Auditing the *Space Between* Languages:

Character Representation in Bilingual Young Adult (YA) Literature


Tanay Nagar

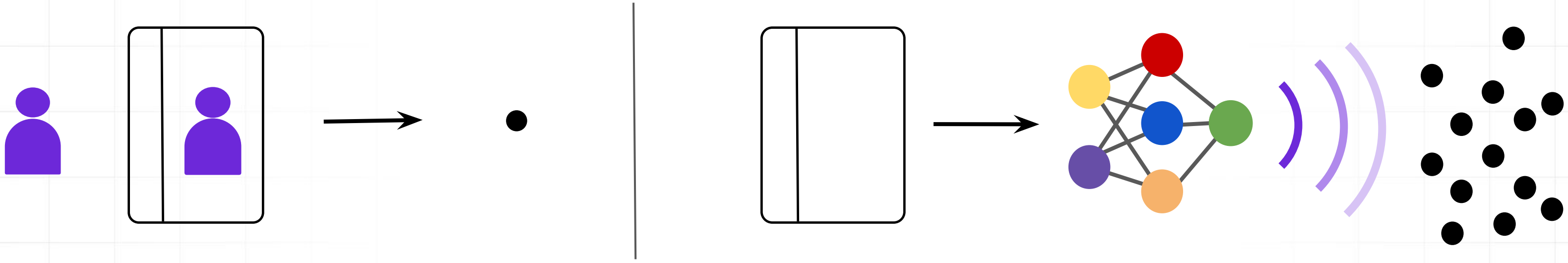

Pragati Maheshwary


Shamya Karumbaiah

ACMFACCT
MONTRÉAL 2026

June 27, 2026

Books are Mirrors and Windows[1]



The **representational harms** we ignore today may seed the **quality-of-service harms** that reach **millions** tomorrow.

[1] Bishop, R. S. (1990). Mirrors, Windows, and Sliding Glass Doors. Perspectives: Choosing and Using Books for the Classroom, 6(3), ix-xi.

Introduction

Two Gaps

Which text gets audited?

- English -> Audited, on its own[2][3]
- Spanish -> Audited, on its own[2][3]
- Code-mixed -> “noise” (filtered out or collapsed)[2][3]

Code-mixing: Two or more languages interspersed in the same sentence

how is a character represented?

WHO gets associated with a language

WHAT that language does to their portrayal

[never measured separately]

70 Bilingual Young Adult (YA) Novels · EN ⟷ ES

The space between languages goes unaudited and who vs. what goes unasked.

[2] Winata, G. I., Aji, A. F., Yong, Z.-X., & Solorio, T. (2023). The Decades Progress on Code-Switching Research in NLP: A Systematic Survey on Trends and Challenges.Findings of ACL 2023.
[3] Doğruöz, A. S., Sitaram, S., Bullock, B. E., & Toribio, A. J. (2021). A Survey of Code-switching: Linguistic and Social Perspectives for Language Technologies. ACL-IJCNLP 2021.

Research Questions

Are characters associated with Spanish positioned differently **than** characters associated with English? (CASTING)

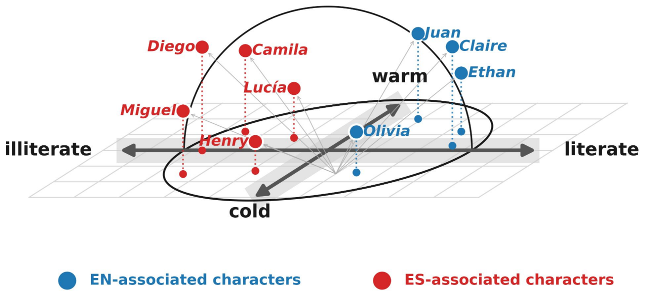
RQ1 (Casting): Specifically, do Spanish-associated character names occupy systematically different positions on sociocultural dimensions such as education, status, and tradition orientation?

Do the same characters get positioned differently **across** English, code-mixed, and Spanish? (FRAMING)

RQ2 (Framing): When a character appears in English-only, code-mixed, and Spanish-only sentences, do their projections onto sociocultural axes differ systematically?

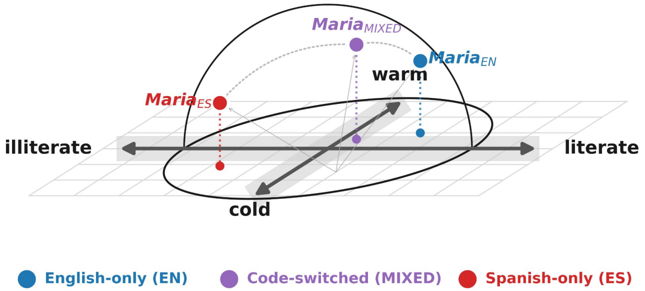
RQ1 — Casting

Different characters occupy different positions in the semantic space.



RQ2 — Framing

The same character occupies different positions by linguistic context.



Methods

From 70 Novels to Character Embeddings

70 Bilingual YA Novels → **619** Character Names → **XLM-RoBERTa** Contextual Embeddings

Each name is embedded **once per sentence**. the sentence is its context window:

name EN
name EN
name MIXED
name ES

the **same name** w/ their sentences labeled EN / MIX / ES

[···]
[···]
[···]
[···]

one embedding **per sentence**

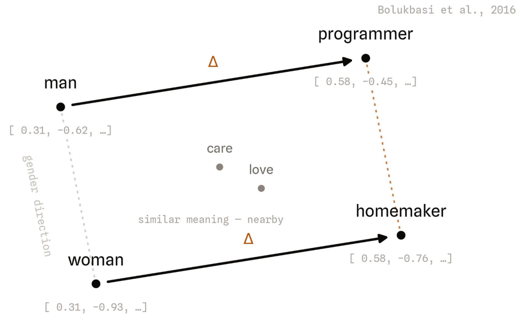
all sentences **averaged together** → [····] **one overall** embedding per character — FRAMING (RQ1)

averaged **within each context** → [····] EN [····] ES [····] MX — CASTING (RQ2)

two ways to pool - one per research question

English and Spanish live in **one shared space** – XLM-RoBERTa

why embeddings? — geometry carries meaning



v(programmer) − v(man) ≈ v(homemaker) − v(woman)

the step from **man** to **programmer**, applied to **woman**, lands on **homemaker** i.e. **geometry carries meaning, and carries bias**[4][5]

If we want to know how a character is positioned, **the embedding already holds that information**

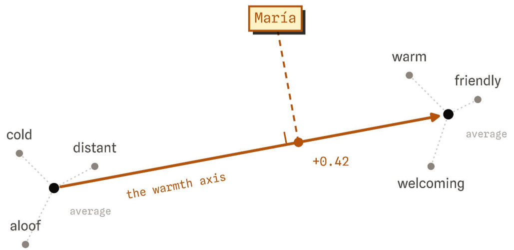
[4] Tolga Bolukbasi, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. 2016. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. Advances in Neural Information Processing Systems 29 (NeurIPS 2016).
[5] Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013. Linguistic Regularities in Continuous Space Word Representations. Proceedings of NAACL-HLT 2013, pages 746–751.

Methods

Measuring Portrayal: 25 Axes, Two Tests

[6] [7]

a character's embedding



warmth(María) = cos(v(María), avg(warm) - avg(cold))

an axis is a ruler laid through the space

schematic

25 Axes. Six Categories:

2	**Kozlowski-validated**	affluence · gender
6	**Hofstede cultural dimensions**	individualism · power distance · stereotypical gender traits · uncertainty avoidance · tradition orientation · indulgence
5	**socioeconomic status**	literacy skill · formal education · work type · status · social mobility
5	**belonging & culture**	immigration status · national origin · place belonging · cultivation · rural-urban
3	**interpersonal / relational**	**warmth** · personalismo · familism
4	**affect**	character agency · praised-criticized · fear · likability

axis word lists: **pole assignment** validated by **two independent raters** – Cohen's κ = 0.94

RQ1 · CASTING

different characters – one overall embedding each

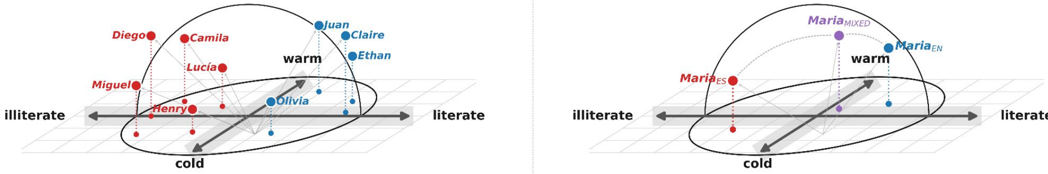
- EN-associated vs **ES-associated** characters
- **Welch's t-test**, per axis
- effect size: **Hedges' g**
- BH-FDR, q = 0.05

A RQ1 — Casting

Different characters occupy different positions in the semantic space.

B RQ2 — Framing

The same character occupies different positions by linguistic context.



● EN-associated characters ● ES-associated characters ● English-only (EN) ● Mixed-context (MIXED) ● Spanish-only (ES)

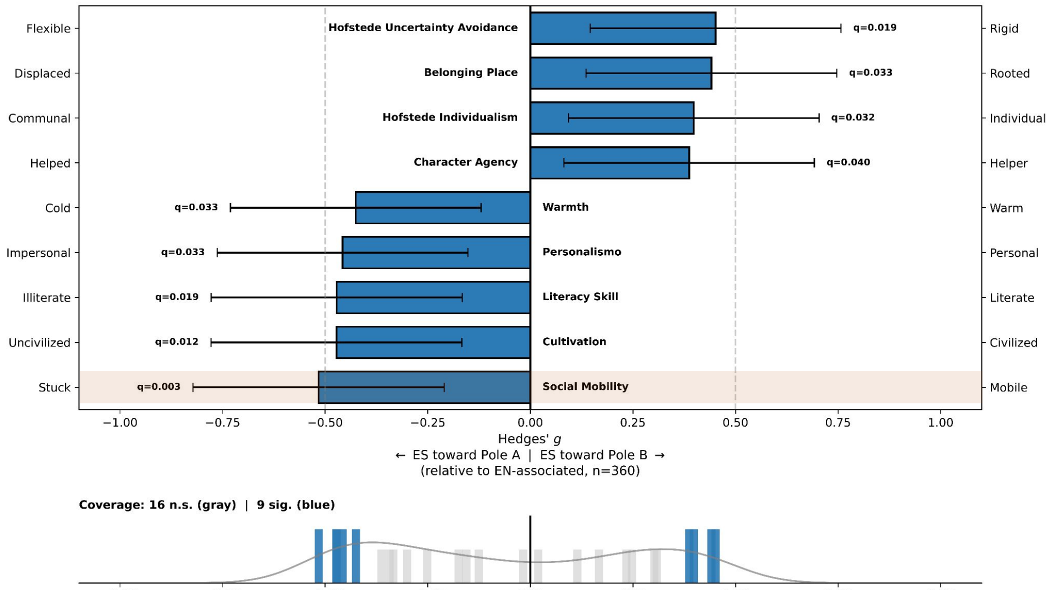
RQ2 · FRAMING

the same character – split by context: EN / MIXED / ES

- each character **compared with itself**
- **paired t-test**, per axis
- effect size: **Cohen's d_z**
- BH-FDR, q = 0.05

[6] Austin C. Kozlowski, Matt Taddy, and James A. Evans. 2019. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. American Sociological Review 84(5), 905–949.
[7] Li Lucy, Divya Tadimeti, and David Bamman. 2022. Discovering Differences in the Representation of People using Contextualized Semantic Axes. Proceedings of EMNLP 2022.

Spanish-associated characters are cast as deficit and agentic at once



(A) Hedges' g with 95% CI for the nine significant axes (q < 0.05, blue); positive g = Spanish-associated characters positioned toward pole B. **(B)** all 25 tested axes; gray = not significant.

9 of 25 axes significant

small-to-medium effects · FDR q ≤ 0.05 · ES n=47 vs EN n=360

DEFICIT framing – 2 AXES

stuck, not mobile | `largest effect: g = -0.52`

uncivilized

illiterate

impersonal

cold

rigid, not flexible

AGENTIC FRAMING - 2 AXES

rooted, not displaced

the helper, not the helped

the ninth – individualism (+0.40) – is culturally ambiguous; see panel A

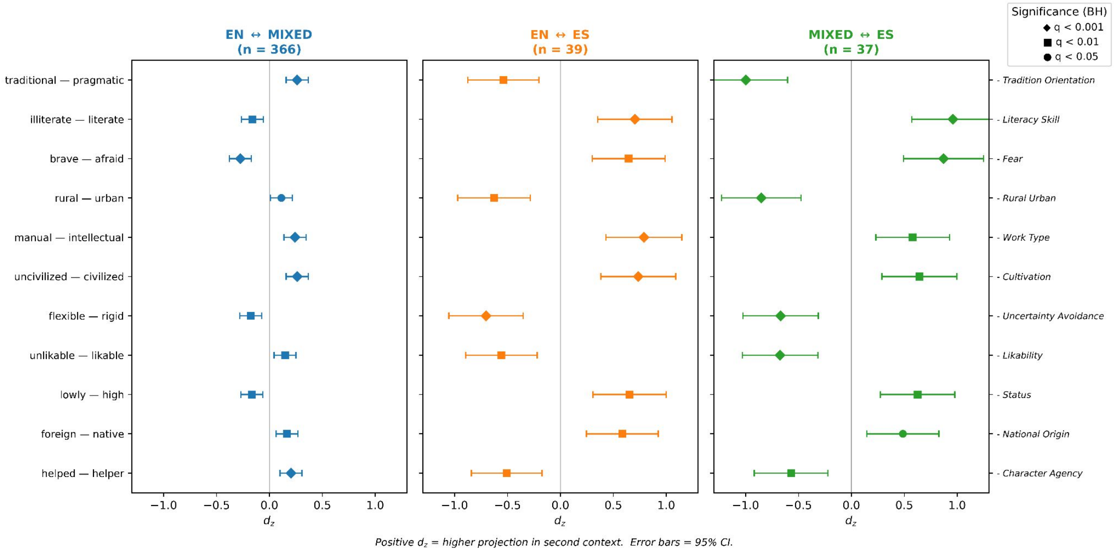
Findings

Characters held fixed - Contradiction persists

the same characters, across linguistic contexts

Elevated	**Diminished**
more **literate**	more **afraid**
more **intellectual**	less **likable**
higher **status**	more **in need of help**

Holds in **both comparisons involving ES** (EN⟷ES and MIXED⟷ES) – different axes than before, **the same ambivalent shape**



Cohen's d_z with 95% CI for the **eleven axes significant in all three pairwise comparisons** ($q < 0.05$, BH-corrected); marker shape = significance level; positive d_z = higher projection in the second context.

the contradiction shows up **between characters, and within them**

Code-Mixed Text Behaves as Its Own Distinct Register

11 of 25 axes significant in **all three** pairwise comparisons (q ≤ 0.05, BH-corrected).

On **7 of the 11**, English sits in the middle – code-mixed and Spanish contexts pull the **same characters in opposite directions**, away from the English baseline.

The Implications

Most NLP pipelines are **`monolingual-first`**: code-mixed text is **filtered out, or collapsed** into one language. Audits built on top of them can return **distorted results**, collapsing can even **flip the direction** of an association.

Our guideline: **audit register-aware**.
Treat code-mixed text as a register in its own right

Axis	Ordering
Opposite directions (7)	
Likability	ES→EN→MX
Character Agency	ES→EN→MX
Tradition Orientation	ES→EN→MX
Rural–Urban	ES→EN→MX
Literacy Skill	MX→EN→ES
Status	MX→EN→ES
Fear	MX→EN→ES
Same direction (4)	
Cultivation	EN→MX→ES
Work Type	EN→MX→ES
Uncertainty Av.	ES→MX→EN
National Origin	EN→MX→ES

Ordering of the three contexts on the 11 robust axes, pole A → pole B. **Blue/orange rows (7):** MIXED and ES diverge from English in opposite directions. **Gray (4):** MIXED falls in between. **Example – Literacy Skill:** the same character reads most *illiterate* in MIXED, intermediate in EN, most *literate* in ES.

Takeaway

Paper

All Links

TRAIL Lab

Audit the Space Between Languages *Someone Might Be Living There.*

Millions of people do. For speakers of **Spanglish, Hinglish, Taglish, and Guaraní-Jopara**, code-mixing is not an edge case. It is the norm of everyday speech. What is rare in the data is still real in the world.

Normal should not always be **normative**.

With thanks —Lauren Ciha, Nathan Han, Madison Lin, Swarit Pakala (early exploratory work) · Alina Guha, Kaycie Barron, Priyal Jain, Shanthi Kuppa (IRR validation)
and a special thank-you to **ACM FAccT**, whose travel award made it possible to present this work in person.

Tanay Nagar · Pragati Maheshwary · Shamya Karumbaiah
tanaynagar7@gmail.com