

Using GIS to Map the Multilingual City

Ljuba Veselinova and Jason Booza

ljuba@ling.su.se

j.booza@wayne.edu

ABSTRACT

Work in two rather distinct disciplines, urban geography and sociolinguistics, readily points out the multiethnic and multilingual character of big cities. However, there is still demand for studies which establish the language structure of modern cities. This paper presents a pilot study where data on multilingualism is collected and presented visually using ArcGIS and ArcIMS. GIS technology together with census data and the results of fieldwork is used to arrive at an adequate description of the spatial distribution of languages currently spoken in the Detroit, Michigan, metropolitan area. Two levels of language used are mapped: the mother-tongue speakers and the dominant languages used in work situations. The study thus suggests one possible model for the initial stages of mapping the multilingual city; moreover, the collected data provides 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.

1. Introduction

Multilingualism is a widespread phenomenon (cf Trudgill 1995, Paulston and Tucker 2003: 379, to name just a few general sources) but this is rarely, if ever, represented on language maps. Following Williams and Van Der Merwe (1996), this paper presents a pilot study where census data are used together with GIS technology for the study of multilingualism. We use data from the 2000 US Census; for the purposes of the current work, we concentrate on the Detroit Metropolitan Area. Administratively, the area covers the territories of three counties, Macomb, Oakland and Wayne in southeastern Michigan (see map section 3). Due to limitations of the census data, and rather time consuming procedures they entailed, at this stage the data had to be restricted to mapping languages in household use only.

The presentation below is organized as follows. In section 2, we discuss the main problems encountered when working with census data. In sections 3.1 and 3.2, we present the linguistic profile of the metropolitan area as a whole and compare it with the linguistic profiles of Detroit City proper together with the linguistic profiles of the two cities it surrounds: Highland Park and Hamtramck. Further analysis and discussion of the spatial distribution of the identified languages can be found in sections 3.3 and 3.4. Conclusions and directions for future research are outlined in section 4.

2. Method, procedures and software used in this work

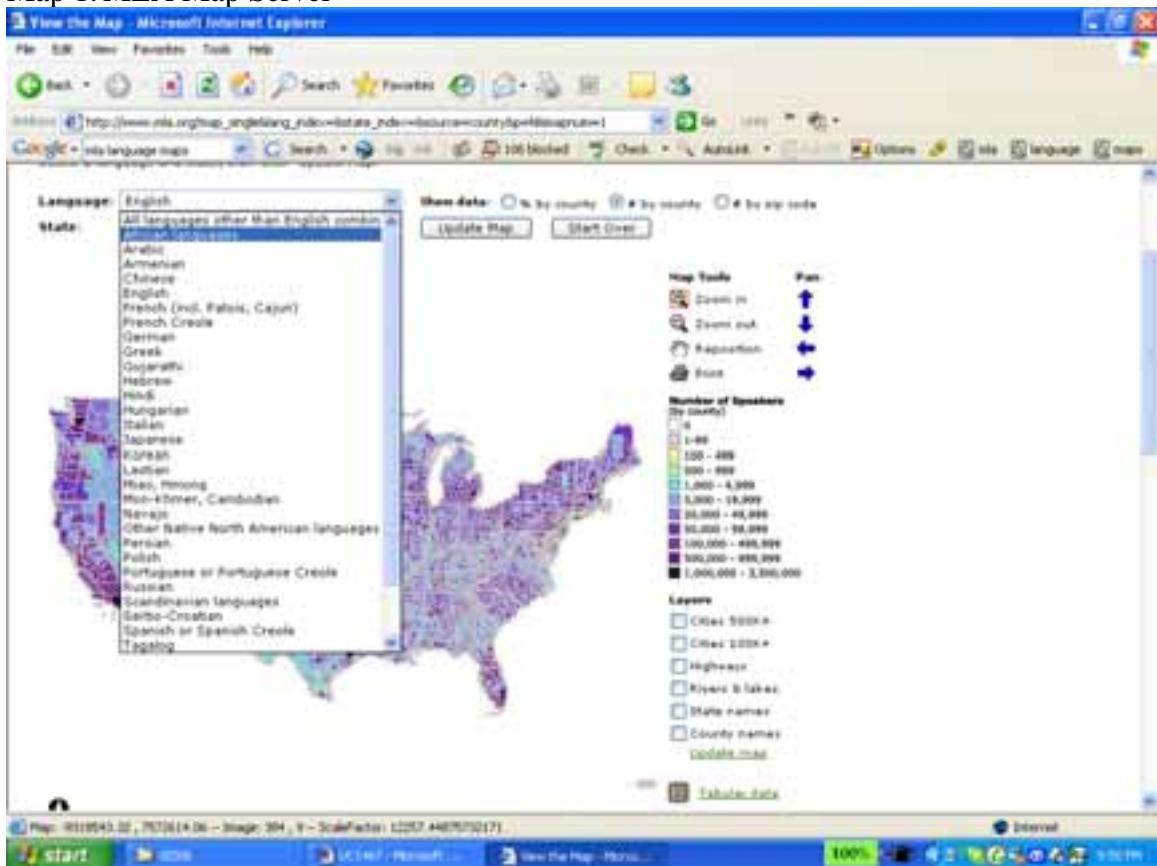
2.1. *Problems with using census data for linguistic studies*

Census data on the languages used in a given area in many ways an indispensable source of information but working with them face value presents problems in several respects.

First, the language data are restricted to home use since the participants in the investigation are asked a question such as ‘Do you use a language other than English in your household?’. If the answer is affirmative, subjects are asked to name the language used. However, the fact that a certain language is used at home tells us nothing about the degree of proficiency of its speakers. Thus census data provide information about languages used in a particular context without any information as regards fluency. While we can hardly correct this, we need to be aware of these limitations.

Second, the identification of languages is highly unsatisfactory. For instance, consider the language map server offered by the Modern Language Association ([http://www\(mla.org/census_map](http://www(mla.org/census_map))>, snapshot of which is presented in Map 1¹.

Map 1. MLA Map Server



The map server cited above is based on census data (the year of the census is not specified). The linguistic profile of the US as shown above is at best misleading.

¹ The reader is encouraged to follow the link

Although it is stated in the introduction that more than 300 languages are spoken in the country, the languages or language groupings shown on the server amount to 33 and it is not clear what criteria were used for this selection. Apart from the limited number of languages shown for the US, many individual languages are presented in an erroneous way. For instance, Spanish and Spanish Creole are mapped together; however, they are completely different languages and consequently should be mapped separately. Similar comments can be made for Portuguese and Portuguese Creole: the term ‘Chinese’ is too general to be truly informative, etc. Finally, a lot of languages are grouped together according to some loose geographical principles, like, for example, 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 mapping languages together just because they stem from there is not justified.

The full set of census data

(http://www.census.gov/population/www/socdemo/lang_use.html), which is not on the MLA map server, includes also groupings which appear to follow, again very loosely, some rather general genealogical classification. For instance, we encounter groupings such as ‘Other Indo-European’ or ‘Other Slavic’. A closer examination of the number of people these groups include reveals that very often a huge amount of speakers is covered by such general denominations. For example, for the Detroit Metropolitan area, the grouping ‘Other Slavic’ covers some 10,000 speakers, it is important to indicate whether these speakers use one and the same language or several different languages.

As already pointed out, awareness of the shortcomings census data is necessary since they bear directly on the implications and the conclusions we draw in such studies. However, we can hardly do studies in multilingualism without census data so we need learn to use such data in the best possible way(s).

2.2. Procedures and software used

For the purposes of this study we examined both the generalized census data (cf. http://www.census.gov/population/www/socdemo/lang_use.html, Summary File 3). The language data in this general source are available on tract level so linking them to shape file which also contain tract information is easy to do. The information given in this file is not sufficient, cf. discussion in 2.1 above. To provide, at least, a partial remedy for its various shortcomings we also used the information provided in http://www.census.gov/population/www/socdemo/lang_use.html, Detailed Language Spoken at Home for the Population 5 Years and Over (STP224) (hereafter ‘Detailed Use’ File).

The collected data were analyzed in ArcMap and displayed for public use in a Google Earth application, <http://lingmap.emich.edu/website/Detroit002.kmz>. An ArcIMS map service for alternative display of these data is currently under development.

The census data had to be split in two groups: one for individual languages, and one to include the various groupings done in the census summaries. 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 call the first one **Genealogical groups** such as ‘Other Indo-European’, ‘Other Slavic’ etc and the second **Geographical groups**, such as ‘Languages from Africa/Asia/Pacific Islands’ etc. The census general summary files do not specify which languages are included in each

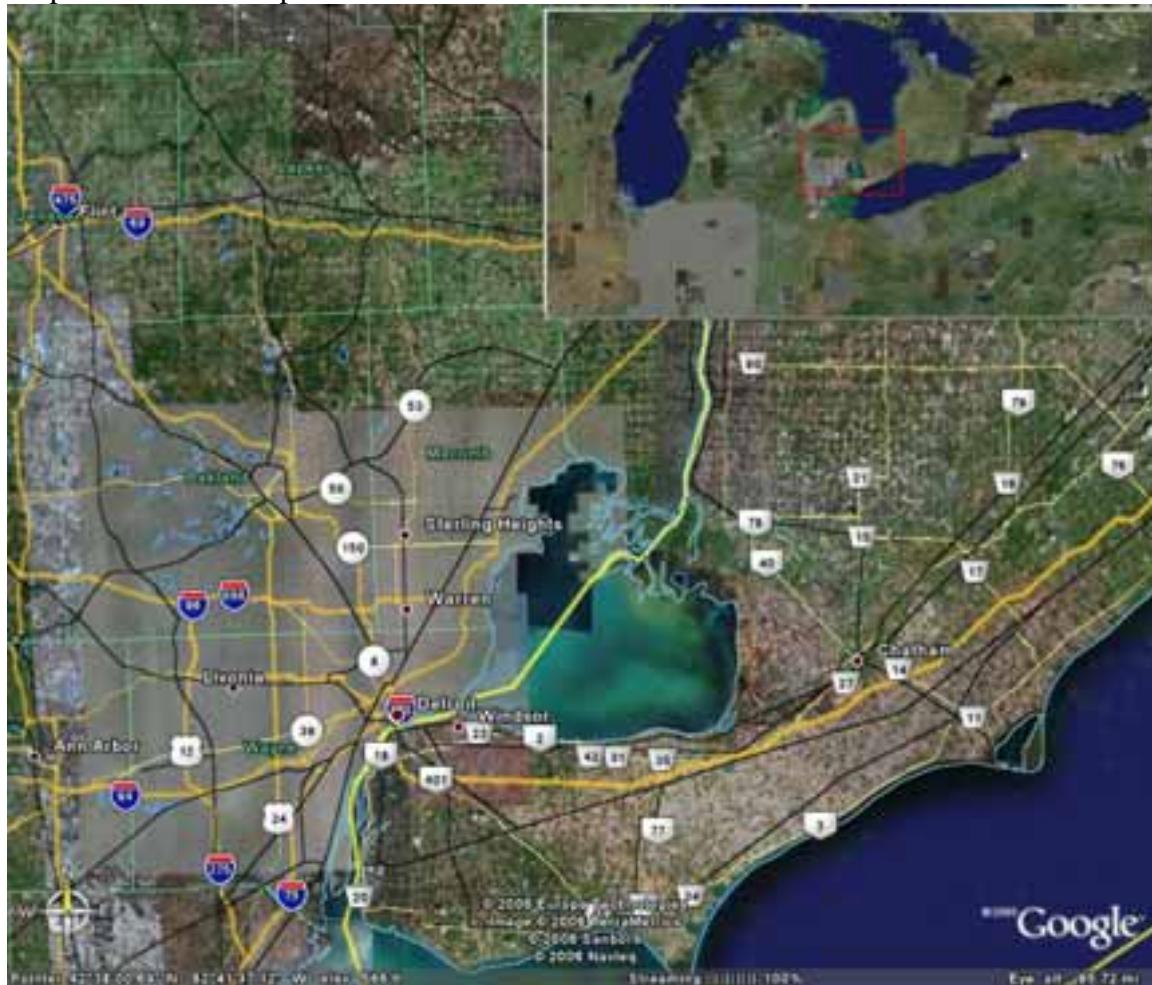
group; to this end, the information found in the Detailed Use File was helpful in that it provided lists of languages included in each group. However, for some languages, there is information on county level only, which in turn makes plotting the approximate location of speakers rather difficult. At this stage of the research, we indicate such languages by additional pop-up windows in the Google Earth application; such explanatory links are also for the upcoming ArcIMS map server which is to reside at <http://lingmap.emich.edu/website/detroit/> and <http://ling-map.ling.su.se/website/detroit/>.

3. The linguistic profile of the Detroit Metropolitan Area

3.1. *The area as a whole*

The Detroit Metropolitan Area includes three counties: Macomb, Oakland and Wayne, all shown on Map 2. below

Map 2. Detroit Metropolitan Area



There are more than 90 different languages spoken in the three counties shown above: there are 89 languages for which we have information in terms of number of speakers together with their location. As stated above, for some languages location of speakers is specified on tract level, for others, this kind of information is only available on county level.

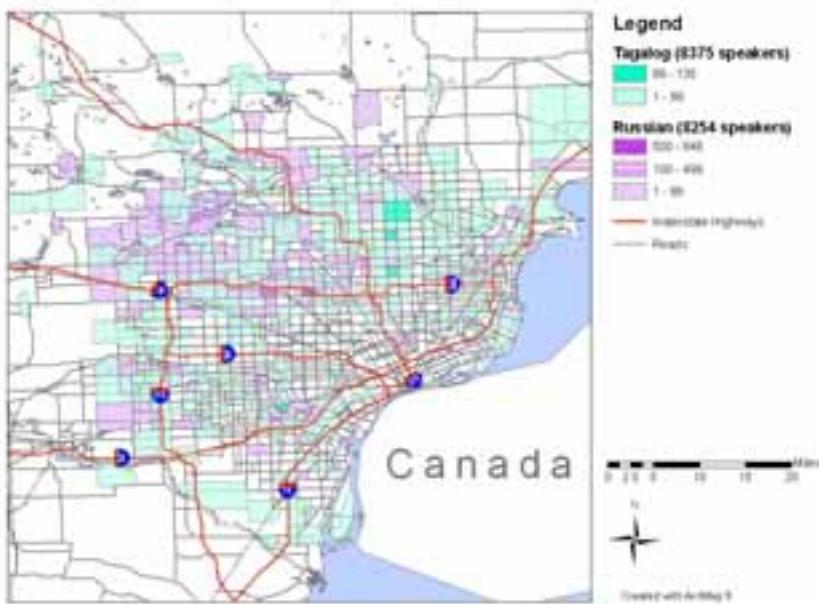
The group of individual languages covers the ones for which most complete information is available; they were further grouped together according to number of speakers, shown in Table 1 below.

Table 1. 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 to 100,000	2	Spanish (92,626) and Arabic (65, 526) ²
10,000 to 50,000	7	Polish, Syriac, Italian, German, French, Chinese, Albanian
5,000 to 10,000	9	Russian, Tagalog, Greek, Serbo-Croatian, Hindi, Japanese, Urdu, Korean, Gujarati
Less than 5,000	11	Hmong Miao, Vietnamese, Armenian, Hungarian, Persian, Portuguese, Thai, Lao, Mon Khmer Cambodian, French Creole, Navajo

Each language listed in the table above is mapped by a separate layer in Google Earth/ArcMap and can be displayed individually or in combination with other languages. Showing all languages simultaneously is impractical, for obvious reasons. The service is illustrated by Map 3 demonstrating the distribution of Tagalog and Russian, which appear in a somewhat complementary distribution. The reader is encouraged to experiment with the other layers (<http://lingmap.emich.edu/website/Detroit002.kmz>).

Map 3. Tagalog and Russian in the Detroit Metropolitan Area



² The table will become rather messy if all languages appear with corresponding number of speakers. This information can be easily obtained from the Google Earth placemark cited above.

There are 4 Genealogical groups and 5 Geographical groups. Similarly to individual languages, the groups are listed according to decreasing number of speakers (and also decreasing amount of information, especially for the geographical groups).

The largest genealogical group is entitled ‘Other Indo-European’. It covers 22052 speakers; specific language information is available for 19065 speakers. The languages with the largest number of speakers are: Romanian (7730), Bengali (5215), Panjabi (2960), Marathi (1140) and Lithuanian (985); these counts are also, at least partially associated with specific tracts which makes a partial plotting of each language possible. However, since the information on tract level is incomplete, they are still listed as a group. Other languages in the group called ‘Other Indo-European’ are Latvian (285), Irish Gaelic (240), Scotts Gaelic (15), Welsh (15), Sindhi (75), Sinhalese (40), Oriya (30), Romany (30), Nepali (25), and Kurdish (35). The spatial information about these languages is on county level only which is too general to make mapping meaningful. This is why these languages do not appear as separate layers on the Google service; these languages and likewise all the other languages without information on tract level are only listed in the pop-up windows for their respective groups.

Other groups which can be described as based on genealogical relationships are ‘Other Slavic’, the biggest languages in it being Macedonian and Ukrainian; another group covers Dutch, Afrikaans and Pennsylvania German (which in the census appears as Pennsylvania Dutch) and finally, Scandinavian languages.

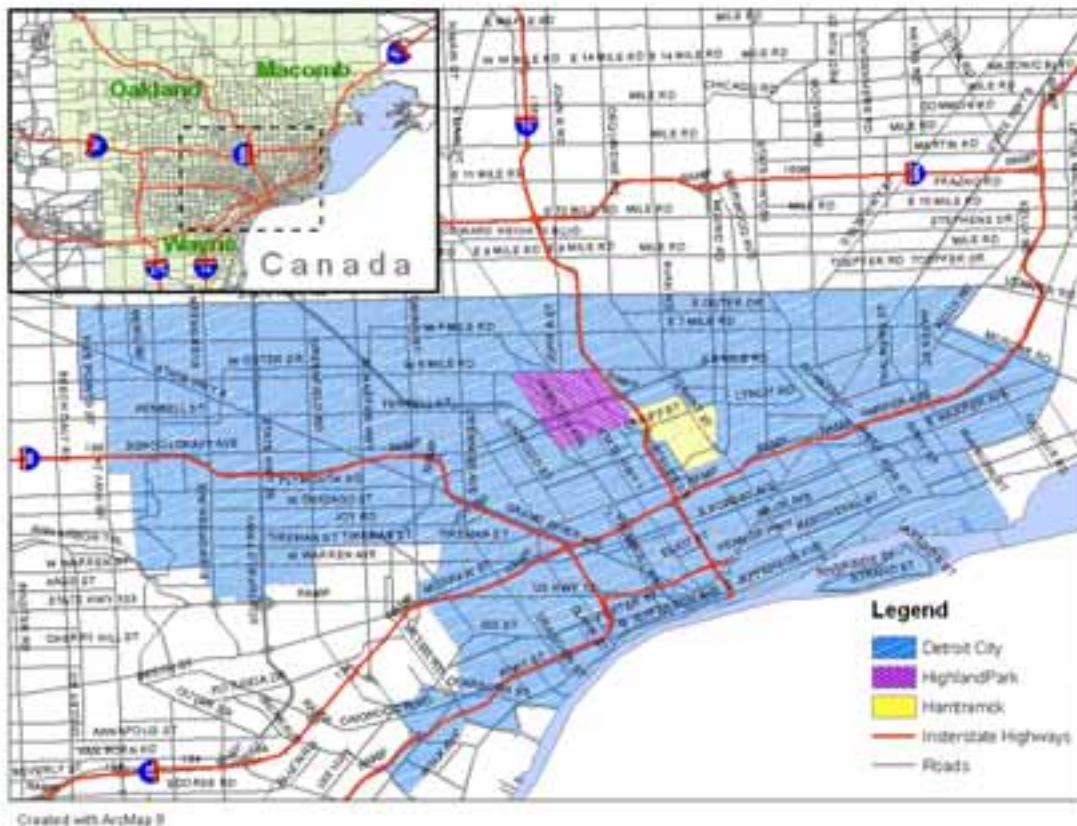
The largest geographical group covers languages from India; the languages found in this group are predominantly Dravidian but since the number of speakers given for them is less than for whole group, it remains unclear what other languages are included in this group. The groups entitled ‘Languages from Africa’ and ‘Languages from Asia’ are very poorly specified even in the detailed census files. The same can be said about the groups entitled ‘Languages from Islands in the Pacific’ and ‘Native American Languages’; the first one appears to include mainly Austronesian languages whereas Algonquian, Iroquois and Muskogean seem to cover half of the speakers included in the Native American group (however, we still need information about the other half).

It should have become clear by now that the Detroit metropolitan area represents a very complex linguistic landscape. The dominant language is, not surprisingly, English, followed by Spanish, Arabic and a wide variety of languages some of which were listed above. The linguistic profile of the metro area is compared with the profiles of Detroit City and the two cities it surrounds in the next section.

3.2. The linguistic profiles of the core areas

The core city in the area is Detroit City which surrounds the smaller cities of Highland Park and Hamtramck as shown on Map 4.

Map 4. Detroit, Highland Park and Hamtramck



In the charts below (Figure 1 and 2), we present the quantitative distribution of the languages in the metro area (the overview map in Map 4 above, see also Map 2 above) together with the profiles of its core areas.

Figure 1. Linguistic Profiles of Detroit and surrounding areas

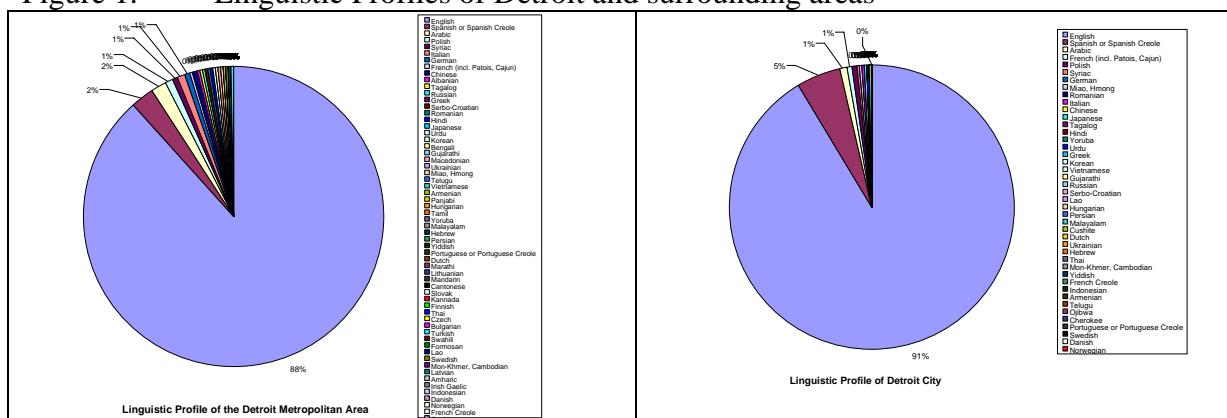
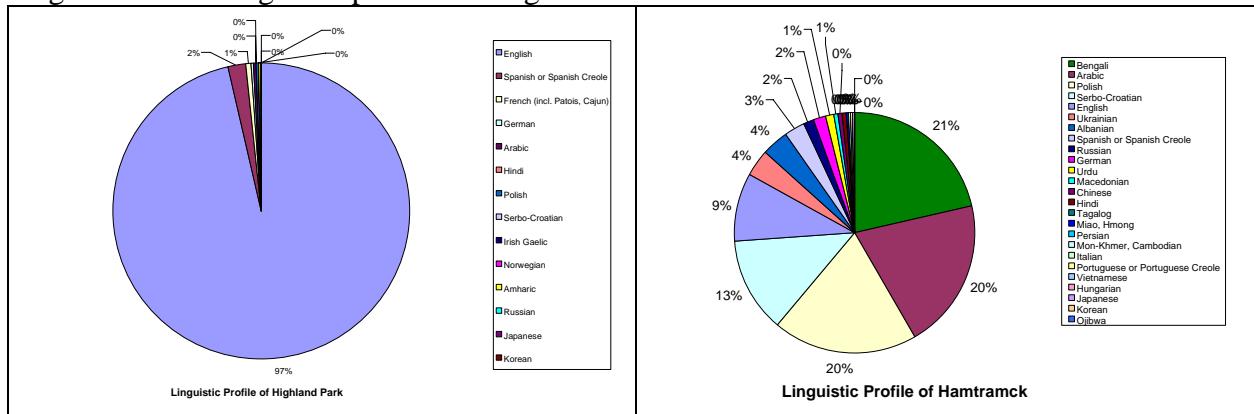


Figure 2. Linguistic profiles of Highland Park and Hamtramck



As we can see, the Metro area and Detroit City have similar profiles with English as a dominant language but in both areas, a huge number of other languages are also spoken. There is a noticeable Spanish presence in Detroit City. Highland Park and Hamtramck, on the other hand, are in stark contrast with each other and to a certain extent with the Metro Area and Detroit City as well. Highland Park is almost entirely monolingual with English. In Hamtramck, the dominant languages are Bengali and Arabic which have only recently come to outnumber the formerly dominant Polish (cf. also Danner and Peart 1999: 266-7). English speaking households appear to hold a minor position in Hamtramck.

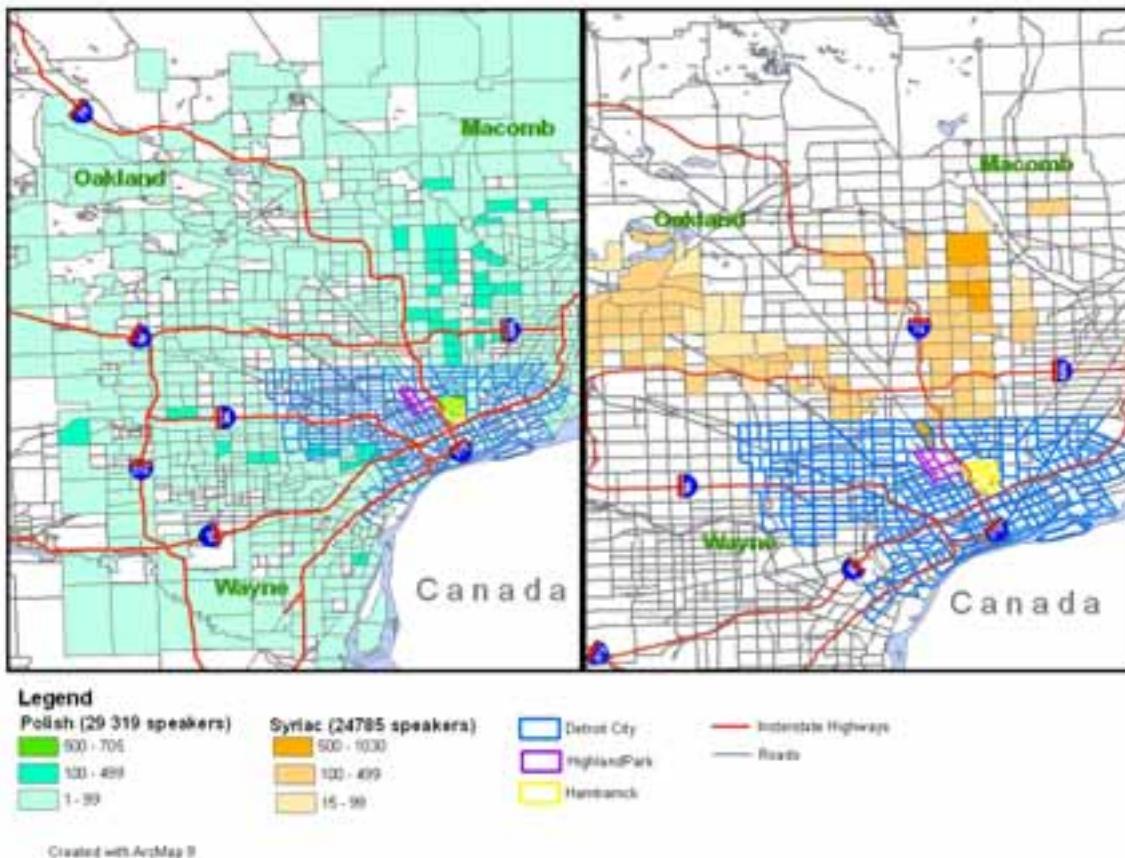
Thus we can see that even for a quantitative ranking of the languages in the area, we still need to bring in a spatial dimension. The charts above also show that there are areas where a high degree of homogeneity is harbored, e.g. Highland Park whereas in Hamtramck, we observe a much greater diversity. Further observations on the spatial distribution of the languages in greater Detroit are offered in the sections below.

3.3. Clustering vs. non-clustering languages

There are languages which form clear dense areas whereas others hardly form any dense areas at all. Consider, for instance the settlement of Polish and Syriac³ speakers shown on Map 5. The two languages are relatively close in terms of number of speakers in the area: around 30,000 for Polish and 25,000 for Syriac.

³ Syriac is a Modern Aramaic language from Iraq. The term **Syriac** in linguistics is used to refer to a dead Aramaic language. However, this is also one of the names native speakers use to refer to their language which is why it is retained in our work as well.

Map 5. Polish and Syriac in Metro Detroit

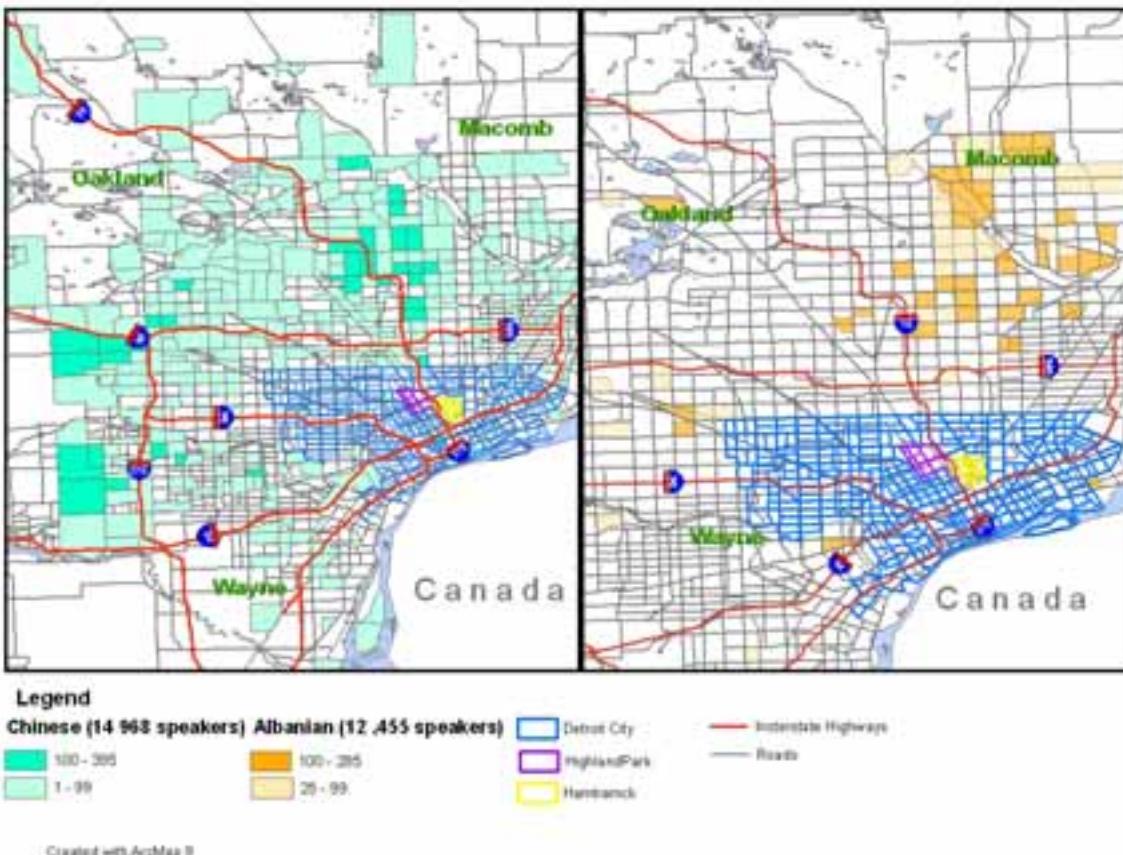


Polish can be easily described as a ‘low density’ language: its speakers have settled all over the three counties, with a clear dis-preference for Detroit proper. The high density areas of Polish speakers are, on the whole, very few. Syriac speakers, on the other hand, form two clearly delimited areas in Oakland and Macomb counties. Moreover, the number of speakers in a dense Syriac 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 1930’s; Syriac immigration is a relatively recent phenomenon, since late 1980’s and onwards.

A similar example can be shown by comparing settling patterns for Chinese and Albanian speakers, shown on Map 6.

Map 6. Chinese and Albanian in Metro Detroit



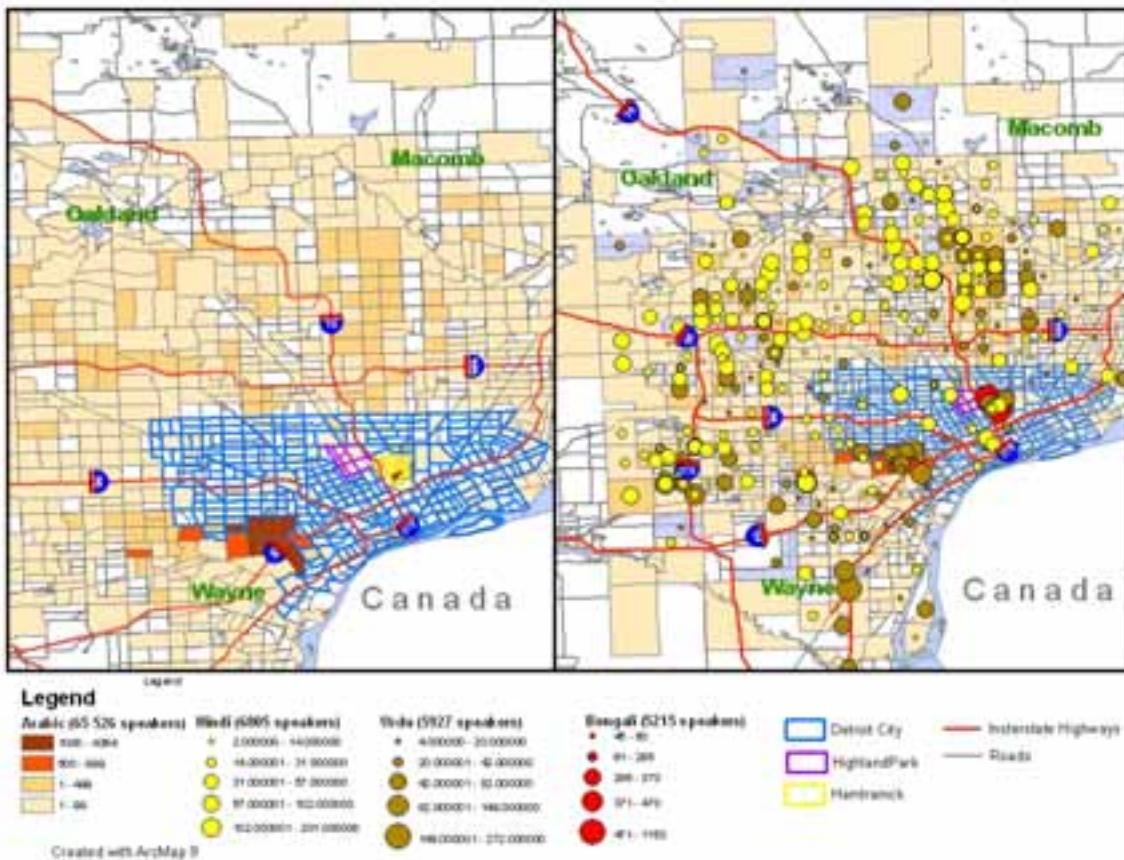
Similarly to Polish and Syriac, the Chinese and Albanian speakers who live in the Detroit Metro area are relatively close in numbers: around 15,000 Chinese speakers and 13,000 Albanian speakers. However, their settling patterns are quite different. As shown on the map above, Chinese speakers tend to live relatively spread all over the area. Albanian speakers appear to live exclusively in Macomb county and that in much higher density tracts than Chinese speakers. Again, between the two groups, Albanians represent the one of more recent arrival in the US, 1990's and onwards.

Another motivation for different settling patterns is cited by Ratliff (2005): 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 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 US as refugees and still tend to live close together. We shall have occasion to return to Hmong Miao in section 3.4.

3.4. Languages which pattern with other languages

There are languages which show close patterns with to other languages. However, when choosing a location, similarity of religion seems to be more important than closeness of ethnicity, and thereby language. Consider the settling patterns of Arabic, Bengali and Urdu speakers shown on Map 7.

Map 7. Arabic, Urdu and Bengali vs. Hindi



As we recall, Arabic is the third largest language in the Detroit Metropolitan Area. The highest concentration of speakers is in Dearborn, a suburb of Detroit, together with nearby areas as well as Hamtramck. However, relatively dense tracts (100-499 speakers) are observed all over the 3 counties.

A closer look at the distribution of Urdu speakers shows that they tend to settle in dense Arabic areas rather than with the closely related Hindi. This can be easily explained by the fact that both Arabic and Urdu speakers tend to be Muslim.

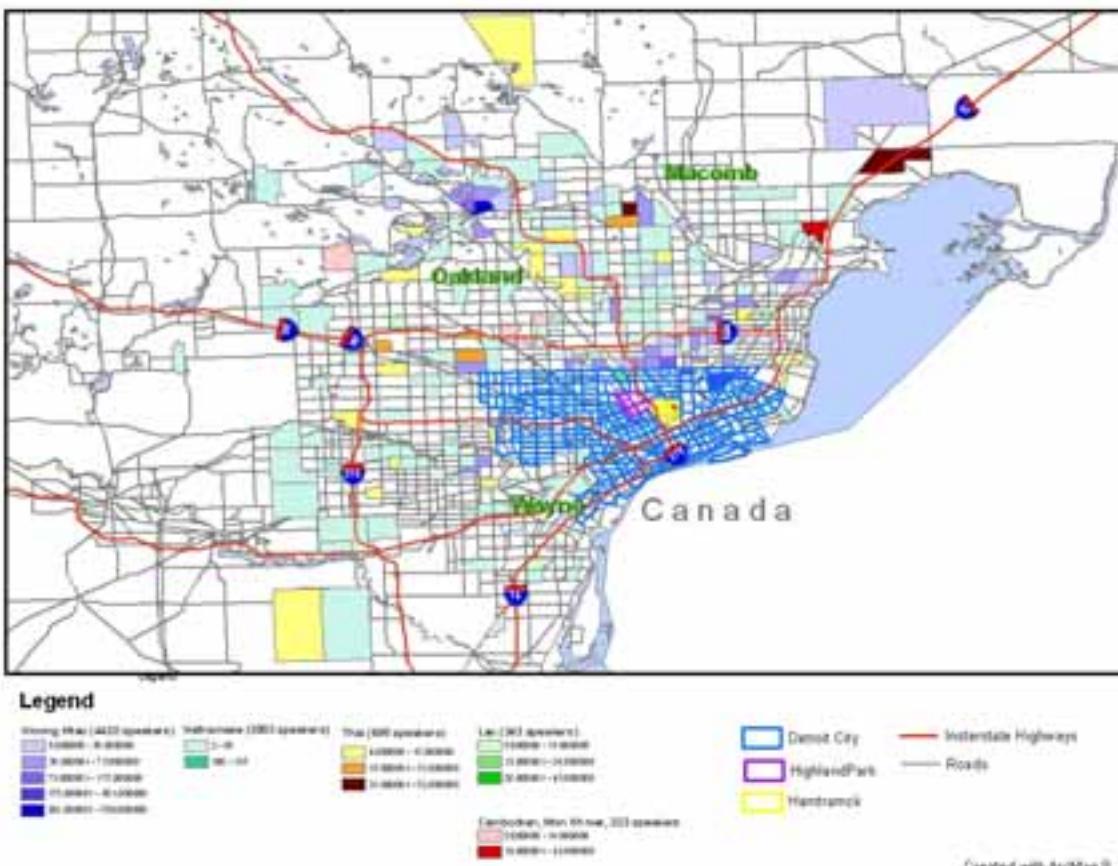
The Bengali are ethnically and linguistically completely separate from both Arabic and Urdu speakers. Yet, they are also Muslim and this appears to play a decisive role when choosing a place to live in a new situation. In fact, most of the Bengali (3530 out of 5215, that is 67%) appears to live in Hamtramck where a number of Arabic and Urdu speakers live as well.

Similarity of religion as a unifying fact in this situation is something that deserves further exploration. At this point, we would also like to bring up the case of South East Asian languages. It is not as convincing as the Arabic-Urdu-Bengali case discussed above but still it presents interesting evidence for our hypothesis.

As already mentioned above, Ratliff (2005) points out that in the state of Michigan, Hmong Miao speakers tend to form close communities while speakers of other South Asian languages tend to settle down in a more scattered fashion. As shown on Map 8, Ratliff's observations are generally valid for the South Asian groups in the Detroit

Metropolitan area as well. However, a spatial analysis of the data brings up additional generalizations.

Map 8. South East Asian Languages in Metro Detroit



Specifically, although each group appears to be scattered all over the three counties, they still seem to come fairly close together and form intra-group communities. For instance, there are several instances on the map above, where Thai and Vietnamese speakers settle in the immediate vicinity of each other. Similarly, Cambodian, Vietnamese and again Thai speakers appear close together. Whether this is due to a shared religion or to a generally close geographical origin remains to be further specified. However, the patterns are recurrent and thus interesting enough for future investigation.

4. Conclusions and directions for future research

In this paper we used census data together with GIS analysis to outline the linguistic profile of the Detroit Metropolitan Area. Apart from identifying the 89 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. Furthermore, we found interesting spatial co-occurrence for completely unrelated languages. We attribute this to shared religion; however, this issue requires more evidence and deserves further investigation. Apart from

religious motives, another motive for different ethnic groups to settle close together in a new country appears to be proximity in the place of origin.

Because of time and space restrictions, there was no room to show correlations with socio-economic variables such as homeownership, medium income, availability of various services in the areas of residence for different language speakers. However, such correlations are an integral part of a study of this kind and more such correlations should be performed as this inquiry develops.

Due to limitations of census data and a rather time consuming process of properly identifying the languages spoken in the area, we could only map languages on household level at this stage. Pending response from the Michigan Board of Education, language use for instruction at schools will be mapped as well. As the project develops, at least the main languages of work situations should be identified.

On the whole, combining census data on language diversity with geographical analysis has proven to be a rewarding and a highly informative enterprise. While this study has only scratched the surface of the potential such a combination has to offer, it has also shown that spatial analysis is crucial to the study of human diversity and is probably the only appropriate way to study multilingualism in context.

References

- Danner, Marcia and Patricia Banker Peart (eds.). 1999. *Global Journeys in Metro Detroit*. Detroit: New Detroit Inc.
- Paulston, Christina Bratt and Richard G. Tucker. 2003. *Sociolinguistics: the Essential Readings*. Malden, MA: Blackwell Pub.
- Ratliff, Martha. 2005. Southeast Asian languages in Michigan and Ohio. In: Joseph, Brian, Preston, Carol G. and Dennis R. Preston. (eds.) *Language Diversity in Michigan and Ohio*. Ann Arbor: Caravan Books.
- Trudgill, Peter. 1995. *Sociolinguistics: An Introduction to Language and Society*. London: Penguin.
- Williams, Colin H. and Izak Van der Merwe. 1996. Mapping the Multilingual City: A Research Agenda for Urban Geolinguistics. *Journal of Multilingual and Multicultural Development* 17: 1, pp. 49-66.