

Language Projections: 2010 to 2020

Presented at the Federal Forecasters Conference, Washington, DC, April 21, 2011
Hyon B. Shin, Social, Economic, and Housing Statistics Division, U.S. Census Bureau
Jennifer M. Ortman, Population Division, U.S. Census Bureau

This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

ABSTRACT

Language diversity in the United States has changed rapidly over the past three decades. The use of a language other than English at home increased by 148 percent between 1980 and 2009 and this increase was not evenly distributed among languages. Polish, German, and Italian actually had fewer speakers in 2009 compared to 1980. Other languages, such as Spanish, Vietnamese, and Russian, had considerable increases in their use. Using data on the language spoken at home from the American Community Survey and the U.S. Census Bureau's 2008 and 2009 National Population Projections, this paper presents projections of what the population speaking a language other than English might look like in 2020, with a focus on the methodology used to produce these projections.

INTRODUCTION

The changing landscape of the population living in the United States over the past several decades can be seen in many areas throughout the country. Whether it is a road sign written in Chinese or a Spanish-language television station, one can see that the language diversity in the United States is rapidly changing. In 2009, 57.1 million people (20 percent of the population 5 years and older) spoke a language other than English (LOTE) at home. In 1980, there were 23.1 million (11 percent of the population 5 years and older) LOTE speakers (Table 1).

The overall 148 percent increase from 1980 to 2009 in the number of LOTE speakers was not evenly distributed among languages. Polish, German, and Italian actually had fewer speakers in 2009 compared to 1980. Other languages, such as Spanish, Vietnamese, and Russian, however, had considerable increases in their use. This paper presents national-level projections of what the LOTE population might look like in 2020, with a focus on the methodology that is used to produce these projections.

BACKGROUND

The United States has always been a country noted for its linguistic diversity. Information on language use and proficiency collected from decennial censuses shows that there have been striking changes in the linguistic landscape. These changes have been driven in large part by a shift in the origins of immigration to the United States. During the late 19th and early 20th centuries, the majority of U.S. immigrants spoke either English or a European language such as German, Polish, or Italian (Stevens, 1999). Beginning in the middle of the 20th century, patterns of immigration shifted to countries in Latin America, the Caribbean, and Asia (Bean and Stevens, 2005). As a result, the use of Spanish and Asian or Pacific Island languages began to grow. By 2000, over 70 percent of the population speaking a LOTE spoke Spanish, Chinese, Japanese, Korean, Vietnamese, or Tagalog (Shin and Bruno, 2003).

Since 1980, the percentage of the population who reported speaking a language other than English at home rose from 23.1 million speakers to 57.1 million speakers in 2009 (Table 2). The largest numeric increase in the population speaking a language other than English at home was for Spanish speakers (increased by 24.4 million speakers) whereas the largest percent increase was for Vietnamese speakers (533 percent increase).

Language use is an indicator of cultural assimilation (Rumbaut, 1997), which is measured by shifts to English as the language usually spoken by U.S. immigrants and their descendants (Stevens, 1994). For most U.S. immigrant groups, the shift to English monolingualism takes place within a few generations (Hurtado and Vega, 2004).

There are many incentives to learn and use English in the American society. Economists have argued that the impetus for language acquisition was for human capital (Chiswick and Miller, 2001) or that potential earnings could be affected by not having a strong command of the English language and, therefore, motivate immigrants to learn English and increase potential earnings (Cohen-Goldner and Eckstein, 2008). Others have argued that the economic view overlooks the social and cultural aspects of learning English in the United

States (Espenshade and Fu, 1997; Mouw and Xie, 1999; Stevens, 1992) such as communication within and outside of one's language group.

The U.S. Census Bureau has collected information about the language characteristics of U.S. residents in every decennial census from 1890 through 2000, with the exception of the 1950 census. Information was collected on English proficiency, mother tongue, and language spoken. The development of a consistent time series of data for the period between 1890 and 1980 is hindered by the considerable variation across censuses in terms of question wording, coding of responses, and the subsets of the population that were asked these questions (Stevens, 1999).

Beginning in 1980, a series of three questions were introduced to gather data on language use and English speaking ability. These questions were developed to satisfy the legislative mandate of the minority language assistance provision of Section 203 in the Voting Rights Act of 1965 and, along with a few other variables, are used to determine which jurisdictions must provide voting rights materials in minority languages.¹ The three questions were asked of the population 5 years and over. The first question asked "Does this person speak a language other than English at home?" If the respondent answered "Yes" to this question, they were then asked "What is this language?" with a write-in field for the answer and then asked "How well does this person speak English?" with the following four answer categories: "Very well," "Well," "Not well," and "Not at all."

These same three questions were asked in the 1980, 1990, and 2000 censuses, providing a consistent time series with which to study changes in language use and English-speaking ability among U.S. residents over time. Since 2001, the language questions, along with all of the other social, economic, and housing questions that were asked in the Census 2000 long-form census questionnaire, are now asked yearly in the American Community Survey. This change allows for these characteristics to be gathered yearly instead of every 10 years. Having the same three questions asked for the last 3 decades gives a good metric for comparing the relative growth or decline of individual languages.

¹ For more information on the Voting Rights Act and how the language questions are used to satisfy the legislative mandate, see the Federal Register at <http://www.census.gov/rdo/pdf/FRN_VotingRightsDeterminations.pdf>.

The language data collected are obtained from the second language question that asks "What is this language?" The languages written in this box are put through a coding procedure that assigns a language code for individual languages or groups of languages. There are 382 language codes and from this list, a standard classification of 39 detailed language groups is available. These 39 languages are further collapsed into four major language groups; Spanish, Other Indo-European languages, Asian and Pacific Island languages, and all other languages. Table 1 shows data from the 2009 American Community Survey for the four- and 39-language groups by English-speaking ability.

DATA AND METHODS

This paper presents a series of national-level language projections developed using data on the language spoken at home from the American Community Survey and the Census Bureau's 2008 and 2009 National Population Projections. The paper discusses the language-projection results using the 2008 National Population Projections numbers only. The results using the 2009 projections are available upon request.

American Community Survey Data

The American Community Survey (ACS) collects data on social, housing, and economic characteristics for demographic groups in the United States. This paper uses the 2006, 2007, 2008, and 2009 ACS files.

Data on language use and English-speaking ability historically collected in the decennial censuses, are now captured every year in the ACS. The ACS was conducted on a test basis from 2000 through 2004 and expanded to full sample size for housing units in 2005 and for group quarters in 2006. To have a complete sample, comparable to Census 2000, we chose to use the ACS data files from 2006 through 2009.²

National Population Projections Data

The U.S. Census Bureau's 2008 and 2009 National Population Projections were created using the cohort-component method and provide projections of the resident population of the United States and demographic components of change (births, deaths, and

² For more information on the ACS, the American Community Survey website provides handbooks for data users. These handbooks are available online at <http://www.census.gov/acs/www/guidance_for_data_users/handbooks/>.

net international migration).³ These projections are based on Census 2000 data. These data are provided by age, sex, race and Hispanic origin for each year from July 1, 2000 to July 1, 2050. The projection series released in 2009 provided four supplemental series of projections with results for different international migration assumptions. The supplemental series included: (1) high migration, (2) low migration, (3) constant migration, and (4) zero migration. Assumptions about future rates of mortality and fertility are the same in all five series. This paper uses data for the years 2010 through 2020 from the 2008 series (U.S. Census Bureau, 2008) and the high, low, and constant series from the 2009 release (U.S. Census Bureau, 2009).

Language Projection Methodology

We produce projections of both the total number of people speaking a language other than English at home (LOTE speakers) and the number of speakers for individual languages with at least 500,000 speakers in 2009. The 13 languages that meet this condition are: Spanish, French, Italian, Portuguese, German, Russian, Polish, Hindi, Chinese, Korean, Vietnamese, Tagalog, and Arabic. These are the most commonly spoken non-English languages and for some, such as Vietnamese and Russian, there has been tremendous growth in the number of speakers in the last few decades. The projections are produced by projecting future LOTE use based on trends in the ACS data and then applying the projected distribution of LOTE speakers to the projected population from the Census Bureau's 2008 and 2009 National Population Projections.

The distributions of LOTE speakers are projected by demographic characteristics. For projections of the overall population speaking a LOTE and the population speaking Spanish, we project by age (single years 5-49 and 50 years and over) and Hispanic origin, resulting in a total of 92 groups for which we project the percent speaking a LOTE and Spanish. Projections of the individual languages other than Spanish are developed by age, resulting in a total of 46 groups for which we project the percent speaking other individual languages.

We have developed three series of language projections, based on assumptions of constant, linear, and logistic change. The first assumption we make is the most basic and simplistic. We held LOTE use constant at currently

observed levels. To do this, we held the percentage of LOTE speakers constant for each age and Hispanic origin group we project for at the level reported in the 2009 ACS. This is represented in equation 1, where P represents the percent speaking a LOTE in a given year.

$$P_{2009} = P_{2010} = P_{2011} = P_{2012} = \dots = P_{2020} \quad [1]$$

The constant model assumes that future LOTE use will remain constant at recently estimated levels, and consequently there would be no change in the distribution of LOTE speakers within age and Hispanic origin groups. In this model, changes in the number of speakers will be driven by changes in the population projections. The percentage of LOTE speakers remains the same through 2020, but we apply these percentages to a population that is changing over time. If the size of a group increases over time, so will the number of speakers.

The other two models we use are a linear model and a logistic model, which are based on the assumption that language use can change over time and are based on trends in LOTE use observed in the four years of ACS data (2006-2009). The linear model assumes that language use in the future will change by the same amount as in the past and is represented by equation 2, where P_t represents the percent speaking a LOTE at time t , a is the estimated intercept, b is the estimated slope, and t is the year of data being projected.

$$P_t = a + b(t) \quad [2]$$

The third series, based on an assumption of logistic growth, is also based on trends in LOTE use from the 2006 through 2009 ACS. In contrast to the assumption of linear growth, the logistic model assumes that growth is constrained by an upper and lower bound. The logistic model is represented by equation 3, where P_t represents the percent speaking a LOTE at time t ; a , b , and c are estimated parameters, and t is the year of data being projected.

$$P = \frac{a}{[1 + (b)(e^{-ct})]} \quad [3]$$

The linear model has the potential to exceed the bounds of the percent distribution, rising above 100 percent or falling below zero, whereas the logistic model will constrain growth as it approaches the upper and lower asymptotes of the distribution. In contrast to the constant model, where changes in the number of speakers will be driven by the population projections, for the linear and logistic models, changes in the number of speakers will be driven by both changes in

³ The 2008 and 2009 National Population Projections do not incorporate 2010 Census results. Projections using the 2010 Census as a base are planned for release in 2012.

the projected percentages of LOTE speakers within each group and by changes in the population projections.

Comparison of Language Projection Models

Figures 1 and 2 provide two examples of what each projection model looks like, based on ACS data for two age and Hispanic origin groups. These groups illustrate two trends we observed in the ACS data. One group shows an increase in the number of LOTE speakers whereas the other group shows a decrease in the number speaking a LOTE.

Figure 1 shows the observed and projected percent speaking a LOTE at home for 36-year old non-Hispanics. This group showed an increase in LOTE use from 2006 to 2009, represented by the blue line in the figure. The red, green, and purple lines show what the projected percent of LOTE speakers will be for each of our three models. The constant series, represented by the red line, sets the projected percent of LOTE speakers for this group to equal the value observed in 2009, which was 13.8 percent. When this projected percent of LOTE speakers is applied to the projected population for this group, we would expect to see an increase in LOTE speakers so long as the projected population for this group increases over time. The green and purple lines show what the projected percent of LOTE speakers would be based on trends in the ACS data. These lines are very close to each other, illustrating that the linear and logistic models produce very similar results. When the percent projected to speak a LOTE is applied to the projected population for this group, we would expect to see an increase in the number of speakers. This increase would be larger than what would result from the constant model.

Figure 2 shows the observed and projected percent speaking a LOTE at home for 19-year old non-Hispanics. This group showed a slight decrease in LOTE use from 2006 to 2009, represented by the blue line in the figure. The projected percent of LOTE speakers for each of our three models is represented by the red, green, and purple lines in the figure. The constant series, represented by the red line in the figure, sets the projected percent of LOTE speakers for this group to equal 9.1 percent, which was the value observed in 2009. When this projected percent of LOTE speakers is applied to the projected population for this group, we would expect to see an increase in the number of LOTE speakers as long as the projected population for this group increases over time. The green and purple lines show what the projected percent of LOTE speakers would be based on trends in the ACS data. As was the case in the first example, the linear and logistic models produce very similar results. When

the percent projected to speak a LOTE is applied to the projected population for this group, we would expect to see a decrease in the number of LOTE speakers. The trend in this example is the trend that we found for a majority of the groups we projected for. As a result, the projected number of LOTE speakers in the constant model will increase over time as long as the population increases, while the linear and logistic models will show either small increases or in some cases a decrease in the number projected to speak a LOTE.

RESULTS

The results are presented in three sections. The first will address the overall use of a language other than English, followed by results for Spanish speakers, and finally the results for the other twelve individual languages we projected. The discussion presented in the paper is for the language projections based on the 2008 National Population Projections. Appendix Table 1 provides the results using the 2008 series. The results for the language projections using the 2009 National Population Projections are provided in appendix tables 2 through 4.

Language Other than English Use

The overall number speaking a LOTE is projected to increase in all three projection models (see Figure 3). We see the largest increase in the constant model, which is based on the simplistic assumption that the percent speaking a LOTE within the age and Hispanic origin groups we project would remain constant. When applying the constant proportions, we see a large amount of growth in the number of LOTE speakers. For the linear and logistic models, where a majority of groups actually showed decreases in the percent speaking a LOTE from 2006 to 2009, the projected increases in LOTE use are much smaller. While the population for these groups is projected to grow, the projected percent speaking a LOTE actually goes down. This results in a smaller increase in the overall number projected to speak a LOTE.

The distribution of the population by language spoken is presented in Figure 4. This figure shows the percent distribution of the population that is projected to speak a LOTE and those that are projected to speak only English in 2010, 2015, and 2020. In each of the three models, there is a small increase in the percent that is projected to speak a LOTE. For all three models, English is projected to remain the only language spoken by a majority of U.S. residents. The constant model does show a slightly larger increase in LOTE use compared to the linear and logistic models. This finding is expected given that the assumption of the constant

model is that the percent speaking a LOTE will remain constant at the levels observed in 2009, rather than to decrease over time as projected for several groups in the linear and logistic models.

Spanish Use

The number of Spanish speakers is projected to increase in all of the projection models (see Figure 5). As was the case for the overall number of LOTE speakers, the largest increase in the number of Spanish speakers occurs in the constant model, whereas for the linear and logistic models, which follow the trends in the ACS, the projected percent of the population speaking Spanish increases, but by a smaller amount. This is to be expected, since a majority of the age and Hispanic origin groups we projected showed a decrease in the percent speaking Spanish. While the projected population increases over time, the percentage speaking Spanish decreased for many groups. This resulted in smaller increases in the overall number projected to speak Spanish in the linear and logistic models, compared to results for the constant assumption.

Figure 6 presents the percent of the total population five years and older that is projected to speak Spanish in 2010, 2015, and 2020. The percent speaking Spanish is projected to increase slightly over the next decade. In 2009, just over 12 percent of the population spoke Spanish at home. Under the assumptions that use of Spanish would remain constant over the next ten years, nearly 16 percent of the population 5 years or older is projected to speak Spanish. The linear and logistic models project a smaller increase, to just over 13 percent in 2020.

Spanish is projected to remain the language spoken by a majority of LOTE speakers (see Figure 7). In 2009, 63 percent of LOTE speakers reported speaking Spanish at home. This increased to almost 68 percent in the constant series, while the percent projected to speak Spanish held steady at just over 62 percent in the linear and logistic models in 2020.

Use of Other Languages

The projected change between 2010 and 2020 in the population speaking French, Italian, Portuguese, German, Russian, and Polish is presented in Figure 8. The constant model shows an increase in the number of speakers for all languages. This is expected because the driver of change for this model is the population projections. In the linear and logistic models, which are based on observed trends, the population speaking French, Italian, German, and Polish is projected to decline. The decline in the number of speakers for these

languages is also consistent with longer term trends observed in the 1980, 1990, and 2000 Census data (Table 2). The population speaking Portuguese and Russian is projected to increase in the linear and logistic models, and the increases are higher than what was projected in the constant model, indicating that trends in the ACS data show growth in the use of these languages.

Figure 9 shows the projected change in the population that speaks Hindi, Chinese, Korean, Vietnamese, Tagalog, and Arabic. With the exception of Korean, use of the non-European languages is projected to increase over the next ten years in all three models. While the number of Korean speakers increased from 1980 to 2000, trends in ACS data show that the use of Korean has declined in recent years. As a result, Korean is projected to decline in the linear and logistic models.

Figures 10, 11, and 12 present the distribution of LOTE speakers by the language spoken for the constant, linear, and logistic models, respectively. Spanish, which was presented in Figure 7, and Chinese are the most commonly spoken languages in all three projections series, followed by French and Tagalog. Polish is the least spoken language among the thirteen languages we projected. In the constant model, all languages, except Spanish, are projected to decrease slightly as a percent of overall LOTE use (see Figures 7 and 10). In the linear model, Russian, Hindi, Tagalog, and Arabic increased slightly as a percent of overall LOTE use, while the other languages were either maintained at levels projected for 2010 or decreased slightly (see Figures 7 and 11). For the logistic model, Hindi, Chinese, Vietnamese, Tagalog, and Arabic all increased slightly as a percent of overall LOTE use, while the other languages were either maintained at current levels, or decreased slightly (see Figures 7 and 12).

CONCLUSIONS

This research suggests that the United States will continue to be a linguistically diverse nation in the coming years. The projections we produced show that the use of LOTE is projected to increase over the next ten years, though English is expected to continue to be the only language spoken by a substantial majority of all U.S. residents 5 years and older. The population speaking Spanish, as well as the populations speaking Portuguese, Russian, Hindi, Chinese, Vietnamese, Tagalog, and Arabic are projected to increase. Spanish is projected to remain the most commonly spoken non-English language. The linear and logistic models suggest that the populations speaking French, Italian, German, Polish, and Korean can be expected to decrease over the next decade.

The assumption of constant growth is likely overly simplistic, as it results in an increase in LOTE use for all languages, even those that are shown to decline in Census and in ACS data. The linear and logistic assumptions are perhaps more realistic, following observed trends, and provide results that are very similar. Since the logistic assumption is constrained within upper and lower bounds, and cannot produce projected percentages below zero or above 100, we may consider adopting the logistic model for use in future work.

As we move forward with this research, we plan to add 2010 ACS data to the time series that provides the basis for these projections, extending the time series to five years. We will also use the 2010-Census based population projections when they become available. Increasing the sample size could reduce variation resulting from sampling variability and improve the robustness of our results. In an effort to increase the sample size of the age and Hispanic origin groups we project, we will consider projecting by age groups instead of single years of age or using three-year ACS files instead of single year files to form the basis of the time series.

We will also consider projecting by birth cohorts instead of by age. A cohort approach will entail following cohorts of individuals as they grow older, instead of comparing language use of the population of the same age at different points in time. Studies have shown that language use can shift and change over the life course (Lutz, 2006; Ortman and Stevens, 2008; Portes and Rumbaut, 2001), which supports the adoption of a cohort approach to projecting language use into the future.

We did not project language use by nativity or generational status. Research shows that the use of non-English languages is strongly linked to immigration and is most frequent among first generation residents (Alba et al., 2002; Rumbaut et al., 2006; Stevens, 1992). The Census Bureau's population projections do not currently separate the population by foreign and native-born status. Should projections by nativity become available, we could further develop our methodology to project by nativity status, which could inform and improve the accuracy of the language projections.

REFERENCES

Alba, Richard, John Logan, Amy Lutz, and Brian Stults. 2002. "Only English by the Third Generation? Loss and Preservation of the Mother Tongue among the Grandchildren of Contemporary Immigrants." *Demography* 39(3): 467-484.

Bean, Frank D. and Gillian Stevens. 2005. *America's Newcomers and the Dynamics of Diversity*. Russell Sage Foundation: New York.

Chiswick, Barry R. and Paul W. Miller. 2001. "A Model of Destination-Language Acquisition: Application to Male Immigrants in Canada." *Demography* 38(3): 391-409.

Cohen-Goldner, Sarit and Zvi Eckstein. 2008. "Labor Mobility of Immigrants: Training, Experience, Language and Opportunities." *International Economic Review* 49(3): 837-872.

Espenshade, Thomas J. and Haishan Fu. 1997. "An Analysis of English-Language Proficiency Among U.S. Immigrants." *American Sociological Review* 62(2): 288-305.

Hurtado, Aída and Luis A. Vega. 2004. "Shift Happens: Spanish and English Transmission Between Parents and Their Children." *Journal of Social Issues* 60:137-155.

Lutz, Amy. 2006. "Spanish Maintenance among English-Speaking Latino Youth: The Role of Individual and Social Characteristics." *Social Forces* 84(3): 1417-1433.

Mouw, Ted and Yu Xie. 1999. "Bilingualism and the Academic Achievement of First- and Second-Generation Asian Americans: Accommodations With or Without Assimilation." *American Sociological Review* 64(2): 232-252.

Ortman, Jennifer M. and Gillian Stevens. 2008. "Shift Happens, But When?: Inter- and Intra-Generational Language Shift among Hispanic Americans." Presented at the Annual Meetings of the Population Association of America, April 17-19, New Orleans, LA.

Portes, Alejandro and Rubén G. Rumbaut. 2001. *Legacies: The Story of the Immigrant Second Generation*. Berkeley: University of California Press.

Rumbaut, Rubén G. 1997. "Paradoxes (and Orthodoxies) of Assimilation." *Sociological Perspectives* 40:483-511.

Rumbaut, Rubén G., Douglas S. Massey, and Frank D. Bean. 2006. "Linguistic Life Expectancies: Immigrant Language Retention in Southern California." *Population and Development Review* 32(3): 447-460.

Shin, Hyon B. and Rosalind Bruno. 2003. "Language Use and English-Speaking Ability: 2000." U.S. Census Bureau, Washington, D.C.

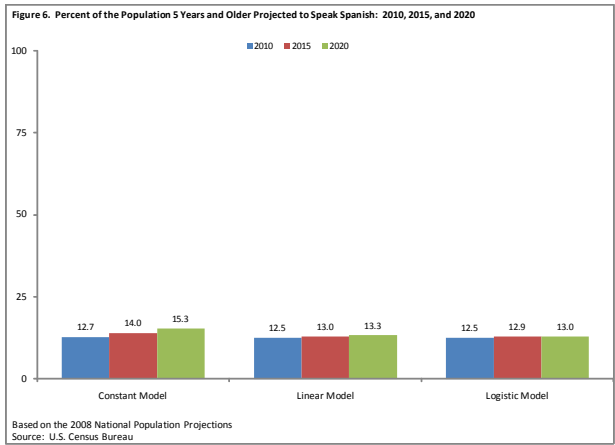
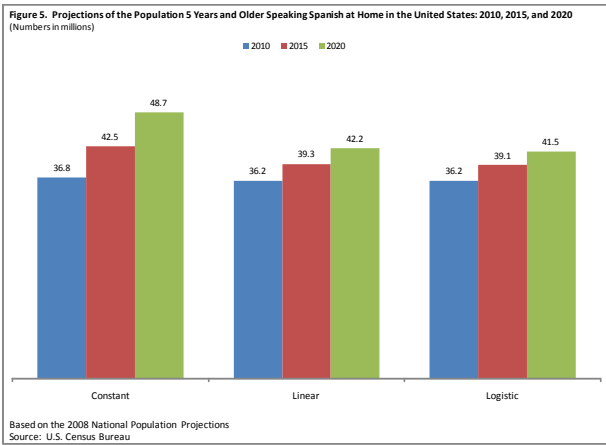
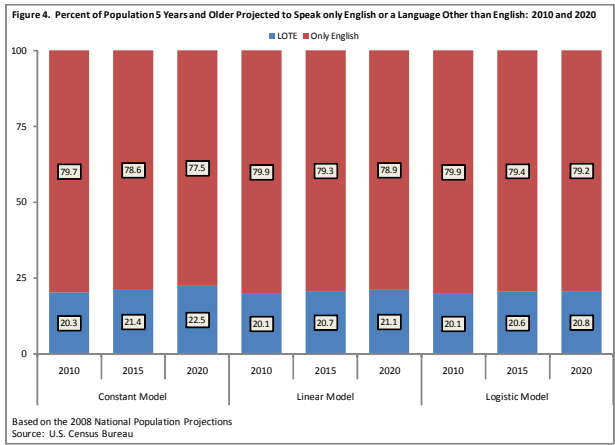
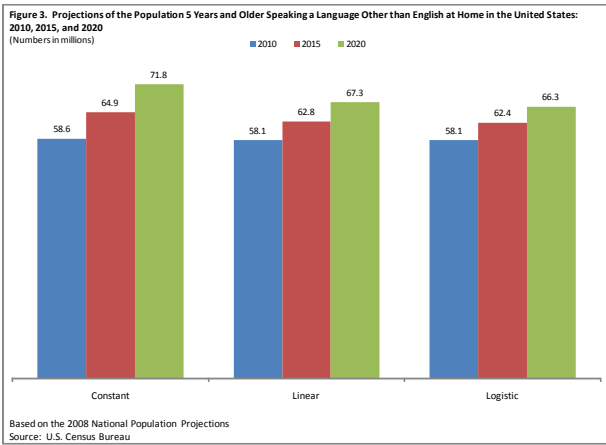
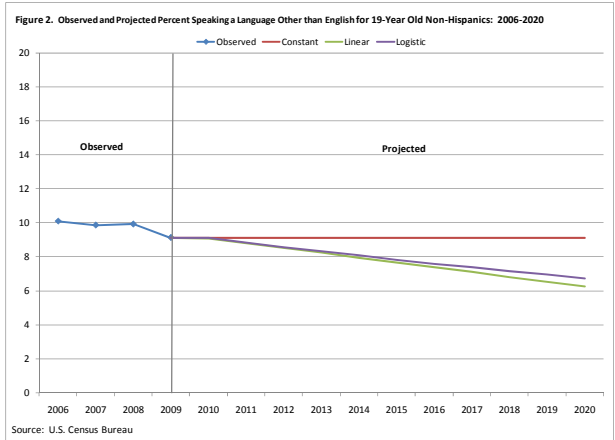
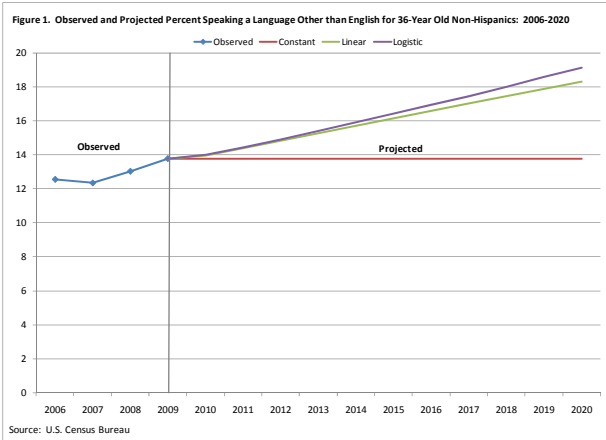
Stevens, Gillian. 1992. "The Social and Demographic Context of Language Use in the United States." *American Sociological Review* 57(2): 171-185.

_____. 1994. "Immigration, Emigration, Language Acquisition, and the English Language Proficiency of Immigrants in the United States." Pp. 163-185 in *Immigration and Ethnicity: The Integration of America's Newest Arrivals*, edited by B. Edmonston and J. S. Passel. Washington, D.C.: The Urban Institute Press.

_____. 1999. "A Century of U.S. Censuses and the Language Characteristics of Immigrants." *Demography* 36(3): 387-397.

U.S. Census Bureau. 2008. "2008 National Population Projections." Washington, DC: U.S. Census Bureau. (<http://www.census.gov/population/www/projections/2008projections.html>).

_____. 2009. "2009 National Population Projections (Supplemental)." Washington, DC: U.S. Census Bureau. (<http://www.census.gov/population/www/projections/2009projections.html>).



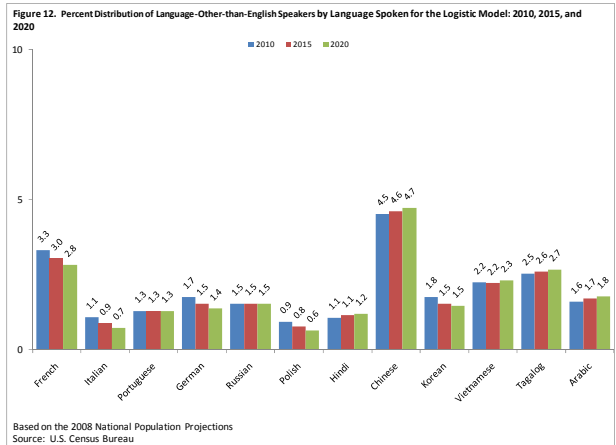
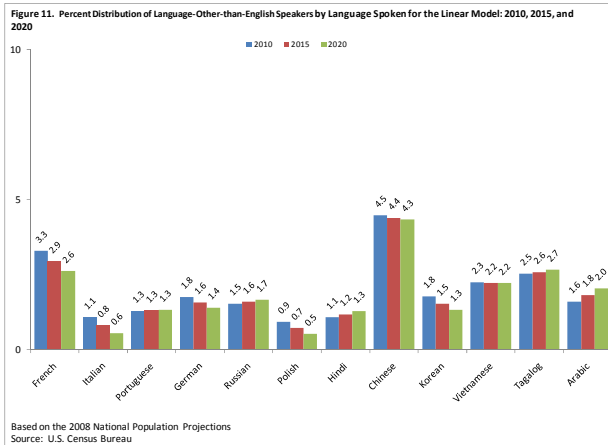
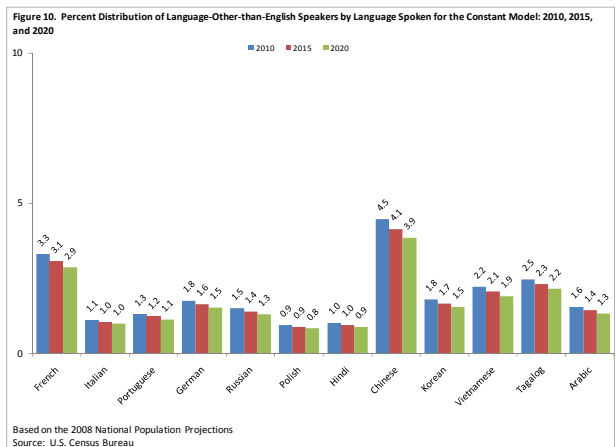
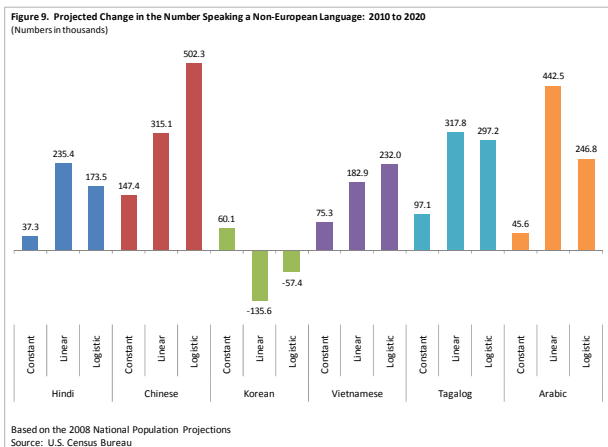
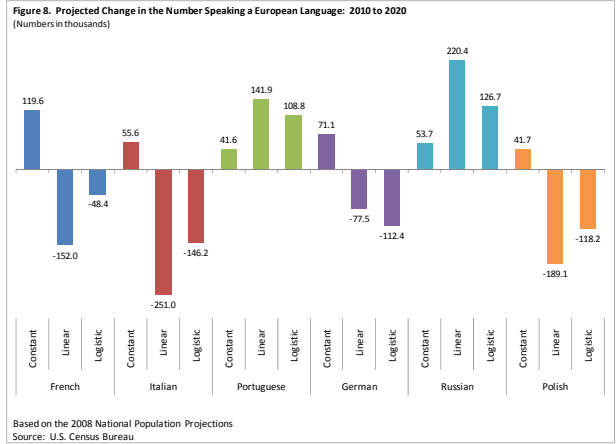
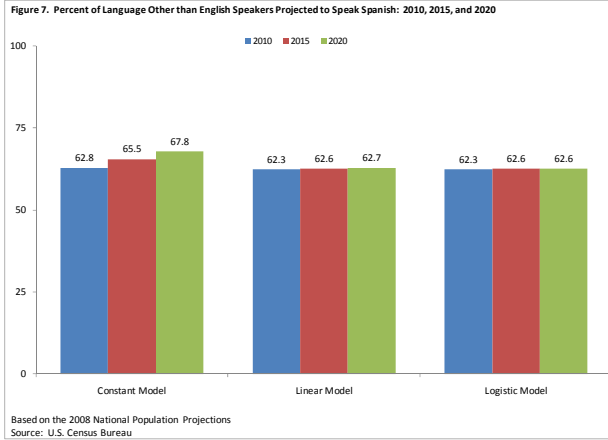


Table 1. Languages Spoken at Home: 1980, 1990, 2000, and 2009

Characteristic	1980	1990	2000	2006	2007	2008	2009	Percentage change 1980-2009
Population 5 Years and older	210,247,455	230,445,777	262,375,152	279,012,712	280,950,438	283,156,079	285,797,349	35.9
Spoke only English at home	187,187,415	198,600,798	215,423,557	224,154,288	225,505,953	227,295,534	228,699,523	22.2
Spoke a language other than English at home	23,060,040	31,844,979	46,951,595	54,858,424	55,444,485	55,860,545	57,097,826	147.6
Spoke a language other than English at home²	23,060,040	31,844,979	46,951,595	54,858,424	55,444,485	55,860,545	57,097,826	147.6
Spanish or Spanish Creole	11,116,194	17,345,064	28,101,052	34,044,945	34,547,077	34,615,394	35,468,501	219.1
French (includes Patois, Cajun, Creole)	1,550,751	1,930,404	2,097,206	1,997,618	1,984,824	1,973,531	1,964,556	26.7
Italian	1,618,344	1,308,648	1,008,370	828,524	798,801	782,173	753,992	-53.4
Portuguese or Portuguese Creole	351,875	430,610	564,630	683,405	687,126	661,120	731,282	107.8
German	1,586,593	1,547,987	1,383,442	1,135,999	1,104,354	1,121,465	1,109,216	-30.1
Russian	173,226	241,798	706,242	823,210	851,174	860,568	881,723	409.0
Polish	820,647	723,483	667,414	640,265	638,059	616,492	593,598	-27.7
Hindi ¹	(NA)	(NA)	317,057	504,607	532,911	562,587	560,983	(NA)
Chinese	630,806	1,319,462	2,022,143	2,492,871	2,464,572	2,473,968	2,600,150	312.2
Korean	266,280	626,478	894,063	1,060,631	1,062,337	1,048,400	1,039,021	290.2
Vietnamese	197,588	507,069	1,009,627	1,207,721	1,207,004	1,236,419	1,251,468	533.4
Tagalog	474,150	843,251	1,224,241	1,415,599	1,480,429	1,496,208	1,513,734	219.3
Arabic	217,529	355,150	614,582	732,519	767,319	780,995	845,396	288.6

NA Not available.

¹ Prior to 2000, Hindi and Urdu speakers were combined in the same language group (Indic languages). Individual estimates of Hindi speakers are not available for 1980 and 1990.

² The total does not match the sum of the 17 languages listed in this table because the total includes other languages that are not listed here. The 13 languages listed in this tables are those language with 500,000 or more speakers in 2009.

Sources: U.S. Census Bureau, 1980 and 1990 Census, Census 2000, and 2006-2009 American Community Survey.

Table A-1. Detailed Languages Spoken at Home by English-Speaking Ability for the Population 5 Years and Older: 2009

	Number of speakers	Percentage of speakers of a non-English language	English-speaking ability			
			Very well	Well	Not well	Not at all
Population 5 years and older	285,797,349	(X)	(X)	(X)	(X)	(X)
Spoke only English at home	228,699,523	(X)	(X)	(X)	(X)	(X)
Spoke a language other than English at home	57,097,826	100.0	56.9	19.6	15.9	7.5
Spoke a language other than English at home	57,097,826	100.0	56.9	19.6	15.9	7.5
Spanish or Spanish Creole	35,468,501	62.1	54.3	18.0	17.9	9.8
Other Indo-European languages	10,495,295	18.4	67.5	19.6	10.1	2.8
French	1,305,503	2.3	79.9	13.9	5.7	0.5
French Creole	659,053	1.2	54.2	25.9	15.1	4.8
Italian	753,992	1.3	72.6	17.5	8.8	1.0
Portuguese	731,282	1.3	58.6	20.8	15.3	5.3
German	1,109,216	1.9	83.3	12.6	3.8	0.2
Yiddish	148,155	0.3	66.2	19.4	11.0	3.4
Other West Germanic languages	271,227	0.5	77.7	18.0	3.7	0.7
Scandinavian languages	126,337	0.2	89.0	9.2	1.8	
Greek	325,747	0.6	75.3	15.1	8.7	0.9
Russian	881,723	1.5	49.8	27.3	17.2	5.7
Polish	593,598	1.0	57.6	25.5	14.0	2.9
Serbo-Croatian	269,333	0.5	61.4	21.7	13.6	3.3
Other Slavic languages	298,094	0.5	61.5	21.7	13.2	3.6
Armenian	242,836	0.4	54.8	22.5	14.7	8.1
Persian	396,769	0.7	62.1	21.8	11.3	4.8
Gujarathi	341,404	0.6	64.2	20.2	11.8	3.9
Hindi	560,983	1.0	78.0	16.0	4.9	1.1
Urdu	355,964	0.6	70.3	18.7	8.7	2.2
Other Indic languages	668,596	1.2	60.6	23.5	11.1	4.7
All other Indo-European languages	455,483	0.8	64.1	23.4	9.3	3.2
Asian and Pacific Island languages	8,698,825	15.2	51.8	25.8	17.1	5.3
Chinese	2,600,150	4.6	45.1	26.0	19.5	9.4
Japanese	445,471	0.8	55.3	27.5	15.6	1.5
Korean	1,039,021	1.8	43.3	28.5	23.6	4.6
Mon-Khmer, Cambodian	202,033	0.4	48.4	23.6	21.9	6.1
Hmong	193,179	0.3	53.5	24.6	15.2	6.6
Thai	152,679	0.3	48.5	32.1	17.0	2.4
Laotian	146,297	0.3	50.2	24.5	20.3	5.0
Vietnamese	1,251,468	2.2	39.2	27.1	26.2	7.5
Other Asian languages	792,756	1.4	68.4	20.8	8.4	2.4
Tagalog	1,513,734	2.7	68.5	24.4	6.6	0.5
Other Pacific Island languages	371,653	0.7	60.8	25.7	12.3	1.2
Other languages	2,435,205	4.3	69.0	20.1	8.5	2.3
Navajo	169,009	0.3	77.3	13.6	6.8	2.3
Other Native American languages	196,372	0.3	84.5	11.3	3.7	0.5
Hungarian	90,612	0.2	67.4	24.1	8.0	0.5
Arabic	845,396	1.5	63.0	22.9	11.3	2.8
Hebrew	221,593	0.4	82.1	14.8	2.8	0.2
African languages	777,553	1.4	67.7	22.2	7.8	2.2
All other languages	125,054	0.2	61.6	17.2	13.9	7.3

X Not applicable.

Sources: U.S. Census Bureau, The 2009 American Community Survey.
For more information on ACS see <http://www.census.gov/acs/www/>

Table A-2. Projected Population Speaking a Language Other than English at Home: 2010, 2015, and 2020, 2008 National Population Projections

	2010	2015	2020	2010	2015	2020	2010	2015	2020
	(In thousands)			(Percent of population ages 5 and over)			(Percent of population speaking a language other than English)		
Population ages 5 and over	286,998	299,378	311,886	100.0	100.0	100.0	(X)	(X)	(X)
Constant Model									
Only English	230,573	238,538	246,736	80.3	79.7	79.1	(X)	(X)	(X)
Language other than English	58,560	64,926	71,805	20.4	21.7	23.0	100.0	100.0	100.0
Spanish	36,780	42,500	48,711	12.8	14.2	15.6	62.8	65.5	67.8
French	1,934	1,994	2,054	0.7	0.7	0.7	3.3	3.1	2.9
Italian	652	681	707	0.2	0.2	0.2	1.1	1.0	1.0
Portuguese	762	782	804	0.3	0.3	0.3	1.3	1.2	1.1
German	1,031	1,066	1,102	0.4	0.4	0.4	1.8	1.6	1.5
Russian	881	908	934	0.3	0.3	0.3	1.5	1.4	1.3
Polish	555	578	597	0.2	0.2	0.2	0.9	0.9	0.8
Hindi	601	621	638	0.2	0.2	0.2	1.0	1.0	0.9
Chinese	2,623	2,694	2,771	0.9	0.9	0.9	4.5	4.1	3.9
Korean	1,051	1,077	1,111	0.4	0.4	0.4	1.8	1.7	1.5
Vietnamese	1,300	1,335	1,376	0.5	0.4	0.4	2.2	2.1	1.9
Tagalog	1,448	1,495	1,545	0.5	0.5	0.5	2.5	2.3	2.2
Arabic	911	932	956	0.3	0.3	0.3	1.6	1.4	1.3
Linear Model									
Only English	231,001	240,692	251,202	80.5	80.4	80.5	(X)	(X)	(X)
Language other than English	58,132	62,772	67,339	20.3	21.0	21.6	100.0	100.0	100.0
Spanish	36,238	39,305	42,229	12.6	13.1	13.5	62.3	62.6	62.7
French	1,912	1,846	1,760	0.7	0.6	0.6	3.3	2.9	2.6
Italian	625	508	373	0.2	0.2	0.1	1.1	0.8	0.6
Portuguese	750	815	891	0.3	0.3	0.3	1.3	1.3	1.3
German	1,023	990	945	0.4	0.3	0.3	1.8	1.6	1.4
Russian	892	999	1,113	0.3	0.3	0.4	1.5	1.6	1.7
Polish	539	452	350	0.2	0.2	0.1	0.9	0.7	0.5
Hindi	627	742	862	0.2	0.2	0.3	1.1	1.2	1.3
Chinese	2,601	2,758	2,916	0.9	0.9	0.9	4.5	4.4	4.3
Korean	1,033	959	898	0.4	0.3	0.3	1.8	1.5	1.3
Vietnamese	1,309	1,394	1,492	0.5	0.5	0.5	2.3	2.2	2.2
Tagalog	1,474	1,624	1,792	0.5	0.5	0.6	2.5	2.6	2.7
Arabic	933	1,147	1,375	0.3	0.4	0.4	1.6	1.8	2.0
Logistic Model									
Only English	231,037	241,055	252,216	80.5	80.5	80.9	(X)	(X)	(X)
Language other than English	58,096	62,409	66,325	20.2	20.8	21.3	100.0	100.0	100.0
Spanish	36,221	39,072	41,525	12.6	13.1	13.3	62.3	62.6	62.6
French	1,921	1,896	1,872	0.7	0.6	0.6	3.3	3.0	2.8
Italian	629	548	482	0.2	0.2	0.2	1.1	0.9	0.7
Portuguese	748	805	857	0.3	0.3	0.3	1.3	1.3	1.3
German	1,017	958	904	0.4	0.3	0.3	1.8	1.5	1.4
Russian	887	961	1,013	0.3	0.3	0.3	1.5	1.5	1.5
Polish	544	480	426	0.2	0.2	0.1	0.9	0.8	0.6
Hindi	621	711	794	0.2	0.2	0.3	1.1	1.1	1.2
Chinese	2,629	2,873	3,131	0.9	1.0	1.0	4.5	4.6	4.7
Korean	1,024	949	966	0.4	0.3	0.3	1.8	1.5	1.5
Vietnamese	1,300	1,381	1,532	0.5	0.5	0.5	2.2	2.2	2.3
Tagalog	1,475	1,619	1,773	0.5	0.5	0.6	2.5	2.6	2.7
Arabic	923	1,065	1,170	0.3	0.4	0.4	1.6	1.7	1.8

X Not applicable.

Source: U.S. Census Bureau