question and a religious affiliation question, and almost 90% are in favor of a language question. Given these facts, i.e., the multiplicity fact for the concepts of both ethnicity and language plus the highest ranking of respondents' preference for a language question, there are remarkable and paradoxical differences in detail between the proposed ethnicity and language questions: For ethnicity there is a whole range of pre-coded groups, including multiplicity, whereas the language question is not worked out in such detail.
- There are at least five good reasons to favor a home language question over a main language question:
 - the proposed question *What is your main language?* is a very ambiguous one because the key informant might take a wide variety of contexts and/or interlocutors into consideration;
 - the question should be depersonalized from heads of households: for this reason, the home language question in many countries is phrased in the form of a passive construction, e.g., *What is/are the main language(s) spoken in your home?*
 - the main language question will predictably lead to a serious underestimate of LOTE in the home;
 - the home language question is widely used in non-European English-dominant immigration countries, which is in support of its validity and which allows for cross-national analyses of its outcomes;
 - the home language question is also widely supported by the informed community of sociolinguistic UK researchers in this field (BAAL 2005).
- There are at least two good arguments for providing a pre-coded list of the top-10 of IM languages next to RM languages in the UK, plus an additional write-in box of *Other*. First of all, the top-10 of IM languages will cover a significant proportion of all the answers given, estimated to be at least 70% or even 80% (Extra & Yağmur 2004: 196–200). Moreover, a pre-coded list of languages will make informants aware of the distinction between language and country (e.g., “Moroccan” or “Congolese” are not languages; Extra & Yağmur 2004: 201–202).
- A deficit perspective on languages other than English should be avoided. LOTE data can have many applications with an intrinsic value, e.g., in terms of making LOTE provisions available in education, in the media, in courts or in medical care/hospitals. Some countries make ethnic broadcasting provisions available on the basis of distributional data on home language use.

Conclusions and outlook

In this chapter, we discussed concepts, paradigms, methods and databases for mapping linguistic diversity in multicultural contexts. In terms of databases, we made a distinction between census data, register data and survey data. Our focus was on census data, in particular on various types of language data. Research at this international crossing of demography and linguistics is referred to as demolinguistics.

We compared the European state of knowledge in this domain with initiatives in other parts of the world. Taken from this international comparative perspective, our focus was on mapping diversity in EU countries and in non-European English-dominant immigration countries, in particular Australia, Canada, South Africa and the USA. We discussed four criteria for the definition and identification of population groups in a multicultural society, in terms of nationality/citizenship, birth country, ethnicity and home language. Both advantages and disadvantages of each of these criteria were presented, and the utilization of these criteria was illustrated for both European and non-European contexts.

In the European context, if censuses take place at all, census data may or may not include language data. If they do, the focus tends to be on mother tongue data and, in the case of minority groups, on RM languages. In the above-mentioned non-European contexts, census data do include language data, with a focus on home language data and, in the case of minority groups, on IM languages. The higher validity of home language data and the increasing presence of IM groups and IM languages across Europe will lead to European databases on population groups that will inevitably follow these longstanding non-European examples. We also presented a case study on census questions in the UK, and illustrated some key issues in the prospective UK census in 2011 with a number of considerations in this chapter.

An important similarity in questions about home language use, whether in census, register or survey planning, is that the outcomes are based on reported rather than observed facts. Answers to questions on home language use may be colored by the language of the questions themselves (which may or may not be the primary language of the respondent), by the ethnicity of the interviewer (which may or may not be the same as that of the respondent), by the (perceived) goals of the sampling (which may or may not be defined by national or local authorities), and by the spirit of the times (which may or may not favor multiculturalism). These problems become even more prominent in a school context in which pupils are the respondents. Apart from the problems mentioned, the answers may be colored by peer-group pressure and they may lead to interpretation problems in attempts to identify and classify languages. The problems referred to are inherent characteristics of large-scale data gathering through questionnaires about language-related behavior and should be compensated by small-scale data gathering through observation of actual language behavior. Such small-scale ethnographic research is not an alternative to large-scale language surveys, but a necessary complement, which can offer rich details and variation at the individual level.

Given the decreasing significance of nationality and birth-country criteria in the European context, the combined criteria of self-categorization (ethnicity) and home language use are promising alternatives for obtaining basic information on the increasingly multicultural composition of European nation-states. Across Europe, it is common practice to present data on RM groups on the basis of (home) language use and/or ethnicity, and to present data on IM groups on the basis of nationality and/or country of birth. However, *convergence* between these criteria for the two groups emerges over time, due to the increasing period of migration and minorization of IM groups in particular countries. As a result of their prolonged/permanent stay, there is strong erosion in the utility of nationality or birth-country

statistics. The added value of home language statistics, whether originating from census data, register data or survey data, derives from at least four different perspectives (Extra & Gorter 2008: 316):

- taken from a *demographic* perspective, home language data play a crucial role in the definition and identification of population groups in a multicultural society;
- taken from a *sociolinguistic* perspective, home language data offer valuable insights into both the distribution and the vitality of home languages across different population groups, and thus raise the public awareness of multilingualism;
- taken from an *educational* perspective, home language data are indispensable tools for educational planning and policies;
- taken from an *economic* perspective, home language data offer latent resources that can be built upon and developed in terms of economic opportunities.

To conclude, home language data put to the test any monolingual mindset in a multicultural society and can function as agents of change (Nicholas 1994) in a variety of public and private domains.

Questions for Further Thought and Discussion

1. Why do questions on language and ethnicity in population research lead to complementary databases?
2. Why are questions on ethnicity in population research more difficult to formulate and to answer than questions on language use?
3. Discuss advantages and disadvantages of census data, register data and survey data on diversity in multicultural societies.
4. Discuss the arguments of proponents and opponents of mapping diversity in multicultural societies. How would you define and justify your own position in this debate?
5. What statistical data are collected on diversity of population groups in your own country and how would you evaluate the questions in your country's data collection system(s) against the questions in Table 3 of this chapter?
6. Discuss the validity of home language questions compared to mother tongue questions in population research.
7. How would you evaluate the added value of home language statistics as presented in this chapter compared to the value of statistics on ethnicity?
8. In what ways can home language data in a multicultural society function as agents of change?

Selected Bibliography

- Alterman, Hyman (1969). *Counting People. The Census in History*. New York: Harcourt.
- BAAL (2005). Letter to Office of National Statistics, printed in *Baal News*, 81: 18–22.
- Baker, Philip, and John Eversley (eds.) (2000). *Multilingual Capital. The Languages of London's Schoolchildren and their Relevance to Economic, Social and Educational Policies*. London: Battlebridge Publications.
- Barni, Monica, and Guus Extra (eds.) (2008). *Mapping Linguistic Diversity in Multicultural Contexts*. Berlin/New York: Mouton de Gruyter.
- Clyne, Michael (1991). *Community Languages: The Australian Experience*. Cambridge: Cambridge University Press.
- Clyne, Michael (2003). *Dynamics of Language Contact*. Cambridge: Cambridge University Press.
- De Vries, John (1990). On coming to our census: a layman's guide to demolinguistics. *Journal of Multilingual and Multicultural Development* 11 (1–2): 57–76.
- European Commission and EuroStat (2004). *Documentation of the 2000 Round of Population and Housing Censuses in the EU, EFTA and Candidate Countries, Parts I-III and Annexes*. Luxembourg: Office for Official Publications of the European Communities.
- European Commission and EuroStat (2005). *Population and Housing Censuses 2001. Results at National and Regional Level with Documentation and Detailed Tables*. Luxembourg: Office for Official Publications of the European Communities (CD-ROM).
- Extra, Guus, and Durk Gorter (eds.) (2001). *The Other Languages of Europe. Demographic, Sociolinguistic and Educational Perspectives*. Clevedon: Multilingual Matters.
- Extra, Guus, and Durk Gorter (eds.) (2008). *Multilingual Europe. Facts and Policies*. Berlin/New York: Mouton de Gruyter.
- Extra, Guus, and Kutlay Yağmur (eds.) (2004). *Urban Multilingualism in Europe: Immigrant Minority Languages at Home and School*. Clevedon: Multilingual Matters.
- Forsyth, Frances, Ben Bradford, Aleks Collingwood and Hamish MacPherson (2008). *Conceptual Basis for Measuring Ethnicity, Identity, Language and Religion*. London: Office for National Statistics.
- Guiguet, Benoît (1998). Citizenship and nationality: Tracing the French roots of the distinction. In *European Citizenship, An Institutional Challenge*, Massimo la Torre (ed.), 95–111. The Hague: Kluwer Law International.
- Kertzer, David, and Dominique Arel (2002). *Census and Identity. The Politics of Race, Ethnicity, and Language in National Censuses*. Cambridge: Cambridge University Press.
- Nicholas, Joe (1994). *Language Diversity Surveys as Agents of Change*. Clevedon: Multilingual Matters.
- Poulain, Michel (2008). European migration statistics: definitions, data and challenges. In *Mapping Linguistic Diversity in Multicultural Contexts*. Monica Barni and Guus Extra (eds.), 43–66. Berlin/New York: Mouton de Gruyter.
- Van der Merwe, Izak (1989). Geolinguistics of Afrikaans in the metropolitan area of Cape Town. *South African Journal of Linguistics* 7: 92–96.
- Van der Merwe, Izak, and Johannes van der Merwe (ed.) (2006). *Linguistic Atlas of South Africa: Language in Space and Time*. University of Stellenbosch: Sun Press.
- Veltman, Calvin (1991). Theory and method in the study of language shift. In *Language and Ethnicity*, Volume II, James R. Dow (ed.), 145–167. Amsterdam/Philadelphia: John Benjamins.