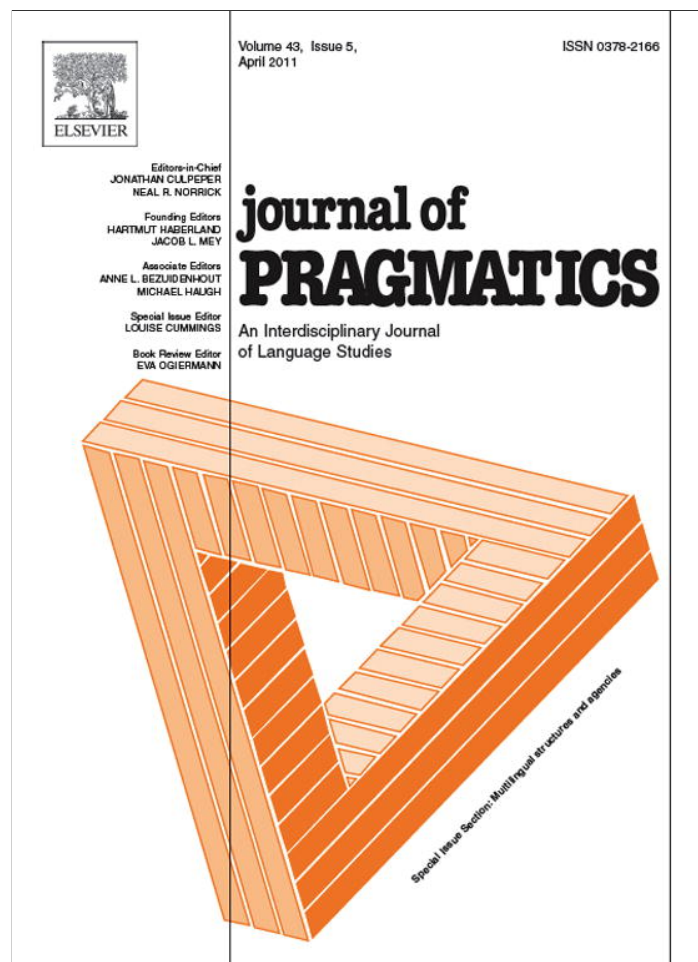


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

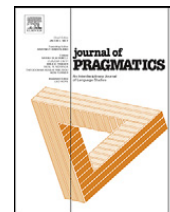
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Journal of Pragmatics

journal homepage: [www.elsevier.com/locate/pragma](http://www.elsevier.com/locate/pragma)

# Urban multilingualism in Europe: Mapping linguistic diversity in multicultural cities

Guus Extra<sup>\*</sup>, Kutlay Yağmur

Tilburg University, Department of Culture Studies, Warandelaan 2, 5037 AB Tilburg, The Netherlands

## ARTICLE INFO

### Article history:

Received 17 February 2009

Received in revised form 5 February 2010

Accepted 5 October 2010

Available online 15 December 2010

### Keywords:

Crosslinguistic perspectives

Data collection and processing

Distribution of languages

Language profiles

Language vitality

Multilingual Cities Project

## ABSTRACT

The focus of this paper is on urban multilingualism in Europe. Crossnational and crosslinguistic perspectives are offered on both the distribution and the vitality of immigrant minority languages among primary school children in six multicultural cities across European nation-states. Going from North to South, these cities are Göteborg, Hamburg, The Hague, Brussels, Lyon and Madrid.

The paper comprises eight sections. The first two sections offer the demolinguistic background against which the project was carried out. Section 1 deals with the European constellation of languages and population groups, section 2 with the importance of the language criterion for identifying population groups. The next sections focus on the rationale and goals of the *Multilingual Cities Project*, carried out under the auspices of the *European Cultural Foundation* (section 3), the method of research in the MCP (section 4), the distribution of languages across cities (section 5), the specification of language profiles and language vitality (section 6), crosslinguistic perspectives on language vitality (section 7), and conclusions (section 8).

© 2010 Elsevier B.V. All rights reserved.

## 1. Introduction

Europe's identity is determined to a great extent by cultural and linguistic diversity (Haarmann, 1995). Within the European Union (henceforward EU), there are large differences in population size amongst EU nation-states. German, French, English, Italian, Spanish and Polish belong to the six most widely spoken national languages in the present EU, whereas Turkish would come second to German in an enlarged EU. There is also a close connection between nation-state references and official state language references. Distinct languages are the clearest feature distinguishing one nation-state from its neighbors (Barbour, 2000), the major exceptions (and for different reasons) being Belgium, Austria, and Cyprus. This match between nation-state references and official state language references obscures the very existence of different types of minority languages that are actually spoken across European nation-states. Many of these languages are indigenous minority languages with a regional base, many other languages stem from abroad without such a base. We will refer to these "other" languages of Europe as regional minority (henceforward RM) languages and immigrant minority (henceforward IM) languages, respectively (Extra and Gorter, 2001, 2008).

A number of issues need to be kept in mind, however. Within and across EU nation-states, some RM and IM languages have larger numbers of speakers than some of the official state languages. Moreover, both RM and IM languages in one EU nation-state may be official state languages in another nation-state. Examples of the former result from language border

<sup>\*</sup> Corresponding author. Tel.: +31 13 466 3122; fax: +31 13 466 2892.

E-mail addresses: [guus.extra@uvt.nl](mailto:guus.extra@uvt.nl) (G. Extra), [k.yagmur@uvt.nl](mailto:k.yagmur@uvt.nl) (K. Yağmur).

crossing in adjacent nation-states, such as Finnish in Sweden or Swedish in Finland. Examples of the latter result from transnational processes of migration and minorization, in particular from Southern to Northern Europe, such as Portuguese, Spanish, Italian or Greek. In particular the context of migration and minorization makes our proposed distinction between RM and IM languages ambiguous. We see, however, no better alternative. It should also be kept in mind that many, if not most, IM languages in particular European nation-states originate from countries outside Europe. In our opinion, the proposed distinction leads at least to awareness raising and may ultimately lead to an inclusive approach in the European conceptualization of “minority” languages.

There have always been speakers of IM languages in Europe, but these languages have only recently emerged as community languages spoken on a wide scale in urban Europe, due to intensified processes of migration and minorization. Turkish and Arabic are good examples of so-called “non-European” languages that are spoken and learned by millions of inhabitants of the EU nation-states. Although IM languages are often conceived of and transmitted as core values by IM language groups, they are much less protected than RM languages by affirmative action and legal measures, for example, in education. In fact, the learning and certainly the teaching of IM languages are often seen by mainstream language speakers and by policy makers as obstacles to integration. At the European level, guidelines and directives regarding IM languages are scant and outdated. Despite the possibilities and challenges of comparing the status of RM and IM languages across European nation-states, amazingly few connections have been made in sociolinguistic, educational and political domains. As yet, we lack a common referential framework for the languages under discussion. Publications which focus on both types of minority languages are rare: examples are the dual volumes on RM and IM languages by Alladina and Edwards (1991), and the integrated volumes by Gogolin et al. (1991), Fase et al. (1992, 1995), Ammon et al. (1995), Ammerlaan et al. (2001), Extra and Gorter (2001, 2008) and Barni and Extra (2008).

Comparative information on population figures in EU member-states can be obtained from the Statistical Office of the EU in Luxemburg (*Eurostat*). 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 (Poulain, 2008). 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 nation-wide census data to administrative registers or to more or less representative surveys. Most importantly, however, the most widely used criteria for IM status – nationality and/or country of birth – have become less valid over time because of an increasing trend toward naturalization and births within the countries of residence. In addition, most residents from former colonies already have the nationality of their country of immigration.

For a discussion of the role of censuses in identifying population groups in a variety of multicultural nation-states, we refer to Kertzer and Arel (2002). Alterman (1969) offers a fascinating account of the history of counting people from the earliest known records on Babylonian clays in 3800 BC to the USA census in 1970. Besides the methods of counting, Alterman discusses at length who has been counted, and how, who not, 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 and Arel, 2002:23–25). Both proponents and opponents of measuring diversity can be found (cf. Blum, 2002 on this debate in France):

- proponents argue in terms of the social or scientific need for population databases on diversity as prerequisites for affirmative action by 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 databases for stereotyping, stigmatization, discrimination or even removal of the “unwanted other”.

Kertzer and Arel (2002:2) show that the census does much more than simply reflect social reality; rather 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. Language rights are often a key demand for minority groups on the basis of (home) language databases.

Decennial censuses became a 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 (Extra and Gorter, 2008).

## 2. The importance of the language criterion for identifying population groups

Complementary or alternative criteria for identifying population groups in a multicultural society have been suggested and used in countries with a longer immigration history, and, for this reason, with a longstanding history of collecting census data on multicultural population groups (Kertzer and Arel, 2002). This holds in particular for non-European English-dominant immigration countries like Australia, Canada, South Africa, and the USA. To identify the multicultural composition of their populations, these four countries employ a variety of questions in their periodical censuses. In Table 1, an overview of the kernel array of questions is provided; for each country the given census is taken as the norm.

**Table 1**

Overview of kernel census questions in four multicultural contexts (Extra and Yağmur, 2004:67).

Kernel questions in the census	Australia 2001	Canada 2001	South Africa 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 Religion	+	+	+	–	3
Total of dimensions	7	11	5	7	30

Both the type and number of questions are different for each of these countries. Canada has a prime position with the highest number of questions. Only three questions have been asked in all countries whereas two questions have been asked in only one country. Four different questions have been 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 (see also Spencer, 2006; Ansell and Solomos, 2008). 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 current 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 be solved easily.

According to Table 1, 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, whereas 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. The reason for this has been mentioned in Table 1: language is not always a core value of ethnicity/identity and multiculturalism may become under-estimated if reduced to multilingualism. For this reason, one or more questions derived from clusters 4–6 in Table 1 are necessary complements of one or more questions derived from clusters 7–10.

Whereas, according to Table 1, “ethnicity” has been mentioned in recent censuses of only two countries, four language-related questions have been asked in one to four countries. Only in Canada has the concept of “mother tongue” been included (census question 7). It has been defined for respondents as *the language first learnt at home in childhood and still understood*, whereas questions 8 and 9 were related to the language *most often* used at home/work. Table 1 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 1 also shows the importance of comparing different groups with equal criteria. Unfortunately, this is often not the case in public or political discourse. Examples of such unequal treatment are references to *Poles vs. Jews, Israelis vs. Arabs, Serbs and Croats 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 is the world a globe?). Equal treatment presupposes reference to equal dimensions in terms of Table 1.

From this overview, it can be concluded that large-scale home language surveys are both feasible and meaningful, and that the interpretation of the resulting database is made easier by transparent and multiple questions on home language use. These conclusions become even more pertinent in the context of gathering data on multicultural *school* populations. European experiences in this domain have been gathered in particular in Great Britain and Sweden. In both countries,

extensive municipal home language statistics have been collected through local educational authorities by asking pupils and/or parents questions about their oral and written skills in languages other than the mainstream language, and about their need for education in these languages.

An important similarity in the questions about home language use in these surveys 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 the ethnicity of the respondent), by the (perceived) goals of the sampling (which may or may not be defined by central state or local authorities), and by the spirit of the times (which may or may not be in favor of multiculturalism). These problems become even more evident in a school-related context in which pupils are 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 on the basis of the answers given. For a discussion of these and other possible effects, we refer to [Nicholas \(1988\)](#) and [Alladina \(1993\)](#). The problems referred to are inherent characteristics of large-scale data gathering through questionnaires about language-related behavior and can only be compensated by small-scale data gathering through observing actual language behavior. Such small-scale ethnographic research is not an alternative to large-scale language surveys, but a valuable and necessary complement. For a discussion of (cor)relations between the reported and measured bilingualism of IM children in the Netherlands, we refer to [Broeder and Extra \(1998\)](#).

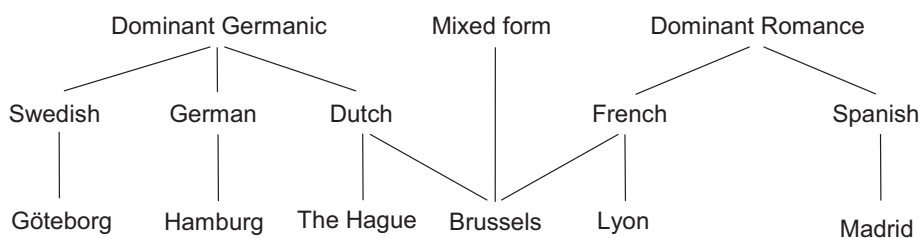
Throughout the EU, it is common practice to present data on RM groups on the basis of (home) language and/or ethnicity, and to present data on IM groups on the basis of nationality and/or country of birth. However, *convergence* between these criteria for the two groups appears over time, due to the increasing period of migration and minorization of IM groups in EU countries. Due to their prolonged/permanent stay, there is strong erosion in the utility of nationality or birth-country statistics. 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 potentially promising alternatives for obtaining basic information on the increasingly multicultural composition of European nation-states. The added value of home language statistics is that they offer valuable insights into the distribution and vitality of home languages across different population groups and thus raise the awareness of multilingualism. Empirically collected data on home language use also play a crucial role in education. Such data will not only raise the awareness of multilingualism in multicultural schools; they are in fact indispensable tools for educational policies on the teaching of both the national majority language as a first or second language and the teaching of minority languages.

### 3. The Multilingual Cities Project: rationale and goals

Given the overwhelming focus on mainstream second language acquisition by IM groups, there is much less evidence on the status and use of IM languages across European nation-states. In contrast to RM languages, IM languages have no established status in terms of period and area of residence. Obviously, typological differences between IM languages across EU nation-states do exist, e.g., in terms of the status of IM languages as EU languages or non-EU languages, or as languages of former colonies. Taken from the latter perspective, e.g., Indian languages are prominent in the United Kingdom, Arabic languages in France, Congolese languages in Belgium, and Surinamese languages in the Netherlands.

Here, we present the rationale, methodology, and kernel outcomes of the *Multilingual Cities Project* (henceforward MCP), a coordinated multiple survey study in six major multicultural cities in different EU nation-states. The project was carried out under the auspices of the *European Cultural Foundation*, established in Amsterdam, and it was coordinated by a research team at *Babylon, Centre for Studies of the Multicultural Society*, at Tilburg University in the Netherlands, in cooperation with universities and educational authorities in all participating cities. The aims of the MCP were to gather, analyze, and compare multiple data on the status of IM languages at home and at school, taken from crossnational and crosslinguistics perspectives. In the participating cities, ranging from Northern to Southern Europe, Germanic or Romance languages have a dominant status in public life. [Fig. 1](#) gives an outline of the project.

The criteria for selecting a city to participate in this multinational study were that it should be prototypical for a multicultural environment with a great variety of IM groups, and that it should offer a university-based research facility that would be able to handle the local data gathering and local data analysis, and the final reporting of the local results. Given the increasing role of municipalities as educational authorities in all partner cities, the project was carried out in close



**Fig. 1.** Outline of the Multilingual Cities Project (MCP).



cooperation between researchers at local universities and local educational authorities. In each partner city, this cooperation proved to be of essential value. The rationale for collecting, analyzing and comparing multiple home language data on multicultural school populations derives from four different perspectives:

- taken from a *demographic* perspective, home language data play a crucial role in the definition and identification of multicultural school populations;
- taken from a *sociolinguistic* perspective, home language data offer valuable insights into both the distribution and 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 chances.

Local reports about the participating cities have been made available for *Göteborg* (Nygren-Junkin and Extra, 2003), *Hamburg* (Fürstenau et al., 2003), *The Hague* (Extra et al., 2001), *Brussels* (Verlot et al., 2003), *Lyon* (Akinç et al., 2004), and *Madrid* (Broeder and Mijares, 2003). For the final crossnational report we refer to Extra and Yağmur (2004).

#### 4. Method of research in the MCP

Except in some countries like Great Britain, Sweden or Switzerland, there is no European tradition of collecting home language statistics on multicultural (school) population groups. In fact, collecting home language data in some countries is even in conflict with present language legislation. This holds in particular for Belgium, where traditional language borders have been allocated and legalized in terms of Dutch, French or German.

Our method of carrying out home language surveys amongst primary school children in each of the six participating cities has profited from experiences in non-European English-dominant immigration countries with nationwide population surveys in which commonly single questions on home language use were asked. In contrast to such questionnaires, our survey was based on multiple rather than single questions on home language use and on crossnationally equivalent questions. In doing this, we aimed at describing and comparing multiple language profiles of major IM communities in each of the cities under consideration.

The questionnaire for data collection was designed after ample study and evaluation of language-related questions in nationwide or large-scale population research in a variety of countries with a longer history of migration and minorization processes. The design of the questionnaire also derived from extensive empirical experiences gained in carrying out municipal home language surveys amongst pupils in both primary and secondary schools in the Netherlands (Broeder and Extra, 1995, 1998; Extra et al., 2001, 2002). For the MCP, the agreed-upon questionnaire was translated into equivalent versions in Dutch, French, German, Spanish, and Swedish. These versions were tested in at least one primary school in each partner city. On the basis of the suggestions of local educational authorities and researchers, the phrasing and wording of the questionnaires were further adapted.

The local questionnaires were printed in multiple copies. Due to the requirements of automatic processing, it was essential that printed rather than photocopied questionnaires were used. Uniformity, both in terms of content and form, was a prerequisite for data processing. Local educational authorities sent out letters of permission to schools and/or parents so that their children could participate in the survey. In each city, the printed questionnaires were distributed to school directors. Both for classroom teachers and for data collection assistants, a manual in the local language was prepared to facilitate interaction with the pupils. The completed questionnaires were delivered by the schools to the researchers at the participating universities. After checks of the total set of questionnaires per school had been made, all delivered questionnaires were sent to Tilburg University in the Netherlands for data processing.

Data processing was centrally done in Tilburg by *Babylon* researchers. Given the large size of the database, an automatic processing technique based on specially developed software (*Teleform*) and available hardware was developed and utilized. By means of these tools about 5000 forms could be scanned per day. Because some questionnaire items were filled-out in handwriting by the pupils, additional verification of these items had to be done using character recognition software; in this way, around 4000 forms could be processed per day. After scanning and verification had been completed, the database for each city was analyzed by using the SPSS program. Table 2 gives an overview of the resulting database, derived from the reports of primary school children in an age range of 4–12 years (only in The Hague were data also collected at secondary schools). The total crossnational sample consists of more than 160,000 pupils.

In order to carry out systematic analyses on the data set, a SPSS syntax file which was developed step-by-step was used in the preparation stage. In the analysis stage, another SPSS syntax file was used in order to achieve uniformity of the findings. The last stage of data processing was transmitting the outcomes of the analyses in a readable format. Given the fact that the research results should be presented in the same format in all six participating cities in the project, a crossnationally uniform format was set up. In presenting the results, *Excel Worksheets* and *Microsoft Graphics* within *Word for Windows* were used. Both the worksheets and the templates for figures within *Microsoft Graphics* were predefined. In this way, a uniform format for all the tables and figures could be achieved, which then needed to be interpreted.

**Table 2**

Overview of the MCP database.

City	Total of schools	Total of schools in the survey	Total of pupils in schools	Total of pupils in the survey	Age range of pupils
Brussels	117 <sup>a</sup>	110 <sup>a</sup>	11,500	10,300	6–12
Göteborg	170	122	36,100	21,300	6–12
Hamburg	231 public 17 catholic	218 public 14 catholic	54,900	46,000	6–11
Lyon	173 <sup>b</sup>	42 <sup>b</sup>	60,000	11,650	6–11
Madrid	708 public 411 catholic	133 public 21 catholic	202,000 99,000	30,000	5–12
The Hague	142 primary 30 secondary	109 primary 26 secondary	41,170 19,000	27,900 13,700	4–12 12–17

<sup>a</sup> Dutch-medium schools only.<sup>b</sup> Réseau d'Education Prioritaire only.

## 5. Distribution of languages across cities

The local language surveys amongst primary school children have delivered a wealth of yet unknown crossnational evidence on the distribution and vitality of IM languages at home. Apart from selecting one or more of the prespecified languages in each of the local surveys, pupils could also opt for self-references to other home languages by filling-out in hand-writing the boxes provided for this objective. The resulting database consists of a huge variety of self-references (types) and their frequencies of mentioning (tokens). In most cases, the pupils referred to entities that could be (re)traced as existing languages. In this context, the regularly updated database of *The Ethnologue* ([www.sil.org/ethnologue](http://www.sil.org/ethnologue); Grimes, 1996) on languages of the world proved to be very helpful. In cases of doubt or lacking information, other resources were used, such as Comrie et al. (2003), Campbell (2000), Dalby (1999/2000), Giacalone Ramat and Ramat (1998), and Crystal (1997). Apart from self-references to known and unknown languages, the pupils also made references to countries that could not reasonably be traced back to languages or to other/unknown categories. In general, however, the resolution level of the language question in the survey was very high, and relatively few references could not be traced back to languages. Table 3 gives a crossnational overview of the data under consideration.

Based on the overview of types and tokens of (re)traced home languages, the distribution of these home languages was specified in a ranked order of decreasing frequency. A common phenomenon in all participating cities, so familiar in type/token studies of word frequencies, was that few languages (types) were referred to often (tokens), and that many languages (types) were referred to rarely (tokens). Therefore, the most frequently mentioned home languages represent a very high proportion of the total number of occurrences/tokens in all cities.

Apart from Madrid, late-comer amongst our focal cities in respect of immigration, the proportion of primary school children in whose homes other languages were used next to or instead of the mainstream language ranged between one third and more than a half. The total number of languages other than Swedish/German/Dutch (The Hague/Brussels)/French/Spanish ranged per city between 50 and 90. The figures were 36% of the total student population in Göteborg, 35% in Hamburg, 49% in The Hague, 82% in Brussels, 54% in Lyon, and 10% in Madrid.

The outcomes of the local surveys were aggregated in one crossnational database. On the basis of the number of references made to home languages, the top 20 of the most frequently mentioned languages in each city were identified. Forty-nine languages were in the group of the top-20 list in the six cities. Out of these 49 languages, 19 languages were represented in 3–6 cities and 30 languages in only 1–2 cities. There were also unique references in the top 20 per city; most of these languages were either languages of neighboring countries, languages of former colonies, or RM languages. For purposes of crossnational and crosslinguistic analyses, 20 of the most frequently mentioned languages in these cities were chosen.

**Table 3**

References made by pupils in terms of types and tokens.

Municipality	Reference to languages		Reference to countries		Other/unknown references	
	Types	Tokens	Types	Tokens	Types	Tokens
Göteborg	75	7598	8	40	10	20
Hamburg	90	16,639	12	229	10	92
The Hague	88	23,435	13	788	17	24
Brussels	54	12,737	9	186	7	11
Lyon	66	6106	17	130	–	–
Madrid	56	2619	x	x	x	x

x, not specified.

**Table 4**

Overview of the numbers of pupils (6–11 years) per reported language and city.

Reported languages	Göteborg	Hamburg	The Hague	Brussels	Lyon	Madrid	Coverage
English	1039	1077	950	676	426	359	6
Arabic	768	464	1391	1608	2789	662	6
Portuguese	88	360	88	77	259	202	6
Italian	51	192	92	361	255	43	6
Turkish	385	4948	2535	606	468	1	5
Spanish	328	431	288	389	353	–	5
German	148	–	156	119	91	45	5
French	118	17	185	7327	–	157	5
Chinese	184	7	180	22	37	160	4
Kurdish	468	197	273	11	36	4	4
Albanian	186	410	5	107	62	3	4
Polish	163	1729	16	33	3	100	4
Russian	70	1652	14	32	11	37	4
Berber	4	–	1334	214	145	37	4
Serbian/Croatian/Bosnian	795	460	46	29	26	6	3
Vietnamese	55	153	14	14	91	–	3
Somali	315	–	135	–	49	–	3
Urdu/Pakistani	27	238	294	32	1	3	3
Armenian	8	82	5	47	41	1	3
Romani/Sinte	51	219	6	8	3	1	2

Two criteria were used to select these 20 languages from the list of 49 languages. Each language should be represented by at least three cities, and each city should be represented in the crossnational database by at least 30 pupils in the age range of 6–11 years. Our focus on this age range was motivated by comparability considerations: This range was represented in the local databases of all participating cities (see Table 2). Romani/Sinte was included in the crossnational analyses because of its special status in our list of 20 languages as a language without territorial status. Two languages had an exceptional status: English “invaded” the local databases as a language of international prestige, and Romani/Sinte was solidly represented in Hamburg and Göteborg only. The concept of language group was based on the pupils’ answers to the question whether and, if so, which other languages were used at home instead of or next to the mainstream language. On the basis of their answer patterns, pupils may belong to more than one language group. Table 4 gives an overview of the resulting database.

As shown in Table 4, eight languages were represented in 5–6 cities, while eleven languages were represented in 3–4 cities. With respect to French, Brussels offers a special case, given the public and private status of both French and Dutch in this city (Verlot et al., 2003). There is a remarkable municipal distribution of two pairs of languages which are often in competition in their source countries, i.e., Turkish and Kurdish in Turkey, and Arabic and Berber in Northern African countries (in particular, Morocco). Only in Göteborg was Kurdish more strongly represented than Turkish, and only in The Hague were Berber and Arabic represented in balance. In our database, Kurdish hardly emerged in Brussels and Madrid. The same holds for Berber in Göteborg and Hamburg.

## 6. Specification of language profiles and language vitality

For all language groups mentioned in Table 4, pseudo-longitudinal and intergenerational profiles were specified and visually represented in graphs and tables. For each language group, three age groups and three generations were distinguished. The age groups consisted of pupils aged 6/7, 8/9, and 10/11 years old. The three generations were operationalized as follows:

- G1: pupil + father + mother born abroad;
- G2: pupil born in country of residence, father *and/or* mother born abroad;
- G3: pupil + father + mother born in country of residence.

The pseudo-longitudinal profiles consisted of age-specific information on:

- *proficiency* in the minority language in terms of language understanding, speaking, reading and writing;
- *choice* of the minority language in interaction with the mother, father, younger and older siblings, and best friends;
- *dominance* in the minority vs. mainstream language;
- *preference* for the minority vs. mainstream language.



In addition, age-specific and generation-specific information was provided on language vitality. The final aim was the construction of a language vitality index (henceforward LVI), based on the outcomes of the four dimensions presented above. Since Giles et al. (1977) introduced the concept of ethnolinguistic vitality, the focus has been on its extra-linguistic determinants rather than on the empirical operationalization of language use. Determinants have been proposed in terms of lists of factors, clustered in status factors, demographic factors, and institutional support factors, e.g., by Giles et al. (1977), or in additional factors such as cultural (dis)similarity, e.g., by Appel and Muysken (1987:32–38). The proposed lists of factors suffer from various shortcomings that cannot be solved easily:

- the lists of factors are neither exhaustive nor mutually exclusive;
- different factors contribute in different ways to (lack of) vitality and may even neutralize each other;
- some of these factors are personal characteristics (e.g., age, gender, or educational level), whereas other factors are group characteristics (e.g., group size or group spread);
- moreover, a distinction has been proposed and found between the objective status of these factors and their subjective perception by minority and/or majority groups (Bourhis et al., 1981; Van der Avoird, 2001);
- finally, no relative quantitative weight has been suggested for the proposed (clusters of) factors, which makes the establishment of a language vitality index and the verification of empirical outcomes unfeasible.

For a comprehensive overview of the origins of the concept “language vitality” and its theoretical and empirical development over time since Weinreich (1953), we refer to Achterberg (2005:23–100), in the context of a case study on Slavonic languages in Germany.

In our research project, we took a different approach by focusing on the empirical operationalization of language vitality rather than on its extra-linguistic determinants. The operationalization of language vitality was derived from the following four reported dimensions:

- language proficiency: the extent to which the minority language under consideration is *understood*;
- language choice: the extent to which this language is commonly spoken at home *with the mother*;
- language dominance: the extent to which the minority language is spoken *best*;
- language preference: the extent to which the minority language is *preferably* spoken.

The focus of the chosen dimensions was on oral skills at home and not on literacy in order to give IM languages a fair chance of emerging in societal contexts in which the acquisition of literacy is rarely promoted, whether at home or at school (see also section 8). Moreover, earlier analyses have shown that the four selected dimensions are highly correlated and lead to reliable scores (Extra et al., 2002:129). The operationalization of the first and second dimension (language proficiency and language choice) was aimed at a maximal scope for tracing language vitality. Language understanding is generally the least demanding of the four language skills involved, and the mother acts generally as the major gatekeeper for intergenerational language transmission (Clyne, 2003).

In the analyses, the four above-mentioned language dimensions were compared as proportional scores, i.e., the mean proportion of pupils per language group that indicated a positive response to the relevant questions. The LVI is, in turn, the mean value of these four proportional scores. This LVI is by definition a value-driven index, in the sense that the *chosen* dimensions with the *chosen* operationalizations are weighted *equally*. The establishment of such an index makes it feasible to carry out crosslinguistic and crossnational comparisons of large databases in which equal criteria for such comparisons are used. On the basis of the established LVI, LVI scores have been calculated per age group and per generation, for each language group. On the basis of this categorization, intergenerational shift can be estimated.

In all cases, the total population of age groups was always larger than the total population of generations. This discrepancy is the result of a predictably larger number of missing values (i.e., non-responses) for generation than for age. In the former case, references have to be made to the countries of birth of the pupil, the father, and the mother; in the latter case, reference has to be made only to the age of the pupil. Language vitality indices for age and generation were calculated only if at least 5 pupils were represented in a particular group. Given the possible non-responses of pupils to any of the questions represented in figures and tables, all figures and tables were presented and interpreted in proportional values. In Table 5, we demonstrate the provided crossnational, pseudo-longitudinal and intergenerational information per language group for the Turkish language group (see Table 4 for the absence of data on Madrid).

## 7. Crosslinguistic perspectives on language vitality

Table 6 gives a crosslinguistic and pseudo-longitudinal overview of the LVI per language group and age group. LVI calculations have only been made if at least 5 pupils were represented in a particular age group and generation. Considering its non-territorial status, it is not surprising that Romani/Sinte emerged with the highest language vitality. English and German ended up in bottom positions given the fact that they often had a higher status at school than at home. When the average scores of the youngest and oldest age groups were compared, 11 language groups showed the highest scores for the former and 5 language groups for the latter. The largest interval between the scores emerged for Romani/Sinte. Strong

**Table 5**

Turkish language group: crossnational numbers of pupils and LVI per age group and per generation.

Cities	Age groups							
	Population				Vitality			
	6/7	8/9	10/11	Total	6/7	8/9	10/11	Mean
Göteborg	124	115	146	385	69	67	66	67
Hamburg	1384	2381	1183	4948	66	62	65	64
The Hague	833	853	849	2535	75	68	65	69
Brussels	225	213	168	606	73	75	71	73
Lyon	146	176	146	468	65	63	68	65
Total/mean	2712	3738	2492	8942	70	67	67	68

Cities	Generations							
	Population				Vitality			
	G1	G2	G3	Total	G1	G2	G3	Mean
Göteborg	51	308	10	369	67	68	43	59
Hamburg	627	3676	205	4508	69	64	49	61
The Hague	539	1842	46	2427	73	68	62	68
Brussels	75	417	42	534	74	74	70	73
Lyon	78	308	24	410	70	64	65	66
Total/mean	1370	6551	327	8248	71	68	58	65

maintenance of language vitality across the youngest and oldest age groups, with intervals of  $-1/0/+1$  only, emerged for 8 out of the 20 language groups.

A different crosslinguistic and pseudo-longitudinal perspective is provided in Table 7, in terms of generations. Table 7 reveals significant differences between language groups in the distribution of pupils across different generations. In most language groups, second-generation pupils were most-represented and third-generation pupils least. Remarkable exceptions to this rule were Armenian and in particular Russian, with mainly first-generation pupils. Third-generation pupils were relatively well represented ( $>20\%$ ) for English, French, German, Italian, Romani/Sinte, and Spanish. In conformity with expectations, Table 7 shows a stronger decrease of language vitality across generations than Table 6 shows across age groups. All language groups show more or less decreasing language vitality across generations. The strongest intergenerational shift between G1 and G3 emerged for Polish (42%), Albanian (38%), Spanish (33%), and Portuguese (30%), whereas the strongest intergenerational maintenance of language vitality occurred for Romani/Sinte and Turkish.

**Table 6**

Language vitality per language group and age group (in %, LVI in cumulative %).

Language group	Total pupils	6/7 years	8/9 years	10/11 years	Average
Romani/Sinte	270	76	71	64	70
Urdu/Pakistani	564	65	70	69	68
Turkish	8942	70	67	67	68
Armenian	170	64	59	65	63
Russian	1791	66	58	57	60
Serbian/Croatian/Bosnian	1285	60	58	59	59
Albanian	765	63	56	58	59
Vietnamese	299	57	60	58	58
Chinese	561	56	58	60	58
Arabic	7682	59	58	58	58
Polish	1925	57	59	53	56
Somali	499	58	54	53	55
Portuguese	1074	54	54	54	54
Berber	1730	51	54	51	52
Kurdish	974	54	47	51	51
Spanish	1789	47	49	47	48
French	7787	47	40	44	44
Italian	994	39	40	39	39
English	4527	37	33	39	36
German	559	35	31	32	33

**Table 7**

Intergenerational distribution (in %) and intergenerational language vitality (LVI in cumulative %) per language group.

Language group	Total pupils	Intergenerational distribution			Intergenerational language vitality		
		G1	G2	G3	G1	G2	G3
Albanian	675	39	56	5	72	51	34
Arabic	7002	21	73	6	64	57	35
Armenian	153	49	42	9	69	55	–
Berber	1656	20	78	2	59	50	45
Chinese	523	22	74	4	72	59	–
English	4045	16	42	41	43	41	28
French	7090	7	45	48	55	43	30
German	506	18	45	38	43	35	22
Italian	916	12	60	28	49	43	29
Kurdish	900	50	49	2	61	43	33
Polish	1837	14	82	4	73	59	31
Portuguese	1004	27	66	8	63	52	33
Romani/Sinte	231	35	41	23	76	66	65
Russian	1616	81	16	3	64	–	–
Serbian/Croatian/Bosnian	1191	38	58	4	71	50	–
Somali	464	38	58	5	70	50	–
Spanish	1570	18	61	21	63	47	30
Turkish	8248	17	79	4	71	68	58
Urdu/Pakistani	534	25	72	3	70	67	–
Vietnamese	270	12	85	3	60	57	–

The top position for language vitality of Romani/Sinte across age groups in Table 6, and its relatively strong maintenance across generations in Table 7, were also observed in earlier and similar research in the Netherlands (Broeder and Extra, 1998 :70). The high vitality of Romani/Sinte was also confirmed by other studies on this language community (Acton and Mundy, 1999; Kyuchukov, 2002). One reason why language vitality is a core value for the Roma across Europe is the absence of source country references as alternative markers of identity—in contrast to almost all other language groups presented in Tables 6 and 7.

## 8. Conclusions

The findings of the *Multilingual Cities Project* have delivered a wealth of hidden evidence on the distribution and vitality of IM languages at home across European cities and nation-states. Apart from Madrid, late-comer amongst our focal cities in respect of immigration, the proportion of pupils in whose homes other languages were used next to or instead of the mainstream language ranged per city between one third and more than a half. The total number of traced “other” languages ranged per city between 50 and 90; the common pattern was that a limited set of languages were often referred to by the pupils and that many languages were referred to only a few times.

The findings show that making use of more than one language is a way of life for an increasing number of children across Europe. Mainstream and non-mainstream languages should not be conceived in terms of competition. Rather, the data show that these languages are used as alternatives, depending on such factors as type of context and interlocutor. The data also reveal that the use of other languages at home does not occur at the cost of competence in the mainstream language. Many children who addressed their parents in another language reported to be dominant in the mainstream language.

Among the major 20 non-national languages in the participating cities, 10 languages are of European origin and 10 languages stem from abroad. These findings clearly show that the traditional concept of language diversity in Europe should be reconsidered and extended. The outcomes of the local language surveys also demonstrate the high status of English among primary school pupils across Europe. Its intrusion in the children's homes is apparent from the position of English in the top 5 of languages referred to by the children in all of the cities (Table 4). This outcome cannot be explained as an effect of migration and minorization only. The children's reference to English also derives from the status of English as the international language of power and prestige. English has invaded the repertoire of all of the national languages under consideration. Moreover, children have access to English through a variety of media, and English is commonly taught in particular grades at primary schools.

In addition, children in all participating cities expressed a desire to learn a variety of languages that are not taught at school. The results of the local language surveys also show that children who took part in instruction in particular non-mainstream languages at school reported higher levels of literacy in these languages than children who did not take part in such instruction. Both the reported reading proficiency and the reported writing proficiency profited strongly from language instruction. The differences between participants and non-participants in language instruction were significant for both

forms of literacy skills and for all the 20 language groups under consideration. In this domain in particular, the added value of language instruction for language maintenance and development is clear. Owing to the monolingual *habitus* (Gogolin, 1994) of primary schooling across Europe, there is an increasing mismatch between language practices at home and at school. The findings on multilingualism at home and those on language needs and language instruction reported by the children should be taken into account by both national and local educational authorities in any type of language policy.

## References

- Achterberg, Jörn, 2005. Zur Vitalität Slavischer Idiome in Deutschland. Eine Empirische Studie zum Sprachverhalten Slovophoner Immigranten. Verlag Otto Sagner, München.
- Acton, Thomas, Mundy, Gary (Eds.), 1999. *Romani Culture and Gypsy Identity*. University of Hertfordshire Press, Hatfield.
- Akinci, Mehmet-Ali, de Ruyter, Jan Jaap, Sanagustin, Floréal, 2004. *Le Plurilinguisme à Lyon. Le Statut des Langues à la Maison et à l'École*. L'Harmattan, Paris.
- Alladina, Safder, 1993. South Asian languages in Britain. In: Extra, G., Verhoeven, L. (Eds.), *Immigrant Languages in Europe*. Multilingual Matters, Clevedon, pp. 55–65.
- Alladina, Safder, Edwards, Viv (Eds.), 1991. *Multilingualism in the British Isles (Vol. 1: The Older Mother Tongues and Europe; Vol. 2: Africa, the Middle East and Asia)*. Longman, London.
- Alterman, Hyman, 1969. *Counting People. The Census in History*. Harcourt, Brace & World, New York.
- Ammerlaan, Tom, Hulsen, Madeleine, Strating, Heleen, Yağmur, Kutlay (Eds.), 2001. *Sociolinguistic and Psycholinguistic Perspectives on Maintenance and Loss of Minority Languages*. Waxmann, Münster/New York.
- Ammon, Ulrich, Mattheier, Klaus, Welde, Peter (Eds.), 1995. *Europasche Identität und Sprachenvielfalt*. Sociolinguistica 9. Max Niemeyer, Tübingen.
- Ansell, Amy, Solomos, John, 2008. *Race and Ethnicity: Key Concepts*. Routledge, London.
- Appel, René, Muysken, Pieter, 1987. *Language Contact and Bilingualism*. Edward Arnold, London.
- Barbour, Stephen, 2000. Nationalism, language, Europe. In: Barbour, S., Carmichael, C. (Eds.), *Language and Nationalism in Europe*. Oxford University Press, Oxford, pp. 1–17.
- Barni, Monica, Extra, Guus (Eds.), 2008. *Mapping Linguistic Diversity in Multicultural Contexts*. Mouton de Gruyter, Berlin/New York.
- Blum, Alain, 2002. Resistance to identity categorization in France. In: Kertzer, D., Arel, D. (Eds.), *Census and Identity. The Politics of Race, Ethnicity, and Language in National Censuses*. Cambridge University Press, Cambridge, pp. 121–147.
- Bourhis, Richard, Giles, Howard, Rosenthal, Doreen, 1981. Notes on the construction of a 'Subjective Vitality Questionnaire' for ethnolinguistic groups. *Journal of Multilingual and Multicultural Development* 2, 145–155.
- Broeder, Peter, Extra, Guus, 1995. *Minderheidsgroepen en Minderheidstalen*. VNG Uitgeverij, Den Haag.
- Broeder, Peter, Extra, Guus, 1998. *Language, Ethnicity and Education. Case Studies on Immigrant Minority Groups and Immigrant Minority Languages*. Multilingual Matters, Clevedon.
- Broeder, Peter, Mijares, Laura, 2003. *Plurilingüismo en Madrid. Las Lenguas de los Alumnos de Origin Inmigrante en Primaria*. Centro de Investigación y Documentación Educativa, Madrid.
- Campbell, George, 2000. *Compendium of the World's Languages, vols. 1 and 2*. Routledge, London.
- Clyne, Michael, 1991. *Community Languages: The Australian Experience*. Cambridge University Press, Cambridge.
- Clyne, Michael, 2003. *Dynamics of Language Contact*. Cambridge University Press, Cambridge.
- Comrie, Bernard, Matthews, Stephen, Polinsky, Maria, 2003. *The Atlas of Languages. Revised Edition*. Eurospan Group, London.
- Crystal, David, 1997. *The Cambridge Encyclopedia of Language*. Cambridge University Press, Cambridge.
- Dalby, David, 1999/2000. *The Linguasphere Register of the World's Languages and Speech Communities, vols. 1 and 2*, Linguasphere Press, Wales-Hebron.
- Extra, Guus, Gorter, Durk (Eds.), 2001. *The Other Languages of Europe. Demographic, Sociolinguistic and Educational Perspectives*. Multilingual Matters, Clevedon.
- Extra, Guus, Gorter, Durk (Eds.), 2008. *Multilingual Europe: Facts and Policies*. Mouton de Gruyter, Berlin/New York.
- Extra, Guus, Yağmur, Kutlay (Eds.), 2004. *Urban Multilingualism in Europe: Immigrant Minority Languages at Home and School*. Multilingual Matters, Clevedon.
- Extra, Guus, Aarts, Rian, Van der Avoird, Tim, Broeder, Peter, Yağmur, Kutlay, 2001. *Meertaligheid in Den Haag. De Status van Allochtone Talen Thuis en op School*. European Cultural Foundation, Amsterdam.
- Extra, Guus, Aarts, Rian, Van der Avoird, Tim, Broeder, Peter, Yağmur, Kutlay, 2002. *De Andere Talen van Nederland. Thuis en op School*. Coutinho, Muiderberg.
- Fase, Willem, Jaspaert, Koen, Kroon, Sjaak (Eds.), 1992. *Maintenance and Loss of Minority Languages*. John Benjamins, Amsterdam/Philadelphia.
- Fase, Willem, Jaspaert, Koen, Kroon, Sjaak (Eds.), 1995. *The State of Minority Languages. International Perspectives on Survival and Decline*. Swets & Zeitlinger, Lisse/Exton.
- Fürstenau, Sarah, Gogolin, Ingrid, Yağmur, Kutlay (Eds.), 2003. *Mehrsprachigkeit in Hamburg. Ergebnisse einer Sprachenerhebung an den Grundschulen in Hamburg*. Waxmann, Münster/New York.
- Giacalone Ramat, Anna, Ramat, Paolo, 1998. *The Indo-European Languages*. Routledge, London.
- Giles, Howard, Bourhis, Richard, Taylor, Donald, 1977. Towards a theory of language in ethnic group relations. In: Giles, H. (Ed.), *Language, Ethnicity, and Intergroup Relations*. Academic Press, London, pp. 307–348.
- Gogolin, Ingrid, 1994. *Der Monolinguale Habitus der Multilingualen Schule*. Waxmann, Münster/New York.
- Gogolin, Ingrid, Kroon, Sjaak, Krüger-Potratz, Marianne, Neumann, Ursula, Vallen, Ton (Eds.), 1991. *Kultur- und Sprachenvielfalt in Europa*. Waxmann, Münster/New York.
- Grimes, Barbara (Ed.), 1996. *Ethnologue. Languages of the World, 13th ed.* Summer Institute of Linguistics, Dallas ([www.sil.org/ethnologue](http://www.sil.org/ethnologue)).
- Haarmann, Harald, 1995. *Europasche Identität und Sprachenvielfalt*. Max Niemeyer, Tübingen.
- Kertzer, David, Arel, Dominique, 2002. *Census and Identity. The Politics of Race, Ethnicity, and Language in National Censuses*. Cambridge University Press, Cambridge.
- Kyuchukov, Hristo (Ed.), 2002. *New Aspects of Roma Children Education*. Ictus, Sofia.
- Nicholas, Joe, 1988. British language diversity surveys (1977–1987). A critical examination. *Language and Education* 2, 15–33.
- Nygren-Junkin, Lilian, Extra, Guus, 2003. *Multilingualism in Göteborg. The Status of Immigrant Minority Languages at Home and at School*. European Cultural Foundation, Amsterdam.
- Poulain, Michel, 2008. European migration statistics: definitions, data and challenges. In: Barni, M., Extra, G. (Eds.), *Mapping Linguistic Diversity in Multicultural Contexts*. Mouton de Gruyter, Berlin/New York, pp. 43–66.
- Spencer, Stephen, 2006. *Race and Ethnicity. Culture, Identity and Representation*. Routledge, London.
- Van der Avoird, Tim, 2001. *Determining language vitality. The language use of Hindu communities in the Netherlands and the United Kingdom (Dissertation)*. Tilburg University, Tilburg.
- Verlot, Marc, Delrue, Kaat, Extra, Guus, Yağmur, Kutlay, 2003. *Meertaligheid in Brussel. De Status van Allochtone Talen Thuis en op School*. European Cultural Foundation, Amsterdam.
- Weinreich, Uriel, 1953. *Languages in Contact. Findings and Problems*. Mouton, The Hague.

**Prof. Guus Extra** holds the Chair of *Language and Minorities* in the Department of Culture Studies of the Faculty of Humanities at Tilburg University. In 1999, he was appointed Director of *Babylon, Centre for Studies of the Multicultural Society* ([www.tilburguniversity.nl/babylon](http://www.tilburguniversity.nl/babylon)). Guus Extra has an extensive record in research and publications on second language acquisition, first language use/maintenance/loss by immigrant minority groups, and multilingualism and education, both at the national (Dutch) and European level. He has also an extensive record in coordinating research in these domains, both at the national and European level.

**Kutlay Yağmur** is an Associate Professor of Multilingualism in the Department of Culture Studies, Tilburg University, The Netherlands. In his current research, he investigates the relationship between the integration ideology of receiving societies and the acculturation processes of immigrants in four national contexts. He has published extensively on Turkish immigrants in Australia, Germany, France and The Netherlands. Apart from his contributions to a variety of international journals, his co-edited book *Urban Multilingualism in Europe. Immigrant Minority Languages at Home and School* (with Guus Extra, *Multilingual Matters*, 2004) has been a major conceptual and methodological contribution to the study of multilingualism.