

Examining Multilingualism in the United States Using ACS Language Write-ins

Presented at the 2024 annual meeting of the
Population Association of America
Columbus, OH
April 17-20, 2024

Daniela Mejía
Social, Economic, and Housing Statistics Division
U.S. Census Bureau
daniela.mejia@census.gov

BACKGROUND AND DATA

- The American Community Survey (ACS) asks if people aged 5 years and older speak a **language other than English (LOTE)** at home.
 - If “Yes”, the language is recorded as a write-in response character string.
 - Only the first LOTE is analyzed, although write-in responses can include multiple languages.
- The ACS releases **1-year estimates** for geographies with populations of 65,000 people or more and **5-year estimates** for all geographic areas down to the census tract and block group levels.
- More information can be found at www.census.gov/acs.

PROJECT GOALS

- Create a text processing algorithm to classify entire write-ins into standard ACS language categories.
- Identify the ten most common sets of languages spoken among multilinguals in the United States.
- Compare the demographic characteristics of bilinguals and multilinguals to English-only speakers using the complete ACS questionnaire.

METHODS

- Classification algorithm tested using 200 cases from the 2016 1-year ACS and then applied to the 2017-2021 5-year ACS ($n=22,647,229$).
- Multilinguals** identified as people who spoke English “Well” or “Very Well” and reported two or more LOTEs spoken at home as write-in response; **bilinguals** identified with one LOTE write-in response in addition to speaking English “Well” or “Very Well”.
- Two binomial logistic regression models compare multilinguals and bilinguals to English-only speakers.

LANGUAGE CLASSIFICATION ALGORITHM

- Remove stop words.**
some Spanish and Iranian Persian → Spanish Iranian Persian
- Run through Hunspell spell checker.**
Spanish Iranian Persian → Spanish Iranian Persian
- Correct for redundant coding and special cases by creating 1-word n-grams.**
Iranian Persian → Iranian Persian
- Analyze string as a series of n-grams of various sizes (1 through up to 4).**
Spanish Iranian Persian → Spanish Iranian Persian
Spanish Iranian Persian → NA

Odds Ratios

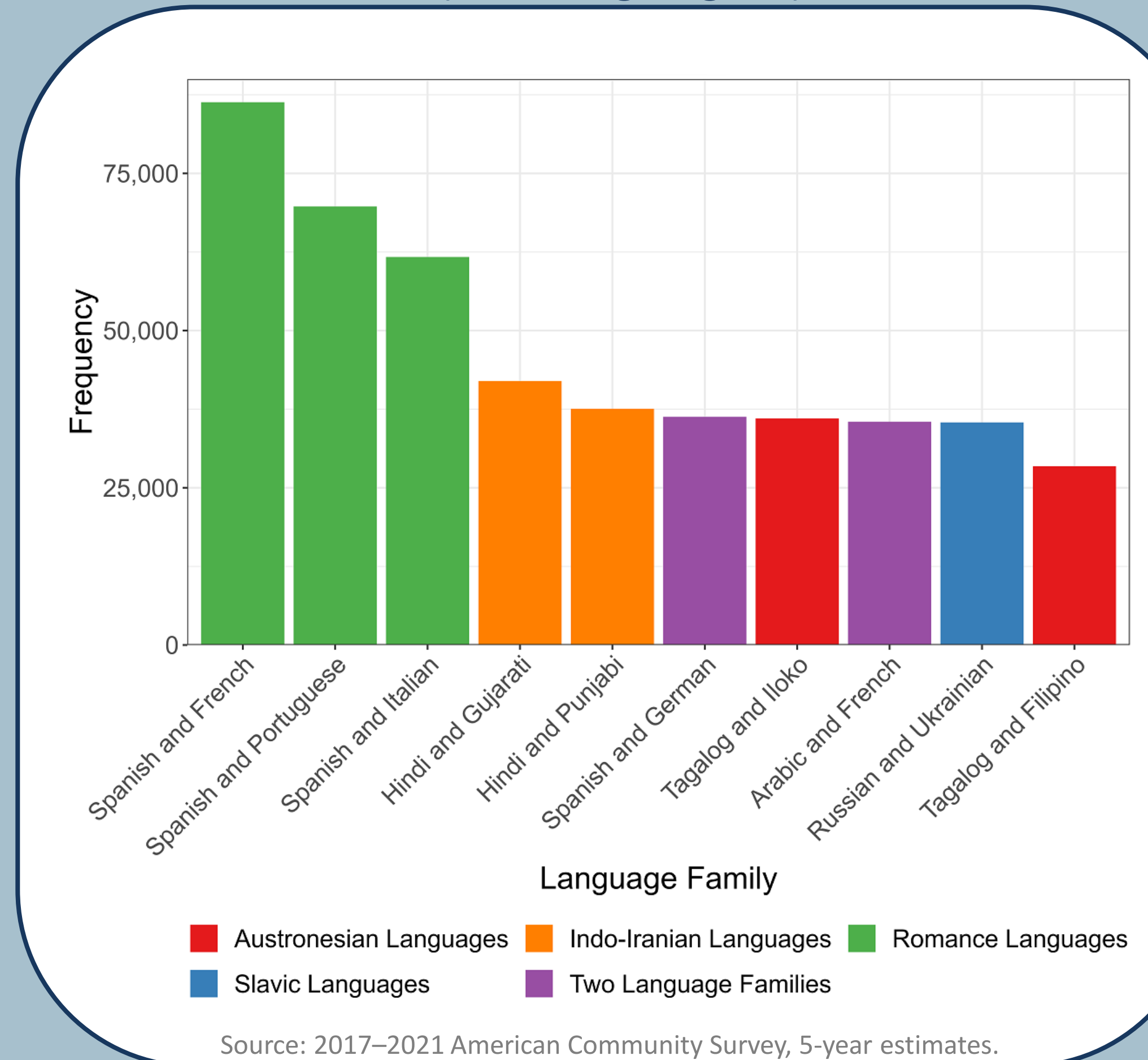
Compared to English-only speakers and controlling for eight additional occupation categories (not shown).

Predictors	Bilingual		Multilingual	
	OR	95% CI	OR	95% CI
Individual income (log)	0.96	(0.95, 0.96)	0.94	(0.93, 0.95)
Age (log)	0.66	(0.66, 0.67)	0.90	(0.86, 0.94)
Male	1.02	(1.01, 1.03)	1.18	(1.15, 1.20)
Bachelor’s degree or higher education (vs. less than Bachelor’s degree)	1.21	(1.20, 1.22)	2.06	(1.99, 2.13)
Race and Hispanic origin				
White	0.68	(0.68, 0.69)	0.76	(0.73, 0.80)
Black	0.56	(0.55, 0.58)	0.75	(0.70, 0.80)
Asian	5.10	(5.00, 5.20)	2.67	(2.53, 2.82)
Hispanic (any race)	28.53	(28.20, 28.85)	1.59	(1.52, 1.67)
Foreign-born	30.07	(29.72, 30.43)	51.83	(50.22, 53.49)
Labor force participant	1.11	(1.10, 1.13)	1.15	(1.11, 1.19)

Source: U.S. Census Bureau, 2017–2021 American Community Survey, 5-year estimates.

RESULTS

Ten Most Common Language Sets (Including English)



Source: 2017–2021 American Community Survey, 5-year estimates.

ACS LANGUAGE QUESTION

14 a. Does this person speak a language other than English at home?

Yes
 No → SKIP to question 15a

b. What is this language?

French and French Creole

For example: Korean, Italian, Spanish, Vietnamese

c. How well does this person speak English?

Very well
 Well
 Not well
 Not at all

Source: 2021 American Community Survey (ACS) questionnaire.

CONCLUSIONS

- Language classification algorithm accuracy: **90.9%**.
- Most common languages spoken at home by multilinguals were **Spanish, French, and English**.
- Most common LOTE sets were in the **Romance language family** (Spanish, French, Portuguese, Italian).
 - Languages from South Asia and the Philippines also common.
- Odds ratios show significant differences across all three speaker groups:
 - Multilinguals were most likely to have a bachelor’s degree or higher education, being **2x** more likely than English-only speakers.
 - Bilinguals were over **28x** more likely to be of Hispanic origin than English-only speakers and about **5x** more likely to be Asian, while English-only speakers were more likely to be White or Black.
 - Even after controlling for all other variables in the model, speaking more than one language was **negatively associated** with individual income for both multilinguals and bilinguals.