

# Abstract

**Background:** Language acquisition is a fundamental aspect of human development, yet the relative contributions of environmental versus biological factors remain underexplored in large-scale empirical studies. This observational study examines whether geographic birthplace predicts primary language spoken across multiple nations.

**Methods:** We analysed national census data from nine countries ( $N > 1.8$  billion individuals) spanning six continents. Primary outcome was concordance between country of residence and dominant national language spoken. Chi-square tests and effect size calculations (Cramér's  $V$ , Cohen's  $h$ ) were conducted.

**Results:** Across all nations examined, geographic residence demonstrated strong concordance with national language acquisition (range: 72.0% - 96.9%). Effect sizes ranged from  $h = 0.46$  to  $h = 1.22$  (mean = 0.93; classified as "medium" to "large" by conventional standards). The observed pattern held regardless of the specific language examined (English, French, Mandarin, Spanish, etc.).

**Conclusions:** Geographic environment appears to be an extraordinarily strong predictor of language acquisition, with effect sizes exceeding those typically observed in behavioural research. The implications of these findings for understanding human behavioural acquisition more broadly warrant further investigation.

**Keywords:** language acquisition, environmental factors, cross-national study, census data

## 1. Introduction

### 1.1 Background

Human language is among the most complex cognitive abilities exhibited by any species. The average adult possesses a productive vocabulary of approximately 20,000-35,000 words, applies grammatical rules unconsciously in real-time, and processes speech at rates exceeding 150 words per minute (Brysbaert et al., 2016). Despite this complexity, healthy children across all cultures acquire language with remarkable consistency.

The question of how language is acquired has been debated extensively. Nativist perspectives emphasise innate language acquisition devices (Chomsky, 1965), while empiricist perspectives highlight environmental exposure and social learning (Tomasello, 2003). However, large-scale empirical studies examining the actual distribution of language outcomes across populations remain limited.

## 1.2 Research Question

This study addresses a straightforward empirical question: To what extent does geographic birthplace predict the primary language an individual speaks?

We then performed a systematic cross-national analyses quantifying this relationship with standardised effect size metrics. Such quantification may provide useful baseline data for understanding environmental contributions to complex human behaviours.

## 1.3 Hypotheses

**H<sub>0</sub> (Null Hypothesis):** Geographic birthplace is not associated with primary language spoken, and language acquisition occurs independently of geographic environment.

**H<sub>1</sub> (Alternative Hypothesis):** Geographic birthplace is associated with primary language spoken, and language acquisition is related to geographic environment.

We set our significance threshold at  $\alpha = 0.05$  and will report effect sizes following conventional interpretations.

## 2. Methods

### 2.1 Design

Cross-sectional observational study using publicly available national census data.

### 2.2 Data Sources

We identified national statistical agencies with publicly available census data on language spoken. Countries were selected based on:

1. Availability of recent census data (2011-2022)
2. Data published in or translatable to English
3. Inclusion of language variables
4. Geographic and linguistic diversity

Final sample included nine nations across six continents (Table 1).

## 2.3 Variables

**Predictor Variable:** Country of residence at time of census (categorical)

**Outcome Variable:** Primary/main language spoken (categorical, operationalised as national language vs. other)

**Control Variables:** None in primary analysis (exploratory design)

## 2.4 Statistical Analysis

For each country, we calculated:

1. Proportion speaking the dominant national language
2. Chi-square goodness-of-fit test against null expectation (50% by chance alone)
3. Effect size (Cohen's  $h$  for proportion comparisons)
4. Cramér's  $V$  for overall association strength

Effect size interpretation followed Cohen (1988):

- Small:  $h = 0.20$
- Medium:  $h = 0.50$
- Large:  $h = 0.80$
- The specific language (English, French, Mandarin, Spanish, German, Hindi)
- Geographic region (Europe, Asia, Oceania, North America, South America)
- Population size (4.7 million to 1.4 billion)

- Economic development level (varied)

---

## 4. Discussion

### 4.1 Summary of Findings

This cross-national observational study examined whether geographic birthplace predicts primary language spoken. Across nine nations representing over 1.8 billion individuals, we found:

1. Strong concordance: Proportions speaking the national language ranged from 72.0% to 96.9%
2. Statistical significance: All chi-square tests significant at  $p < 0.001$  | Very Large\* |

\*Extended interpretation; Cohen's original scale did not extend beyond 0.80.

---

#### Michigan Real-Life Trial Dataset:

<https://web.eecs.umich.edu/~zmohamed/resources.html>

#### DePaulo et al. (2003) – Cues to Deception:

<https://psycnet.apa.org/record/2002-11509-006>

(Also freely available at:

<https://smg.media.mit.edu/library/DePauloEtAl.Cues%20to%20Deception.pdf> )

#### GDRT (2006) – A World of Lies:

<https://journals.sagepub.com/doi/10.1177/0022022105282295>

(Also on Academia.edu: [https://www.academia.edu/6463281/A\\_world\\_of\\_lies](https://www.academia.edu/6463281/A_world_of_lies) )

#### Rizzelli et al. (2021) – Language of Criminal Confessions:

<https://wclawr.org/index.php/wclr/article/view/58>

(Thesis with raw data: [https://academicworks.cuny.edu/jj\\_etds/117/](https://academicworks.cuny.edu/jj_etds/117/) )

**Bond & DePaulo (2006) – Accuracy of Deception Judgments:**

[https://journals.sagepub.com/doi/10.1207/s15327957pspr1003\\_2](https://journals.sagepub.com/doi/10.1207/s15327957pspr1003_2)

**Lim et al. (2022) – Autistic Adults Perceived as Deceptive:**

<https://pmc.ncbi.nlm.nih.gov/articles/PMC8813809/>

**Bagnall et al. (2023) – Police Interviews with Autistic Adults:**

<https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2023.1117415/full>

**Kassin (2017) – False Confessions:**

<https://doi.org/10.1037/amp0000195>

All open access or with freely available versions. Did you get your DOI?

*Manuscript submitted: February 2026*

*Conflicts of interest: None.*