AUTHOR'S QUERY SHEET

Author(s): L.N. Veselinova and J.C. Booza RMMM 358415

Article title:

Article no:

Dear Author

The following queries have arisen during the editing of your manuscript and are identified on the proofs. Unless advised otherwise, please submit all corrections using the CATS online correction form.

- AQ1 Please spell out [GIS] in full at the first mention.
- AQ2 Please check the mention of Section [0] under section 'Linguistic diversity'.
- AQ3 I have changed the reordering of Map from [6] to [5] and [7] to [6]. Please can you confirm that this is OK?
- AQ4 Please provide caption for Figure [1].
- AQ5 Please provide complete details.
- AQ6 Please provide page numbers.
- AQ7 Please provide details of the publisher.

Studying the Multilingual City: a GIS-based approach

Ljuba Nikolova Veselinova^{a*} and J.C. Booza^b

^a*Department of Linguistics, Stockholm University, Stockholm, Sweden;* ^b*Department of Geography and Urban Planning, Wayne State University, Detroit, USA*

(Received 11 December 2008; final version received 12 October 2008)

Work in two distinct disciplines, urban geography and sociolinguistics, readily points out the multiethnic and multilingual character of metropolitan areas. However, there is still demand for studies which establish the language structure of modern cities. For the purposes of this pilot study, we focus on the Detroit Metropolitan Area (DMA), Michigan. GIS technology together with census data were used to arrive at an adequate description of the spatial distribution of languages currently spoken in Detroit and its immediate surroundings. Data from the 2000 US Census are entered into a GIS system and presented visually with a subsequent analysis of the emerging spatial patterns. Due to limitations of the data, we had to restrict the mapping to languages used at home. The study suggests one possible model for the initial stages of mapping the multilingual city; moreover, the data analysed here provide the infra-structure necessary for further research on phenomena such as language shift and language death as well as other aspects of a dynamic multilingual situation.

Keywords: language mapping; geolinguistics; mapping multilingualism; sociolinguistics; urban geography

Introduction

The multiethnic and multilingual character of metropolitan areas is commonly acknowledged by urban geographer and sociolinguists alike. However, there is still demand for studies which establish the language structure of modern cities. This long neglected subject deserves attention for a number of reasons. One pertains to descriptive language geography. Specifically, except for a small number of works, as for instance (Grobler, Prinsloo, and Van der Merwe 1990, 47, in passim), and also (Aitchinson and Carter 1985, 12, in passim), the multilingual composition of urban areas is generally ignored in most printed language atlases (cf. (Breton 1997; Heine and Möhlig 1980; Moseley and Asher 1994), just to name a few). However, half of the world's population lives in cities, so ignoring their multilingual composition gives, at best, a skewed picture of the spatial distribution of the languages around the globe. Additionally, as Van der Merwe (1993, 411) points out, cities are most likely to be the demographic locus of the future. A more detailed knowledge about their ethno-linguistic composition helps city planning and also advances the study of subjects such as language contact, multilingualism and language distribution according to functional domains, such as language choice in different social settings e.g. home

Corresponding author. Email: ljuba@ling.su.se

versus work, leisure, religious practice. Finally, in the sociolinguistic literature, cities have been mainly presented as places where dialects converge into a more prestigious standard and where class and occupational dialects (sociolects) emerge. However, Van der Merwe (1993, 411) following (Angle 1981) and (Mikesell and Murphy 1991) highlights the hypothesis that cities may be the best places for preservation of minority languages as segregated neighbourhoods offer the possibility of close contact (Van der Merwe 1993, 412–3). Currently, there are around 7000 language spoken around the globe and most of them are in danger of extinction (cf. <http://www.ethnologue.com>). Thus, studying the city is potentially also important for language preservation as well.

So far the research on languages in cities has been focussed on the following topics:

1. Current and changing demographics of large urban areas, cf. (Bains 2005; Barni 2006; Pamuk 2006; Van der Merwe 1993; Williams 2000), see also (Higgs, Williams, and Dorling 2004). For the most part, these studies are based on census data on ethno-linguistic groups combined with some use of GIS technology. Some of the cities studied in such a fashion are Cape Town, London, San Francisco, New York, Boston, Montreal, a neighbourhood in Rome (cf. Barni 2006).
2. Language choice in various domains. This topic is the subject of the studies presented in Extra and Yağmur (2004). The data used in this work have been collected by means of direct interviews with speakers. The cities studied this way so far are Gothenburg, Hamburg, The Hague, Lyon Brussels and Madrid.
3. The *linguistic landscape*. This term is used to describe the (study of) language of street signs. The data come from taking photographs and subsequently analysing street signs (cf. Gorter 2006). In an earlier work, Landry and Bourhis (Landry and Bourhis 1997) combine the study of the linguistic landscape with a study of language attitudes and language choice in various domains. The research works mentioned above cover cities such as Montreal, Bangkok, Tokyo, Ljouwert and Leeuwarden in Friesland and Donostia and San Sebastian in Basque Country.

A holistic study of the language situation in large urban areas should make use of all three directions and the methodologies they imply. Additionally, as suggested by Williams and Van der Merwe (Williams and Van der Merwe 1996), we should make better use of GIS, a technology that is becoming pervasive and easily accessible to the general user. The main advantage of including GIS in research on multilingual urban areas is the flexibility GIS tools offer. As is well known, GIS provide the possibility to assemble and visualise different kinds of data in separate layers. This way, we can, potentially, draw a variety of correlations between linguistic and non-linguistic variables – something that sociolinguists have been trying to do but have usually been rather constrained due to the use of technologically less advanced means. Large-scale studies presuppose, however, substantive project funding and work in teams that consist of linguists, geographers and statisticians. The current work is performed on a much more modest scale but we hope to make a contribution to the field of urban geolinguistics and to further promote the awareness of the need for more studies of this kind.

Following Williams and Van der Merwe (1996), this paper presents a pilot study where census data from the US census year 2000 are used together with GIS technology for studying the multilingual composition of the Detroit Metropolitan Area (DMA).¹ Apart from the descriptive goal of the paper, we use the collected data to test and further discuss the language vitality model suggested by Giles, Bourhis, and Taylor (1977).

Choosing the DMA was largely a matter of convenience as both authors resided in the area at the time the study was conducted. However, one of main points of this paper, in addition to the ones stated above, is that any densely populated urban area deserves to be studied in its own right. In the sociolinguistic literature, Detroit is mainly known from studies on African American Vernacular English (cf. (Green 2002; Smitherman 1977; Wolfram 1969) for a few examples), and not as an area of great ethnic diversity. A careful examination of its overall population brings up multiethnic aspects that have received only local attention, at best. Thus, even an area known as a fairly homogeneous one shows in fact a great deal of diversity. When mapped, the diversity of this area shows spatial patterns are that are in some ways similar to diversity patterns in other cities, and in others specific to this area.

This paper is organised as follows. Section 'Terminology and Method' gives an outline of our terminology and method. In Section 'The Language Composition of the Detroit Metropolitan Area (DMA)', we present the languages of the DMA as identified by the census, together with some basic GIS analysis of the kinds of settlement patterns observed for different ethnic groups. Relating these settlement patterns with economic factors as well as with the spatial distribution of foreign language instruction (FLI) is the subject matter of Section 'Using GIS to Relate Ethno-linguistic Distribution with Other Variables'. Summary and conclusions are presented in the last section.

Terminology and method

Terminology

In the presentation below we use the following terms:

Language

For the purposes of this work, we consider a language the entity defined as such by linguistic and socio-political criteria and typically listed by at least one language classification. Here we use the Ethnologue (<http://www.ethnologue.com>). Further divisions of languages into dialects are not taken into account.

Language composition

The list/enumeration of languages spoken in a given area. We use the term *language structure* with the same content.

Language vitality or ethno-linguistic vitality

A notion that originates from research on psychology of language. A language of 'high' vitality is spoken and used widely. This notion is highly relevant in discussions

of language shift or language death (cf. Meyerhoff 2006, 107–8, in passim, also Giles, Bourhis, and Taylor 1977, 308–18).

Linguistic diversity

AQ2

For this notion we follow Greenberg (1956) (see also Section 0 below). That is the term linguistic diversity is used to refer to the quantitative measure (index) of the probability for any two individuals from a given area to speak the same or different languages. The term linguistic diversity as used here does not presuppose geographical concepts such as index of segregation/intergration; index of dissimilarity, isolation, exposure or interaction measures.

Density

The number of items per unit area, such as persons per square kilometre. Also known as population density.

Distribution

Spatial pattern of where an observed feature appears within a particular area.

Segregation

The location of particular groups of people into distinct areas separate from the general population, usually based on race, religion or economic circumstances.

Data collection, procedures and software used

Data on language use in the DMA comes from the US Census 2000, Detailed Language Spoken at Home for the Population 5 Years and Over (STP224) (http://www.census.gov/population/www/socdemo/lang_use.html).

For the purposes of applying the language vitality model, we analysed Census 2000 data on property values, home ownership and median rent.² We also collected data on FLI in the schools of the area and finally its linguistic landscape.

Data on FLI are not readily available. Despite repeated efforts the Michigan Board of Education did not respond to our requests to provide a list of schools with foreign languages on their curricula. We had better luck with the Detroit school district, thanks to Mattie Majors, Detroit Public Schools Media Relations Director, who provided us with a list of schools with FLI in this district. For the rest of the area, we used data from the online database of the National Center for Education Statistics, <http://nces.ed.gov/> which we searched for schools in the three counties under investigation: Oakland, Macomb and Wayne. This search produced a list of 1168 schools. Only 229 of the 1168 schools had information on FLI available online. The state of FLI and its spatial distribution based on our data were presented in Section ‘Ethno-linguistic Distribution and Foreign Language Instruction (FLI)’.

The collected data were into a GIS/ArcMap for visualisation and analysis. The map files were then published online via Google Earth (<http://ling-map.ling.su.se/website/Detroit.kmz>). An ArcIMS map service for alternative display of these data is

currently under development, the addresses where the service will appear in winter 2009 is <http://ling-map.ling.su.se/website/detroit/>.

The language composition of the Detroit Metropolitan Area (DMA)

Using census data for linguistic research

Census data on the languages used in a given area are in many ways an indispensable source of information but working with them at face value presents problems in several respects. One concerns the kind of information collected by the census, and subsequently its handling and distribution; here we will only discuss these issues with regard to the US census but they are problems that deserve attention on a global scale.

First, it is questionable whether the data collected by the census are completely reliable. As Pienemann and Keßler point out (2007, 252) respondents may answer in a way that reflects their desired status in society (which often includes language use) rather than the actual situation.

Second, this study had to be restricted to languages used in households since the US census only asks respondents about their primary language use in the home.³ Survey respondents are asked whether they speak a language other than English at home and if so, what language they speak. Furthermore, respondents are asked to indicate how well they speak English if they do speak another language. The latter question can be used to gauge English proficiency in the household, but the US Census tells us nothing about the degree of proficiency for languages other than English. Furthermore, the data and subsequently any maps based on them, do not give any information about language use in contexts outside of the home environment or about attitude(s) towards using the identified languages in various situations. These issues are central to any sociolinguistic inquiry on multilingualism but they cannot be investigated properly using the US census data.⁴

Third, from the point of view of the language geographer, language names are recorded by the census in a way that is often completely unsuitable for linguistic analysis. For instance, Spanish and Spanish Creole are put in the same group; however, they are entirely different languages and consequently should be given separate records. The same can be said about Portuguese and Portuguese Creole. The census uses the term *Chinese* which is too broad a notion to be truly informative. The term *Syriac* is somewhat unfortunate since in linguistics it is more commonly used to refer to a dead Aramaic language from Iraq. Obviously, in the census context the term *Syriac* refers to a Modern Aramaic language from Iraq but if a potential user should look up this language (which is otherwise not very well known) in a linguistically informed language catalogue, such as, for instance, *The Ethnologue* (<http://www.ethnologue.com>), the user will be quite puzzled. The more commonly used name for this language is *Assyrian* and it is used in the presentation below. In addition to erroneous reference to individual languages, many languages are grouped together according to some loose geographical or genealogical principles. For example, there is a group called *African languages*. Such a grouping is completely unsatisfactory from a linguistic point of view since Africa is a huge continent, with a great linguistic diversity, thus, grouping languages together just because they stem from there is not justified. On a general level, licensing such labels on an authoritative

site such as the US Census reinforces their validity with the general public, which at best, helps for spreading misinformation.

Since these problems pose very serious limits to any linguistic inquiry, many researchers simply discard the census as a data source for language studies. In our case, we decided to use it nonetheless, while making up for its shortcomings to the extent possible.⁵

Mapping the census data

We identified 88 different languages spoken in the households of Oakland, Macomb and Wayne counties. The dominant language is not surprisingly, English (3,322,071, 88.4% of the population), and the two big minority languages are Spanish (92,626, 2.5%) and Arabic (65, 1.7%), cf. Appendices 1 and 2 for an alphabetical list of all languages as well as their placement on the map layers of the Google application. It is worth noting that Arabic in this area shows the highest concentration of speakers in the whole of the USA. For most languages, the location of their speakers is specified on tract level. For languages with a relatively small number of speakers, the confidentiality procedures of the census have been applied to the data and consequently, information is available on county level only.⁶

Linking the census data to a GIS system enables us to visualise the following: (i) the spatial distribution of each ethno-linguistic group individually; (ii) different kinds of settlement patterns; (iii) clusters of unrelated languages (Map 1).

The greatest concentration of Spanish speakers is at the west end of Detroit City whereas the greatest concentration of Arabic speakers is in Dearborn, right outside Detroit as well as in Hamtramck, a separate city in the middle of Detroit territory.

Map 1. The distribution of monolingual English households together with the core areas of Spanish and Arabic.

As also shown on the map there are areas where none of these three languages is the dominant one. We will return to these areas in Section 'Measuring Linguistic Diversity' when we discuss the measurement of linguistic diversity.

There are languages which form clear density areas whereas others hardly form any density areas at all. For instance, consider the settlement of Polish and Assyrian speakers shown on Map 2. The two languages are relatively close in terms of number of speakers in the area: around 30,000 for Polish and 25,000 for Assyrian.

Polish can be easily described as a 'low density' language: its speakers have settled all over the three counties, with a clear dispreference for Detroit proper. The high-density areas of Polish speakers are, on the whole, very few. Assyrian speakers, on the other hand, form two clearly delimited areas in the counties of Oakland and Macomb. Moreover, the number of speakers in a dense Assyrian tract is much higher than the number of speakers in a dense Polish tract.

This clear differentiation in settling patterns can be correlated with recency of immigration. Polish immigration to the area dates back to at least the 1930s; the Assyrian population (primarily Iraqi Chaldean), while it also has historical roots, is a relatively recent phenomenon, since late 1980s and onwards. A very similar observation can be made when comparing the spatial distribution of Chinese and Albanian. Again, immigration from China has been going on for a long time and Chinese speaking households are relatively spread out in the area. Albanian speakers started to arrive to the ears in the 1990s; they are still very much concentrated to the Macomb county.

Another motivation for different settling patterns is cited by Ratliff (Ratliff 2005, 258), namely, reasons for immigration for the different groups. Ratliff observes that people who come to the US out of life-threatening necessity, e.g. refugees or asylum

Map 2. Polish and Assyrian in Detroit Metro.

seekers, tend to settle down close together. Immigrants with other motives, for instance, economic immigration, tend to integrate on a faster pace and generally do not commonly form closely knit communities. Ratliff cites the examples of Hmong Miao most of who came to the USA as refugees and still tend to live close together.

Additionally, shared religion appears to be another reason for the clustering of unrelated linguistic groups. Consider the distribution of Arabic, Urdu and Bengali speakers, shown on Map 3.

As shown on the map, the highest concentration of Arabic speakers is in Dearborn followed by Hamtramck. When Urdu speaking households are compared to that of Arabic speakers, a spatial relationship is visible. The same observation can be made for most Bengali speakers who are ethnically and linguistically completely separate from both Arabic and Urdu speakers. Yet, being Muslim appears to be an important factor in their spatial patterns. In fact, most of the Bengali (3530 out of 5215, that is 67%) appear to live in Hamtramck where a number of Arabic and Urdu speakers live as well. A major reason for the spatial clustering of linguistic groups based on religious affiliation is the few number of mosques in the Detroit area. Most mosques are located either in the Dearborn and Hamtramck areas. Other important factors in explaining the spatial clustering of these groups include socio-economic, family and employment similarities. Furthermore, Dearborn and Hamtramck serve as immigrant gateways for lower socio-economic immigrant groups to the DMA.

Measuring linguistic diversity

A quantitative ranking of languages where only number of speakers is taken into account does not present a realistic estimate of the actual linguistic diversity.

Map 3. Clustering of different languages: Arabic, Urdu and Bengali.

When estimating diversity, we need to know the following:

1. The number of languages spoken in a given area.
2. The proportion their speakers represent from the total population (both for each individual language and for all languages together).

Neither (1) nor (2) alone are sufficient indicators of diversity. For instance, stating that there are 88 languages spoken in Detroit and vicinity tells us nothing about the proportional distribution of these languages among the population. Conversely, stating that 12% of the population speaks a language other than English at home, hides the number of languages actually spoken in the area. Similarly, looking at each language and its proportions individually, for instance, stating that Arabic is spoken at home by 1.7% of the population in Detroit Metro does not say much about the overall linguistic diversity of the region.

In his seminal article from 1956 Joseph Greenberg discusses several methods for measuring linguistic diversity. In the simplest one, referred to as the monolingual non-weighted method, Greenberg proposes a way to calculate the probability for any two individuals from one and the same area to speak the same or different languages. This method can be also used to calculate an index that shows the diversity of the area in an objective way. The procedure suggested by this author can be outlined as follows:

1. Identify all the languages spoken in an area, say, M, N, O and the number of their speakers.
2. Calculate the proportion of the speakers of languages M, N, O from the total population. The calculated proportions are referred to as m , n , o .
3. Use the values obtained in (1) and (2) to calculate a diversity index A .

$$A = 1 - \sum (m^2 + n^2 + o^2)$$

A value of 1 obtained using this index indicates the greatest amount of diversity with each language group representing an equal proportion of the total languages spoken in the area. A value of 0 represents no diversity; this value would be obtained in case only one language was spoken in the area. This index allows us to not only make observations about linguistic diversity in Detroit but also allows others to compare diversity to other areas of the USA and the world. An example of its application is shown below.

As the census data describes language use on tract level, the formula above was applied to each tract by entering the data in a spreadsheet and thus, an index of diversity was calculated for each tract of the area. The index values obtained this way were then used to produce the diversity map presented below (Map 4).

The diversity index values of individual tracts were grouped into four categories into order to make visual representation possible. Those are also indicated on the map legend above: low diversity areas have diversity index 0 through 0.13; mid-low 0.14–0.25, mid-high 0.26–0.43, high 0.44–0.81. As we can see the city of Detroit shows low diversity values whereas high and mid-high diversity areas are form a belt-like shape around the city; the encircled city of Hamtramck show a very high diversity values as well. The identified areas of high, mid-high, mid-low and low diversity can be further related with other variables. Examples of such correlations

Map 4. Applying Greenberg's measurement of linguistic diversity.

are presented in Section 'Using GIS to Relate the Ethno-linguistic Distribution with Other Variables' below.

Using GIS to relate the ethno-linguistic distribution with other variables

Spatial distribution of ethno-linguistic groups and its relation to economic factors

The areas of varying ethno-linguistic diversity were checked for spatial relations with economic factors such as property values, median rent and household home owner occupancy. The clearest relations are between particular diversity areas and areas associated with specific property values and median rent. Somewhat less clear is a relation between household owner occupancy and diversity areas. The distribution of property values and median rent is presented on Map 5; the diversity areas according to Greenberg together with the distribution of household owner occupancy are presented on Map 6.

As we can see the approximate distribution of mid to high property values and mid to high media rents lie north of the Detroit City. This distribution is very similar to the high and mid-high language diversity areas. Conversely, areas of low diversity appear to pattern with areas of low property values and low median rent.⁷ One is almost tempted to state that if one speaks many languages, one is likely to live in a more expensive household. However, probably what this pattern points to is the sad reality of monolingual African American population who live in the poorest districts of the metropolitan area.

Conversely, the groups concentrated in high diversity areas are: Assyrian (56% of the whole group), Arabic (54%), Telugu (48%), Serbo-Croatian (35%) and Albanian (34%). Most Chinese speaking households are in mid-high (39%) and mid-low

Map 5. Distribution of property values and median rent values.

Map 6. Owner Occupant and Diversity areas according to Greenberg's diversity index.

(33%) diversity areas; the Tagalog speaking households are also located in mid-high (25%) and mid-low (40%). The Spanish-speaking households, on the other hand, are almost evenly split between high diversity (32%), mid-low (25%) and low diversity (27%) areas. Thus, the groups living in the moderately priced districts (as defined by property values and median rents) appear to be the Chinese and Tagalog speakers.

The results reported above are very interesting per se but it would be much better if they can be compared with corresponding values for other cities. Pamuk (2006), 139, in passim) reports that in metropolitan areas such as Los Angeles, San Francisco and New York, there is a general trend for China and Filipino-born to own moderately priced homes and to form clusters in major suburban areas. Mexico-born, on the other hand, show significant housing problems such as living in overcrowded households, with living expenses that tend to be very high relative to their income. In Los Angeles, Mexico-born tend to live in areas with affordable rental housing. So if we assume that the households groups just mentioned correspond to the ethnic groups investigated by Pamuk, then we have a similar distribution in Detroit.

Ethno-linguistic distribution and foreign language instruction (FLI)

General description of foreign language instruction (FLI) in Detroit Metro

As stated in Section 'Data Collection, Procedures and Software Used' above, we located and tried to obtain information on FLI from 1168 schools in the area. The quantitative breakdown of public versus private schools follows in Table 1 and 2 below.

We take the proportions presented above, specifically 19% of FLI for public schools and 22% of FLI for private schools as indicators that at least posting of information on FLI does not depend on whether the school is private or public. The 191 public schools plus the 38 private schools with posted information on FLI are the schools which present the basis for our discussion of the state of the art of FLI in the area.

A full-fledged discussion of bilingual education or FLI in the area lies outside of the scope of this paper. For a detailed discussion on bilingual education in the USA, the reader is referred to (Crawford 1999, in passim); for a rather rich and balanced overview of FLI in US schools, one of the best sources is (Reagan 2002, in passim). Some of the observations made in the book are especially relevant here:

1. Foreign languages are often marginalised as school subjects in the following respects:
 - a. they typically figure as electives in the curriculum;
 - b. students often receive a 'having been present' grade, that is, there is no proper grading from A to F.

Table 1. Public schools with online information on FLI.

	Number of schools	Percentage (%) of total
Public schools with FLI	191	19
Public schools without FLI	804	81
Total	995	100

Table 2. Private schools with online information on FLI.

	Number of schools	Percentage (%) of total
Private schools with FLI	38	22
Private schools without FLI	135	78
Total	173	100

2. FLI typically starts in middle and secondary school for most students.
3. FLI is very Eurocentric in that, Spanish is followed by French, German and Latin as the languages most commonly taught.
4. FLI often lacks external institutional support (Reagan 2002, 33–9).

The following can be said about the DMA when considered in comparison with the general US situation: we have no data on (4) above so we cannot comment on the availability of external support for FLI; likewise, we have no data on (1b) whether the grading of foreign languages is predominantly of the kind quoted by Reagan. As regards (2), similarly to the general US picture, most schools with FLI in Detroit and vicinity seem to offer them as electives. As regards (3), the Eurocentric bias is present here as well in that Spanish, French and German are the dominant languages in FLI. However, the language following them in popularity and spread is Japanese, something that is not mentioned in other general studies on FLI in the USA (Amman, Norbert, and Mattheier 2006; Crawford 1999; Meyerhoff 2006; Reagan 2002). Whether this is typical of Detroit or reflects an emerging trend in US schools remains to be checked.⁸ All in all, there are 12 languages that figure in World Languages programmes; we present them in Table 3 below. The figures in the columns do not add up to the totals 229 for schools with FLI because there may be more than one foreign language taught at one and the same school.

Based on the table above, the following can be said about a quantitative correlation between the particular languages taught and the groups who speak them. Chinese, Russian and Hebrew can be seen as languages which show almost matching

Table 3. Schools with foreign language instruction in Detroit Metro.

Language name	Number of schools where it's taught	Percentage (%) of all schools with FLI (229)	Percentage (%) of the ethnic group from all non-English speaking households (458,582 people)
Spanish	202	88	21
French	122	53	4.3
German	72	31	4.6
Japanese	30	13	1.4
Latin	12	5.2	None
Arabic	11	4.8	14.9
American sign language (ASL)	11	4.8	No information
Chinese	9	3.9	3.4
Russian	4	1.7	1.9
Italian	3	1.3	5.4
Classical Greek	2	0.87	None
Hebrew	2	0.87	0.87

correspondence of terms of the proportion of schools where they are taught to the proportion of the foreign population their minority represents (however, see more on Chinese below). Languages that are over-represented in terms of being taught as compared to being used at home are Spanish, French, German and Japanese. The languages that are clearly under-represented in teaching when compared with the number of people who speak them at home are Arabic and Italian. Here we are leaving the consideration of whether a language has a special status as a cosmopolitan language for a later discussion.

Relating the spatial distribution of foreign language instruction (FLI) with the settlement patterns of the groups who speak these languages at home

The spatial distribution of the schools where foreign languages are taught in Detroit Metro is not random. Generally, such schools tend to appear in low and mid-low diversity areas. This means either that schools in areas with less immigrants are better organised to disseminate information about their programmes or that there is a stronger institutional support for foreign languages in these areas. The spatial distribution of the teaching of specific languages shows various degrees of ties to the ethnic population who use them at home. For instance, Arabic, Russian, Italian and Hebrew show a very strong correlation; Japanese shows a similar, though less strong correlation as it is also taught in a very high number of schools outside of areas with Japanese households in them. Conversely, schools where Mandarin is taught show no connection whatsoever to areas with Chinese speaking households. For the most frequently taught languages, the only clear correlation appears to be for Spanish, in that schools that teach it are outside of the areas with highest number of Spanish speaking households in them (which is an essence a reiteration of the general tendency for FLI schools to be outside of high diversity areas).

To conclude, there are in all 11 foreign languages plus American sign language (ASL) included in school curricula in Detroit Metro. This is in sharp contrast with the actual ethno-linguistic diversity of the area where 88 different languages are spoken in various households. This situation is, sadly, rather typical of what goes on in foreign language studies both in the USA and around the world. For instance, Mikael Parkvall, linguist at Stockholm University, shared with us his data collection on second language learning around the world. According to his current dataset, the languages most frequently learned as foreign languages in 192 countries of the world are as follows: English (29%), French (23%), German (14%), Arabic (7%), Spanish (6%), Russian (5%), Italian (4%), Mandarin/Portuguese/Japanese (2%). For the USA, Reagan (2002, 41) quotes Walker (1989, 111) who states that ‘ninety-one percent of the academic study of foreign languages is directed towards languages used by twelve to thirteen percent of humanity’.

Minority languages of Detroit: how strong are they?

Giles, Bourhis, and Taylor (1977) propose to model the vitality (strength) of language vitality where three main factors are considered of greatest importance: socio-economic status of the speakers of a language, demographic strength and institutional support. The model is typically summarised by the diagram below (Figure 1).

AQ4 Figure 1.

Naturally, the three general factors shown above, together with sub-factors are highly dependent on one another. For instance, ethno-linguistics groups of high economic and social status tend to have a strong institutional support and even full control of media, education and government. Thus, their languages are considered very strong.

The following can be said about the main minority languages of Detroit when viewed in terms of this model. Spanish is demographically strong both nationally and in the area. It enjoys institutional support in that it is the most widely taught foreign language. However, it is also associated with low status economic immigrants, something that may contribute to a general lower status of the language. Arabic, on the other hand, shows local demographic strength and its social and economic status are, again locally, on the low side of the scale. The institutional support it receives in terms of language teaching is very closely connected to areas where Arabic speakers reside. This in turn shows that that sheer numbers of speakers may trigger institutional support and thus, make the language stronger, albeit within a very confined geographic space.

On the other hand, it is mostly socio-economic status that plays a decisive role for the prominence of French, German, Japanese and Mandarin. For in instance, in the case of Japanese: its ethnic group ranks among the smaller groups in the area; however, Japanese companies are very prominent in the car industry; likewise, American-Japanese exchange is vibrant. These factors seem in turn to trigger institutional support and interest for teaching Japanese at schools. The size of the actual population seems to be least important in this case.

Summary and conclusions

In this paper we used census data together with GIS technology to map the language structure of the DMA.

Apart from identifying the 88 languages spoken in the area, we also offered quantitative and spatial analysis. The latter showed that there are two main settlement patterns for non-English speakers: settlement in close communities or living in a more spread fashion. These different settlement patterns can be used as indicators of recency of immigration but are also probably due to different motivation(s) for immigration in the first place.

Using the Greenberg's index for measuring linguistic diversity, we were able to outline zones on scale of very low to very high diversity. These zones could in turn be shown to relate, at least to a certain extent with economic factors such as property values and median rent. The settlement patterns of the different ethnic groups were further compared and related with the economic factors mentioned above as well as with FLI.

While the census data show a number of limitations, they can still be used in combination with careful use of GIS tools for initial, infra-structure setting studies such as this one to provide the basis necessary for further research on language in the city.

Notes

1. The Detroit Primary Metropolitan Statistical Area (PMSA) as defined by the US Office of Management and Budget (OMB) is defined as Wayne, Macomb, Oakland, St. Clair, Monroe and Lapeer Counties. For the purposes of this study we have decided to only use the core counties of the metropolitan area (Wayne, Oakland and Macomb), which account for 91% of the metropolitan area's population and 40% of the State of Michigan's population.
2. We opted to these variables as they were in a tabulated format that could be easily geocoded. Data on income, occupation and level of education were not available in a format that we could work with.
3. The questions about language in the 2000 Census are grouped under number 11 (a, b and c) in the long form of the census questionnaire, which can be found at <http://www.census.gov/dmd/www/pdf/d02p.pdf>. This and subsequent URLs are valid as of 14 March 2007.
4. The Canadian Census is much better suited to provide data for proper sociolinguistic research but unfortunately, full access to them is limited to Canadian researchers. So if a comparative cross-linguistic study is to be done, one has to establish a multi-national project(s); something that is potentially very promising but also involves much more complex administration thus, can be more difficult to implement.
5. For instance, we used the most detailed data set possible. Language groups that had no grounding in either historical or areal linguistics were discarded as well, again, to the extent possible.
6. For more details on confidentiality procedures, please consult http://factfinder.census.gov/jsp/saff/SAFFInfo.jsp?_pageId=su5_confidentiality.
7. These generalisations can be further specified by numeric data which can be obtained by applying the Select by Location Operation in a GIS tool such as ArcMap.
8. Reagan (2002), 41) reports the following statistics for students in American public schools: in 1996 some six million students in middle and secondary school studied a foreign language. Among them, 65% studied Spanish, 22% French, 6% German and 3.5% Latin. The remaining 3.5% of the students studied a variety of other languages such as ASL, Arabic, Chinese (both Cantonese and Mandarin), Greek, Haitian Creole, Hawai'ian, Hebrew, Russian, Italian, Japanese, Korean, Norwegian, Polish, Portuguese, Russian, Vietnamese, and a variety of both African and Native American languages.

Since we currently do not have reliable data on the numbers of students in the schools we investigated we are not in a position to make a comparison with the proportions reported by Reagan. However, we hope that the proportions reported here, which are based on counting schools rather than the student population, will give an initial idea of the state of FLI in Detroit Metro.

References

- Aitchinson, J.W., and H. Carter. 1985. *The Welsh language 1961–1981: An interpretive atlas*. Cardiff, Wales: University of Wales Press.
- Amman, U., D. Norbert, and K.J. Mattheier, eds. 2006. *Sociolinguistics: An international handbook of the science of language and society*. Berlin, Germany: Mouton de Gruyter.
- Angle, J. 1981. The ecology of language maintenance – data from nine US metropolitan areas. *Urban Affairs Quarterly* 17: 219–31.
- Bains, B. 2005. *DMAG briefing 2005/12: Ethnic diversity indices*. London: Greater London Authority, Data Management and Analysis Group.
- AQ5** Barni, M. 2006. From statistical to geolinguistic data: Mapping and measuring linguistic diversity. Vol. 53. Fondazione Eni Enrica Mattei Note di Lavoro Seris Index. 12.
- AQ5** Breton, R. 1997. *Atlas of the languages and ethnic communities in South Asia*. Crawford, J. 1999. *Bilingual education: History, politics, theory and practice*. Los Angeles, CA: Bilingual Educational Services.
- Extra, G., and K. Yağmur. 2004. *Urban multilingualism in Europe: Immigrant minority languages at home and school: Multilingual matters*. Clevedon, UK: Multilingual Matters.
- Giles, H., R.Y. Bourhis, and D.M. Taylor. 1977. Towards a theory of language in ethnic group relations. In *Language, ethnicity and intergroup relations*, ed. Howard Giles, 307–48. London: Academic Press.
- Gorter, D., ed. 2006. *Linguistic landscape: A new approach to multilingualism*. Clevedon, UK: Multilingual Matters.
- Green, L. 2002. *African American English: A linguistic introduction*. Cambridge, UK: Cambridge University Press.
- Greenberg, J.H. 1956. The measurement of linguistic diversity. *Language* 32: 109–15.
- Grobler, E., K.P. Prinsloo, and I.J. Van der Merwe. 1990. *Language atlas of South Africa: Language and literacy patterns*. Pretoria: Human Sciences Research Council.
- Heine, B., and W. Möhlig, eds. 1980. *The Non-Bantu languages of Kenya. Language and dialect atlas of Kenya*. Vol. 2. Berlin, Germany: Dietrich Reimer Verlag.
- Higgs, G., C.H. Williams, and D. Dorling. 2004. Use of the census of population to discern trends in the Welsh language: An aggregate analysis. *Area* 36: 187–201.
- Landry, R., and R.Y. Bourhis. 1997. Linguistic landscape and ethnolinguistic vitality: An empirical study. *Journal of Language and Social Psychology* 16: 23–49.
- Meyerhoff, M. 2006. *Introducing sociolinguistics*. London: Routledge.
- Mikesell, M.W., and A.B. Murphy. 1991. A framework for comparative study of minority group aspirations. *Annals of the Association of American Geographers* 81: 581–604.
- Moseley, C., and R.A. Asher. 1994. *The atlas of world's languages*. London: Routledge.
- Pamuk, A. 2006. *Mapping global cities*. Redlands, CA: ESRI.
- Pienemann, M., and J.-U. Kefler. 2007. Measuring bilingualism. In *Handbook of multilingualism and multicultural communication*, ed. P. Auer and L. Wei, 247–75. Berlin, Germany: Mouton de Gruyter.
- AQ6** Ratliff, M. 2005. Southeast Asian languages in Michigan and Ohio. In *Language diversity in Michigan and Ohio*, ed. B. Joseph, C.G. Preston, and D.R. Preston. Ann Arbor, MI: Caravan Books.
- Reagan, T. 2002. *Language, education and ideology: Mapping the linguistic landscape of U.S. schools*. London: Praeger.
- Smitherman, G. 1977. *Talking and testifying: The language of Black America*. Boston, MA: Houghton Mifflin.
- Van der Merwe, I.J. 1993. The urban geolinguistics of Cape Town. *GeoJournal* 31: 409–17.
- AQ7** Walker, G. 1989. The less commonly taught languages in the context of American pedagogy. In *Northeast conference on the teaching of foreign language: Shaping the future*, ed. H. Lepke, 111–37. Middlebury, VT.

Williams, C.H. 2000. *Language revitalization: Policy and planning in Wales*. Cardiff, Wales: University of Wales Press.

Williams, C.H., and I. Van der Merwe. 1996. Mapping the multilingual city: Research agenda for urban geolinguistics. *Journal of Multilingual and Multicultural Development* 17: 49–66.

Wolfram, W. 1969. *A sociolinguistic description of Detroit Negro speech: Urban language series*, no. 5. Washington, DC: Center for Applied Linguistics.

Appendix 1. Alphabetical list of all languages spoken in the households of the Detroit Metropolitan Area (DMA)

Language Name	Number of speakers	Percentage (%) of Total (3,759,434)
Afrikaans	85	0.0023
Albanian	12,455	0.3313
Amharic	245	0.0065
Arabic	65,526	1.7429
Armenian	3025	0.0804
Assyrian	24,785	0.6592
Bengali	5215	0.1387
Bisayan	80	0.0021
Bulgarian	515	0.013691
Cantonese	845	0.02247
Cebuano	80	0.0021
Chamorro	40	0.0010
Cherokee	90	0.0023
Chinese	14,968	0.3981
Choctaw	50	0.0013
Cushite	130	0.0034
Czech	590	0.0156
Danish	215	0.0057
Dutch	1470	0.0391
Efik	90	0.0023
English	3,322,071	88.366
Estonian	105	0.0027
Finnish	780	0.0207
Formosan	415	0.0110
French (including Patois, Cajun)	18,747	0.4986
French Creole	183	0.0048
Fulani	65	0.0017
German	20,146	0.5358
Greek	7993	0.2126
Gujarathi	5209	0.1385
Hawaiian	35	0.0009
Hebrew	2075	0.0551
Hindi	6805	0.1810

Appendix 1 (Continued)

Language Name	Number of speakers	Percentage (%) of Total (3,759,434)
Hungarian	2916	0.0775
Ilocano	35	0.0009
Indonesian	240	0.0063
Irish Gaelic	240	0.0063
Italian	23,839	0.6341
Jamaican Creole	110	0.0029
Japanese	5950	0.1582
Kannada	825	0.0219
Korean	5716	0.1520
Krio	20	0.0005
Kurdish	35	0.0009
Lao	343	0.0091
Latvian	285	0.0075
Lithuanian	985	0.0262
Macedonian	4455	0.1185
Malayalam	2265	0.0602
Mandarin	875	0.0232
Marathi	1140	0.0303
Miao, Hmong	4420	0.1175
Mon-Khmer, Cambodian	323	0.0085
Muskogee	55	0.0014
Navajo	12	0.0003
Nepali	25	0.0006
Norwegian	210	0.0055
Ojibwa	85	0.0022
Oriya	30	0.0007
Ottawa	25	0.0006
Panjabi	2960	0.0787
Pennsylvania German	20	0.0005
Persian	1903	0.0506
Polish	29,319	0.7798
Portuguese or Portuguese Creole	1514	0.0402
Romanian	7730	0.2056
Romany	30	0.0007
Russian	8254	0.2195
Scott Gaelic	15	0.0003
Serbo-Croatian	7906	0.2102
Sindhi	75	0.0019
Sinhalese	40	0.0010
Slovak	840	0.0223
Slovene	55	0.0014
Spanish or Spanish Creole	92,626	2.4638
Swahili	475	0.0126
Swedish	330	0.0087
Tagalog	8375	0.2227
Tamil	2560	0.0680
Telugu	4205	0.1118

Appendix 1 (*Continued*)

Language Name	Number of speakers	Percentage (%) of Total (3,759,434)
Thai	695	0.0184
Turkish	495	0.0131
Ukrainian	4440	0.1181
Urdu	5927	0.1576
Vietnamese	3883	0.1032
Welsh	55	0.0014
Yiddish	1755	0.0466
Yoruba	2365	0.0629

Appendix 2. Languages arranged according to decreasing number of speakers

The census data had to be split in two groups: one called *Individual languages*, and one to include the various groupings done in the census summaries; the latter appears in the Google application as *Census Groups*. The groups made by the census can be said to be of two kinds: those which correspond to language families (or parts thereof) and those which correspond to large geographical areas. We refer to the first one as *Genealogical groups*; it covers groups entitled 'Other Indo-European', 'Other Slavic' etc. and the second *Geographical groups*; it includes groups such as 'Languages from Africa/Asia/Pacific Islands', etc. For languages with a very small number of speakers, information is available on county level only. We indicate such languages by additional pop-up windows in the Google Earth application; such explanatory links are also planned for the upcoming ArcIMS map server application (see addresses for both in Section 'Data Collection, Procedures and Software Used' above).

The group entitled *Individual Languages* includes languages which are covered one by one in the census files and are not included in any of the census groups. For these languages, the available information is most complete in that it typically covers both the number of speakers for each languages as well as spatial distribution on tract level. A classification according to number of speakers is presented in Table 1 below.

Classification of individual languages according to number of speakers.

Group	Number of languages	Language Name(s)
More than 100,000	1	English (3,322,071)
50,000–100,000	2	Spanish (92,626) and Arabic (65,526)
10,000–50,000	7	Polish, Assyrian, Italian, German, French, Chinese, Albanian
5000–10,000	9	Russian, Tagalog, Greek, Serbo-Croatian, Hindi, Japanese, Urdu, Korean, Gujarati
Less than 5000	11	Hmong Miao, Vietnamese, Armenian, Hungarian, Persian, Portuguese, Thai, Lao, Mon Khmer Cambodian, French Creole, Navajo

In the tables below, we present the contents of the two kinds of census groups. Each group is listed according to decreasing number of speakers (and also decreasing amount of information, especially for the geographical groups).

Genealogical groups

Group	Number of speakers	Number of languages: Language names
Other Indo-European	22,062	15: Lithuanian, Latvian, Irish Gaelic, Scottish Gaelic, Welsh, Bengali, Panjabi, Marathi, Sindhi, Sinhalese, Oriya, Romany, Nepali; Kurdish, Romanian No specific information for 2997 speakers
Other Slavic	10,974	5: Macedonian, Ukrainian, Slovak, Czech, Bulgarian No specific information for 79 speakers
Other West Germanic	1576	3: Dutch, Afrikaans, Pennsylvania Dutch
Scandinavian	811	3: Swedish, Danish, Norwegian No specific information for 56 speakers

Geographical groups

Group	Number of speakers	Number of languages: Language names (number of speakers in the parentheses)
Languages from India	11,397	4: Telugu, Tamil, Malayalam, Kannada No specific information for 1542 speakers
Other languages from Asia	10,446	3: Mandarin, Cantonese, Turkish No specific information for 8251 speakers
Languages from Africa	4010	6: Amharic, Efik, Fulani, Swahili, Yoruba No specific information for 640 speakers
Languages from islands in the Pacific	960	7: Formosan, Indonesian, Cebuano, Ilocano, Chamorro, Hawaiian, Bysayan group. No specific information for 40 speakers
Native languages of North America	598	5: Ojibwa, Ottawa, Cherokee, Choctaw, Muskogee No specific information for 293 speakers

Finally, a group called *Other and unspecified languages* includes languages such as Assyrian, Finnish, Estonian, and two Creoles, Jamaican Creole and Krio. The original census group is said to cover 25,082 speakers. However, 20,760 are speakers of Assyrian; there is information about them on tract level so there is really no reason for them to be in this group, other than the language name, which can be subject to dispute. The remaining 4322 speakers fall into groups as follows: Finnish (780), Estonian (105), Jamaican Creole (110) and Krio (20). There is no specific information for the remaining 3307 speakers.