

## **Geographic Clustering of Language Spoken**

Meetings of the Language Working Groups of the  
Race and Ethnic Advisory Committee (April 27 - 28, 2006)  
and the 2010 Census Advisory Committee (May 11 - 12, 2006)

Deborah Griffin and Pamela McGovern  
U.S. Census Bureau

---

This draft document is being provided to the Census Bureau's Advisory Committees prior to upcoming meetings. It is preliminary in nature and in the early stages of development. As such, it is subject to revision. Our intent in making this working document available at this time is to inform ongoing discussions related to 2010 planning and the American Community Survey.

**U S C E N S U S B U R E A U**

H:\language\Geographic Clustering of Language Spoken

## 1. Introduction

In developing specific ACS and 2010 Census tools to provide assistance to households with limited English proficiency, it is important to understand the geographic distribution of language needs. Census 2000 data found that nearly 1-in-5 people, or 47 million U.S. residents age 5 and older, spoke a language other than English at home in 2000.<sup>1</sup> Analysis of Census 2000 data included the identification of areas with the highest percentage of people (age 5 or older) who spoke a language other than English at home. Thematic graphs, available at American Factfinder, display this information at county and other sub-state levels.

Spanish is the predominant non-English language spoken in the United States. In Census 2000 we estimated that over 28 million of the nearly 47 million speakers of a language other than English (about 60 percent) spoke Spanish.<sup>2</sup> The remaining languages (including, Chinese, French, German, Tagalog, Vietnamese, Italian, Korean, Russian, Polish, and Arabic) make up a relatively small proportion of language speakers. Chinese, for example, represents only about 4 percent of the people who spoke a language other than English at home in 2000. This makes the available data on geographic distributions largely a reflection of the clustering of Spanish-speakers.

This paper explores the geographic distribution (and clustering) of language spoken by looking separately at data for people who speak Spanish, Chinese, Korean, Vietnamese, and Russian. In contrast to earlier analyses that focused on population, these data are studied at the household level. This analysis is needed to understand if sufficient language clustering occurs for a given language, thus suggesting the potential to identify areas for specific language tools. These five languages were chosen because Census 2000 data showed they had the greatest numbers of linguistically isolated households.<sup>3</sup> We assume these groups would therefore have the greatest difficulty completing a English questionnaire. A different set of languages could be considered. Identification of these five languages does not suggest that they are the only language groups being considered.

A series of questions are included in section 5. We would appreciate the working group's feedback on these specific issues as well as any other general comments on the paper.

## 2. Background

---

<sup>1</sup>Shin, Hyon B. and Bruno, Rosalind. 2003. "Language Use and English-Speaking Ability: 2000". U.S. Census Bureau. Census 2000 Brief.

<sup>2</sup>Shin, Hyon B. and Bruno, Rosalind. 2003. "Language Use and English-Speaking Ability: 2000". U.S. Census Bureau. Census 2000 Brief.

<sup>3</sup> A linguistically isolated household is one in which no household member age 14 or older reported speaking English at least "very well".

The ACS mails questionnaires in English to all sample addresses in the United States and questionnaires in Spanish to all sample addresses in Puerto Rico. Spanish assistance (including Spanish questionnaires) is available upon request in the United States. English assistance (including English questionnaires) is available upon request in Puerto Rico. Telephone and personal visit follow-up activities are conducted in many languages. If language needs were found to be geographically clustered, the ACS could consider providing additional language materials by mail to sample addresses in specific areas.

The 2010 Census tested mailing a bilingual Spanish/English questionnaire in the 2005 National Content Test and is currently evaluating alternative mailing strategies for the bilingual form in the 2010 Census. As with the ACS, the 2010 Census will conduct follow-up activities in many languages. We plan to mail a Spanish/English bilingual questionnaire in the 2010 Census but have not finalized the criteria for selecting the areas which will receive the bilingual form as part of the initial mailing. We are currently assessing options for questionnaires and language guides in other non-English languages. This examination of geographic clustering is relevant for the planning and implementation of the 2010 Census.

### **3. Methodology**

#### **3.1 Sources**

The Census 2000 Sample Edited Data File (SEDF), containing Census 2000 long form sample data, was used in this analysis. All occupied tracts were eligible for inclusion in this study, regardless of size. All data used to produce the tables in this paper were edited and weighted using the person weight of the householder. No sampling errors were calculated for these estimates. A similar analysis will be possible using multi-year estimates of language spoken at home from the ACS.

The tables in this paper were produced at the tract level to assess if tracts could be identified with concentrations of households that speak Spanish, Chinese, Korean, Russian, and Vietnamese.<sup>4</sup> Comparable tables were also produced at block group levels. The block group data are based on relatively small sample sizes and the estimates are therefore less reliable. The general conclusions reached from the block group data parallel these tract-level findings

#### **3.2 Measuring Geographic Clustering**

---

<sup>4</sup>Tracts are small, relatively permanent statistical subdivisions of a county or statistically equivalent entity delineated by local participants as part of the U.S. Census Bureau's Participant Statistical Areas Program. Census tracts generally have between 1500 and 8000 people. A block group is a cluster of census blocks within a census tract. Block groups generally contain between 600 and 3000 people.

For each tract two summaries of geographic clustering are presented. These data are displayed in the first three data columns of Tables 1 through 5 (Attachment 1). First, a determination was made of the proportion of households in each tract that reported speaking Spanish, Chinese, Korean, Russian, and Vietnamese. Second, a similar analysis was undertaken restricting the universe to linguistically isolated households. Tables 1 through 5 display these results for each of these five language groups. Information is summarized on the number and the percentage of all tracts with any households speaking each of these languages. For example, Table 1 shows that of the 64,956 tracts in the United States, Spanish-speakers were found in 98.2 percent of those tracts. Table 1 also shows that in 60.8 percent of the tracts Spanish speakers were found who were determined to be “linguistically isolated.”

This series of tables also identifies 8 strata based on the concentration of language speakers (and linguistically isolated language speakers) in each tract. The first stratum includes all tracts in which 70 percent or more of the households in the tract reported speaking a specific language (or speaking a specific language and also are linguistically isolated). Table 1 shows that the proportion of Spanish-speakers in a tract was less than 10 percent in 47,228 (72.7 percent) of all tracts. In 1414 tracts over 70 percent of the households in the tract reported speaking Spanish at home.

A cumulative percent column sums these percentages starting with the strata with the greatest concentrations. This column allows you to estimate the percent of all tracts with, for example, a minimum of 50 percent of the population speaking a specific language. In Table 1 that value would be 4.7 percent.

### **3.3 Measuring the Distribution of Households Speaking a Language Other Than English at Home**

To help interpret these data, the total number of Spanish-speaking, Chinese-speaking, Korean-speaking, Russian-speaking, and Vietnamese-speaking households were determined for each tract. These data are displayed in the last three data columns of the tables. Tables 1 through 5 distribute these households across the 8 geographic clustering strata. For example, Table 1 shows that in Census 2000 a total of 10,771,134 households reported speaking Spanish at home. About 1,602,540 (14.9 percent) of these households lived in tracts with the highest concentration of Spanish-speakers (70 percent or greater). At the opposite end, 28.8 percent of all Spanish-speaking households lived in tracts in which less than 10 percent of all households reporting that they speak Spanish at home.

The cumulative percent column cumulates these proportions starting with the strata with the highest concentration. This provides an estimate of the proportion of all Spanish-speaking households that would be impacted if methods were developed to select tracts with a minimum concentration of Spanish-speakers. For example, if a threshold was set

of selecting tracts with 40 percent or more of the households speaking Spanish, such

criteria would include 35.4 percent of all Spanish-speaking households.

Similar data are presented for linguistically isolated households. For example, Table 1 shows that Census 2000 identified over 2.5 million Spanish speaking households who were also linguistically isolated. About 2339 linguistically isolated households fall in the 9 tracts with 70 percent of more of the households in the tracts reporting to speak Spanish at home and to be linguistically isolated.

When used in combination, the tables can provide information on the number of tracts that would need to be selected in order to impact a certain proportion of households that spoke a language other than English or were linguistically isolated. For example, referring to Table 2, a plan that selected all tracts in which more than 10 percent of the households reported speaking Chinese would involve about 1.3 percent of all tracts and would reach about 32.1 percent of all Chinese-speaking households. A plan that selected all tracts with more than 10 percent of the total households in the tract speaking Chinese and being linguistically isolated would involve less than 1 percent of all tracts and would reach about 25.5 percent of all Chinese linguistically isolated households.

### **3.4 Limitations**

There are several limitations that should be considered in reviewing these data. In an effort to isolate individual language needs, this analysis does not identify the tracts with multiple language needs. It should be assumed that some tracts could qualify as having more than 10 percent of the households speaking more than one language. The development of any selection criteria would need to consider this factor.

As noted earlier, sampling errors were not produced for this analysis. The allocation of tracts to strata is imperfect in that it uses the estimates without consideration of these sampling errors.

Finally, all occupied tracts were analyzed, regardless of size. Future work should consider establishing a minimum tract size to avoid identification of very small tracts as the tracts with high proportions of language-speakers.

## **4. Analysis**

Our initial analysis focused on census tracts. Because tracts average about 4000 people we did not identify many tracts in which more than 10 percent of the households reported speaking Chinese, Korean, Vietnamese, or Russian. Similarly, the data for linguistically isolated households identified a limited number of tracts with more than 10 percent of the households in the tract speaking this language and being linguistically isolated.

We explored block group data because block groups are smaller - averaging about 1500 people. The results, however, are quite similar. Using either block group or tract-level data to identify

areas with a minimum threshold of 10 percent unfortunately does not cover a significant proportion of households that spoke a language other than English or Spanish. If a 10 percent threshold were established for language-speakers, the highest coverage of Chinese-speakers is only 32.1 percent. The levels for Korean-speakers is 13.7 percent, Vietnamese-speakers, 14.6 percent, and Russian-speakers, 24.3 percent.

If linguistic isolation is taken into account with a similar 10 percent threshold, the coverage of linguistically isolated households would be 25.5 percent (Chinese), 12.4 percent (Korean), 10.1 percent (Vietnamese), and 27.7 percent (Russian). This indicates that selection criteria based on geographic clustering would not provide language assistance to most Chinese, Korean, Vietnamese, or Russian speakers. Additional methods would be needed to supplement this approach as a majority of households with language needs would not be included in the areas identified as having even 10 percent clustering.

## 5. Questions

- a. The Census Bureau's analysis (Griffin, McGovern) includes summaries based on all households who report speaking a language other than English at home and on "linguistically isolated households". We have considered the prevalence of persons who speak a language other than English and who do not speak English "at all" or "not well" (i.e., linguistically isolated) in the 2006 census test planning and distribution of language guides. Which of these options do you think best in our development of criteria for selecting areas for mailing a 2010 bilingual questionnaire?
- b. The data for Spanish-speakers is very different from that for the other languages. What do the Spanish language data suggest to you about the viability of identifying census tracts to receive a Spanish (or a bilingual) mailing?
3. The data for Chinese, Korean, Russian, and Vietnamese households suggest limited geographic clustering. What do these data suggest to you about providing assistance and materials in these languages?

<b>Table 1. Geographic clustering of Spanish-speaking households</b>						
<b>Spanish-speaking</b>	<b>Number of Tracts</b>	<b>Percent of Total Tracts</b>	<b>Cumulative Percent</b>	<b>Number of Households</b>	<b>Percent of Households</b>	<b>Cumulative Percent</b>
Total in the Nation	64,956	100.0	---	---	---	---
<b>Total with Spanish-speakers</b>	63,782	98.2	---	10,771,134	100.0	---
<b>Percent of households in the tract who speak Spanish</b>						
70 percent or greater	1,414	2.2	2.2	1,602,540	14.9	14.9
60 - 69.9 percent	731	1.1	3.3	669,915	6.2	21.1
50 - 59.9 percent	905	1.4	4.7	715,841	6.6	27.7
40 - 49.9 percent	1,201	1.8	6.5	828,816	7.7	35.4
30 - 39.9 percent	1,739	2.7	9.2	922,012	8.6	44.0
20 - 29.9 percent	2,943	4.5	13.7	1,138,867	10.6	54.6
10 - 19.9 percent	7,621	11.7	25.3	1,786,296	16.6	71.2
Less than 10 percent	47,228	72.7	98.2	3,106,847	28.8	100.0
<b>Linguistically isolated Spanish-speaking</b>	<b>Number of Tracts</b>	<b>Percent of Total Tracts</b>	<b>Cumulative Percent</b>	<b>Number of Households</b>	<b>Percent of Households</b>	<b>Cumulative Percent</b>
Total in the Nation	64,956	100.0	---	---	---	---
<b>Total with linguistically isolated Spanish-speakers</b>	39,484	60.8	---	2,571,597	100.0	---
<b>Percent of households in the tract who speak Spanish and are linguistically isolated</b>						
70 percent or greater	9	<0.1	<0.1	2,339	0.1	0.1
60 - 69.9 percent	13	<0.1	<0.1	12,882	0.5	0.6
50 - 59.9 percent	50	0.1	0.1	48,512	1.9	2.5
40 - 49.9 percent	174	0.3	0.4	101,820	4.0	6.4
30 - 39.9 percent	509	0.8	1.2	236,548	9.2	15.6
20 - 29.9 percent	1,254	1.9	3.1	428,141	16.6	32.3
10 - 19.9 percent	2,863	4.4	7.5	604,682	23.5	55.8
Less than 10 percent	34,612	53.3	60.8	1,136,673	44.2	100.0

Source: U.S. Census Bureau. Census 2000 - SEDF

<b>Table 2. Geographic clustering of Chinese-speaking households</b>						
<b>Chinese-speaking</b>	<b>Number of Tracts</b>	<b>Percent of Total Tracts</b>	<b>Cumulative Percent</b>	<b>Number of Households</b>	<b>Percent of Households</b>	<b>Cumulative Percent</b>
Total in the Nation	64,956	100.0	---	---	---	---
<b>Total with Chinese-speakers</b>	24,227	37.3	---	780,316	100.0	---
<b>Percent of households in the tract who speak Chinese</b>						
70 percent or greater	12	<0.1	<0.1	17,649	2.3	2.3
60 - 69.9 percent	5	<0.1	<0.1	4,282	0.5	2.8
50 - 59.9 percent	19	<0.1	0.1	12,906	1.7	4.5
40 - 49.9 percent	25	<0.1	0.1	12,395	1.6	6.1
30 - 39.9 percent	74	0.1	0.2	34,841	4.5	10.5
20 - 29.9 percent	196	0.3	0.5	67,513	8.7	19.2
10 - 19.9 percent	492	0.8	1.3	100,870	12.9	32.1
Less than 10 percent	23,404	36.0	37.3	529,860	67.9	100.0
<b>Linguistically isolated Chinese-speaking</b>	<b>Number of Tracts</b>	<b>Percent of Total Tracts</b>	<b>Cumulative Percent</b>	<b>Number of Households</b>	<b>Percent of Households</b>	<b>Cumulative Percent</b>
Total in the Nation	64,956	100.0	---	---	---	---
<b>Total with linguistically isolated Chinese-speakers</b>	12,324	19.0	---	294,374	100.0	---
<b>Percent of households in the tract who speak Chinese and are linguistically isolated</b>						
70 percent or greater	1	<0.1	<0.1	1,142	0.4	0.4
60 - 69.9 percent	3	<0.1	<0.1	2,748	0.9	1.3
50 - 59.9 percent	7	<0.1	<0.1	7,535	2.6	3.9
40 - 49.9 percent	12	<0.1	<0.1	6,888	2.3	6.2
30 - 39.9 percent	17	<0.1	0.1	6,649	2.3	8.5
20 - 29.9 percent	42	0.1	0.1	13,578	4.6	13.1
10 - 19.9 percent	203	0.3	0.4	36,382	12.4	25.5
Less than 10 percent	12,039	18.5	19.0	219,452	74.6	100.0

Source: U.S. Census Bureau. Census 2000 - SEDF

<b>Table 3. Geographic clustering of Korean-speaking households</b>						
<b>Korean-speaking</b>	<b>Number of Tracts</b>	<b>Percent of Total Tracts</b>	<b>Cumulative Percent</b>	<b>Number of Households</b>	<b>Percent of Households</b>	<b>Cumulative Percent</b>
Total in the Nation	64,956	100.0	---	---	---	---
<b>Total with Korean-speakers</b>	18,727	28.8	---	375,704	100.0	---
<b>Percent of households in the tract who speak Korean</b>						
70 percent or greater	0	0.0	0.0	39	<0.1	<0.1
60 - 69.9 percent	0	0.0	0.0	0	0.0	<0.1
50 - 59.9 percent	0	0.0	0.0	0	0.0	<0.1
40 - 49.9 percent	5	<0.1	<0.1	4,178	1.1	1.1
30 - 39.9 percent	16	<0.1	<0.1	8,499	2.3	3.4
20 - 29.9 percent	27	<0.1	0.1	9,750	2.6	6.0
10 - 19.9 percent	144	0.2	0.3	29,127	7.8	13.7
Less than 10 percent	18,533	28.5	28.8	324,111	86.3	100.0
<b>Linguistically isolated Korean-speaking</b>	<b>Number of Tracts</b>	<b>Percent of Total Tracts</b>	<b>Cumulative Percent</b>	<b>Number of Households</b>	<b>Percent of Households</b>	<b>Cumulative Percent</b>
Total in the Nation	64,956	100.0	---	---	---	---
<b>Total with linguistically isolated Korean-speakers</b>	7,848	12.1	---	141,567	100.0	---
<b>Percent of households in the tract who speak Korean and are linguistically isolated</b>						
70 percent or greater	1	<0.1	<0.1	33	<0.1	<0.1
60 - 69.9 percent	0	0.0	<0.1	0	0.0	<0.1
50 - 59.9 percent	0	0.0	<0.1	0	0.0	<0.1
40 - 49.9 percent	1	<0.1	<0.1	266	0.2	0.2
30 - 39.9 percent	3	<0.1	<0.1	1,381	1.0	1.2
20 - 29.9 percent	16	<0.1	<0.1	5,934	4.2	5.4
10 - 19.9 percent	51	0.1	0.1	9,955	7.0	12.4
Less than 10 percent	7,776	12.0	12.1	123,998	87.6	100.0

Source: U.S. Census Bureau. Census 2000 - SEDF

<b>Table 4. Geographic clustering of Russian-speaking households</b>						
<b>Russian-speaking</b>	<b>Number of Tracts</b>	<b>Percent of Total Tracts</b>	<b>Cumulative Percent</b>	<b>Number of Households</b>	<b>Percent of Households</b>	<b>Cumulative Percent</b>
Total in the Nation	64,956	100.0	---	---	---	---
<b>Total with Russian-speakers</b>	15,137	23.3	---	323,654	100.0	---
<b>Percent of households in the tract who speak Russian</b>						
70 percent or greater	0	0.0	0.0	0	0.0	0.0
60 - 69.9 percent	3	<0.1	<0.1	3,690	1.1	1.1
50 - 59.9 percent	2	<0.1	<0.1	2,066	0.6	1.8
40 - 49.9 percent	9	<0.1	<0.1	7,997	2.5	4.2
30 - 39.9 percent	31	<0.1	0.1	17,216	5.3	9.6
20 - 29.9 percent	58	0.1	0.2	19,688	6.1	15.7
10 - 19.9 percent	114	0.2	0.3	28,009	8.7	24.3
Less than 10 percent	14,920	23.0	23.3	244,988	75.7	100.0
<b>Linguistically isolated Russian-speaking</b>	<b>Number of Tracts</b>	<b>Percent of Total Tracts</b>	<b>Cumulative Percent</b>	<b>Number of Households</b>	<b>Percent of Households</b>	<b>Cumulative Percent</b>
Total in the Nation	64,956	100.0	---	---	---	---
<b>Total with linguistically isolated Russian-speakers</b>	5,411	8.3	---	137,319	100.0	---
<b>Percent of households in the tract who speak Russian and are linguistically isolated</b>						
70 percent or greater	0	0.0	0.0	0	0.0	0.0
60 - 69.9 percent	0	0.0	0.0	0	0.0	0.0
50 - 59.9 percent	1	<0.1	<0.1	1,259	0.9	0.9
40 - 49.9 percent	1	<0.1	<0.1	791	0.6	1.5
30 - 39.9 percent	4	<0.1	<0.1	3,177	2.3	3.8
20 - 29.9 percent	29	<0.1	0.1	11,576	8.4	12.2
10 - 19.9 percent	87	0.1	0.2	21,197	15.4	27.7
Less than 10 percent	5,289	8.1	8.3	99,319	72.3	100.0

Source: U.S. Census Bureau. Census 2000 - SEDF

<b>Table 5. Geographic clustering of Vietnamese-speaking households</b>						
<b>Vietnamese-speaking</b>	<b>Number of Tracts</b>	<b>Percent of Total Tracts</b>	<b>Cumulative Percent</b>	<b>Number of Households</b>	<b>Percent of Households</b>	<b>Cumulative Percent</b>
Total in the Nation	64,956	100.0	---	---	---	---
<b>Total with Vietnamese-speakers</b>	16,176	24.9	---	320,265	100.0	---
<b>Percent of households in the tract who speak Vietnamese</b>						
70 percent or greater	0	0.0	0.0	0	0.0	0.0
60 - 69.9 percent	0	0.0	0.0	0	0.0	0.0
50 - 59.9 percent	1	<0.1	<0.1	707	0.2	0.2
40 - 49.9 percent	3	<0.1	<0.1	1,684	0.5	0.7
30 - 39.9 percent	18	<0.1	<0.1	8,510	2.7	3.4
20 - 29.9 percent	34	0.1	0.1	10,468	3.3	6.7
10 - 19.9 percent	139	0.2	0.3	25,528	8.0	14.6
Less than 10 percent	15,981	24.6	24.9	273,368	85.4	100.0
<b>Linguistically isolated Vietnamese-speaking</b>	<b>Number of Tracts</b>	<b>Percent of Total Tracts</b>	<b>Cumulative Percent</b>	<b>Number of Households</b>	<b>Percent of Households</b>	<b>Cumulative Percent</b>
Total in the Nation	64,956	100.0	---	---	---	---
<b>Total with linguistically isolated Vietnamese-speakers</b>	8,182	12.6	---	139,804	100.0	---
<b>Percent of households in the tract who speak Vietnamese and are linguistically isolated</b>						
70 percent or greater	0	0.0	0.0	0	0.0	0.0
60 - 69.9 percent	0	0.0	0.0	0	0.0	0.0
50 - 59.9 percent	0	0.0	0.0	0	0.0	0.0
40 - 49.9 percent	0	0.0	0.0	0	0.0	0.0
30 - 39.9 percent	1	<0.1	<0.1	168	0.1	0.1
20 - 29.9 percent	11	<0.1	<0.1	3,051	2.2	2.3
10 - 19.9 percent	58	0.1	0.1	10,859	7.8	10.1
Less than 10 percent	8,112	12.5	12.6	125,726	89.9	100.0

Source: U.S. Census Bureau. Census 2000 - SEDF

