Auditing the Space Between Languages: Character Representation in Bilingual Young Adult Literature

TANAY NAGAR, Independent Researcher, India
PRAGATI MAHESHWARY, University of Wisconsin-Madison, USA
SHAMYA KARUMBAIAH, University of Wisconsin-Madison, USA

Bilingual young adult (YA) literature shapes how Latine youth understand language, culture, and identity, while also serving as training data for large language models. This dual role raises questions about what semantic associations bilingual fiction encodes, and how these associations vary across linguistic contexts. We present a socio-technical audit of character representation in 70 Spanish–English YA novels. Using contextual embeddings from XLM-RoBERTa and 25 interpretable semantic axes, we analyze 619 character names across English-only, Spanish-only, and Mixed language contexts. Our framework separates casting (which characters are associated with which linguistic contexts) from framing (how the same character's portrayal shifts by language context). We find that Spanish-associated characters occupy a contradictory semantic space, simultaneously positioned toward deficit framings (e.g., lower literacy, lower warmth) and agentic framings (e.g., rootedness, helper roles). Within-character analyses reveal systematic framing shifts: characters in Spanish-only contexts appear more literate and higher status yet also more afraid and less likable. Crucially, Mixed language contexts behave as a distinct semantic register rather than an interpolation between English and Spanish. These results demonstrate that register-aware audits are essential for accountable multilingual NLP and for understanding how bilingual narratives construct social meaning.

CCS Concepts: • **Computing methodologies** → *Lexical semantics*; • **Social and professional topics** → **Cultural characteristics**.

Additional Key Words and Phrases: Bias auditing, Multilingual NLP, Sociotechnical evaluation, Code-mixed language, Contextual embeddings, Literary corpora

ACM Reference Format:
Tanay Nagar, Pragati Maheshwary, and Shamya Karumbaiah. 2026. Auditing the Space Between Languages: Character Representation in Bilingual Young Adult Literature. In *The 2026 ACM Conference on Fairness, Accountability, and Transparency (FAccT '26), June 25–28, 2026, Montreal, QC, Canada.* ACM, New York, NY, USA, 24 pages. https://doi.org/10.1145/3805689.3806471

1 Introduction

Bilingual Young Adult (YA) literature occupies a unique position in the literary landscape. For many Latine youth, these novels represent formative encounters with code-mixed language in written form through narratives that interweave Spanish and English to reflect the linguistic realities of their communities [25, 35]. Beyond entertainment, these texts shape how young readers perceive language, culture, and identity [44]. These books model which characters speak which languages, in what contexts, and with what narrative consequences.

Authors' Contact Information: Tanay Nagar, Independent Researcher, Mumbai, India, tanaynagar7@gmail.com; Pragati Maheshwary, University of Wisconsin-Madison, Madison, USA, pmaheshwary@wisc.edu; Shamya Karumbaiah, University of Wisconsin-Madison, Madison, USA, shamya.karumbaiah@wisc.edu.

FAccT '26, Montreal, QC, Canada

ACM ISBN 979-8-4007-2596-8/2026/06
https://doi.org/10.1145/3805689.3806471

However, we know very little about what semantic associations these texts systematically encode. When a character speaks Spanish, what cultural dimensions are implicitly activated? When the same character appears in English-only vs. Spanish-only vs. code-mixed language contexts, does their portrayal shift? These questions carry weight precisely because bilingual YA literature serves as both a mirror and a window for its readers—reflecting their experiences while shaping their perceptions of self and community [6].

Answering these questions requires recognizing that language choice in bilingual literature is not incidental but structured. Most computational approaches to bias auditing treat language as a single, stable channel, while sociolinguistic research has long established that code-mixing is not random noise but a meaningful communicative practice that indexes social identities, relationships, and stances [1, 40]. In literary contexts specifically, scholars distinguish between "natural" code-switching (reflecting bilingual speech patterns) and "staged" code-switching, where authors deploy language alternation for aesthetic and narrative purposes [11, 23][1]. This distinction is critical: authors make deliberate decisions about which characters speak Spanish, in what scenes, and to what effect. Two distinct phenomena warrant exploration from this structure. The first is *casting*: some characters occupy Spanish-associated contexts more than others, and these casting decisions may systematically differ by character role, status, or narrative function. The second is *framing*: the same character may be portrayed differently depending on the linguistic context in which they appear—a character named in a Spanish-only passage may carry different semantic associations than when that same character appears in English prose. Standard bias auditing approaches that compute a single global embedding per entity collapse these two phenomena, obscuring systematic, context-conditioned variation in character representation [8, 10].

To audit bilingual YA literature from casting and framing lenses, we draw on methods from computational semantic analysis that treat word embeddings as windows into the semantic associations encoded in text. Caliskan et al. [10] introduced the Word Embedding Association Test (WEAT), demonstrating that static embeddings replicate implicit human biases across dimensions including race, gender, and age. Kozlowski et al. [32] extended this insight by developing a *semantic axis method* that defines cultural dimensions as directions in embedding space, computed from antonym pairs, allowing any entity to be projected onto interpretable dimensions such as affluence, status, or warmth. We apply this framework to contextual embeddings from XLM-RoBERTa [16], which generates different representations for the same token depending on surrounding context. This context-sensitivity is essential for our research questions: we can compute separate embeddings for the same character name when it appears in English-only, code-mixed, or Spanish-only sentences, then compare their positions on semantic axes to detect systematic framing shifts.

The imperative to conduct such an audit stems from two converging concerns that place this work adequately within the domain of computational systems and their social impacts. From an NLP perspective, literary corpora are not merely objects of study but they act as training data. BookCorpus, a collection of approximately 7,000 self-published books, was used to train foundational models including GPT and BERT [2, 53]. Recent audits have revealed significant biases in this corpus, including genre skews and problematic content [2]. More broadly, large language models ingest literary text through web scrapes, digitized books, and curated collections, implying that associations encoded in fiction—including bilingual fiction—propagate into systems used for generation, classification, sentiment analysis, and content moderation [3]. For instance, if characters appearing in Spanish context in literary corpora are systematically stereotyped or framed as less educated or more rural, these associations may surface in downstream applications that process or generate text about Latine individuals and communities. From a reader-impact perspective, bilingual youth who encounter deficit framings of characters like themselves may internalize these representations [50]. The stereotype content model suggests that repeated exposure to portrayals emphasizing particular traits—warmth, competence, status—shapes social perception

[1]Throughout this paper, we use *code-switching*, *code-mixing*, and *code-mixed language* interchangeably to refer to sentences containing material from both English and Spanish—our operational MIXED category; our usage is descriptive only of the corpus feature we measure.

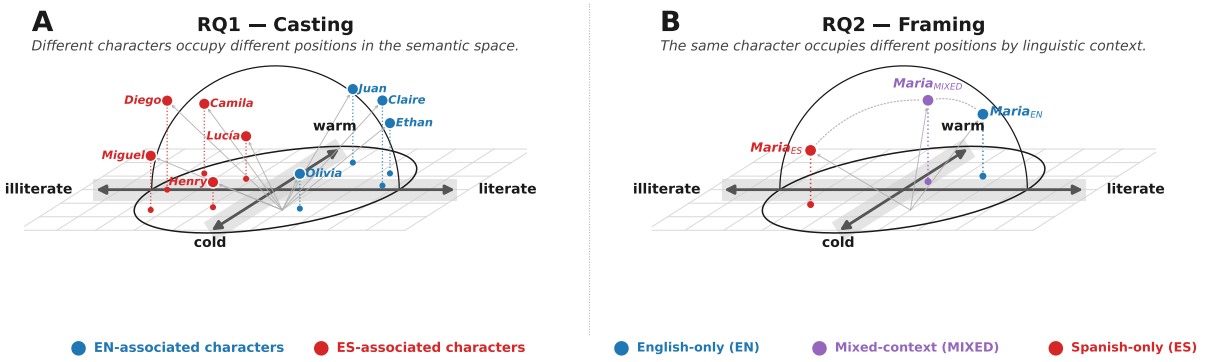

Fig. 1. **(A)** Different characters appear in different regions of the semantic space, depending on the language context they're used. **(B)** The same character embedded in three contexts (English, Mixed, Spanish) lands in three different positions.

along predictable dimensions [21]. Bilingual YA literature thus sits at a socio-technical intersection: the same texts that shape young readers' self-perception also feed the computational systems that increasingly mediate information access and social interaction.

Despite this dual significance, prior work on embedding-based bias measurement exhibits three limitations that leave bilingual literary texts under-examined. First, the vast majority of bias auditing research focuses on monolingual English text, with comparatively little attention to multilingual corpora or code-mixed corpora [30, 47]. When non-English languages are examined, they are typically treated in isolation rather than in contact situations. Second, existing approaches typically compute aggregate representations—a single embedding per entity—that obscure within-entity variation across contexts. A character who appears in both English and Spanish sentences receives one averaged representation, collapsing potentially meaningful differences in how that character is framed across linguistic contexts. Third, and most critically, no prior work disentangles *who gets associated with a language* from *what that language does to their portrayal*. Code-mixed language, when addressed in NLP at all, is typically treated as noise to be filtered, normalized, or collapsed into monolingual categories rather than as a meaningful register warranting distinct analysis [20, 52]. This leaves a methodological gap: we lack tools to bridge computational auditing methods with the sociolinguistic reality that code-mixed language carries social meaning and may systematically alter how characters are framed.

We address this gap through a systematic audit of character representation in 70 Spanish-English young adult novels featuring Latine characters and themes. We extract contextual embeddings using XLM-RoBERTa [16], a multilingual transformer model capable of representing multilingual text without requiring language-specific preprocessing. We then project character name embeddings onto 25 semantic axes spanning socioeconomic status, cultural dimensions, and narrative role, drawing on validated dimensions from Kozlowski et al. [31], Hofstede's cultural framework [27], and others mentioned in section 2.5. Our key methodological innovation is a *three-lect* analytical lens that treats English-only (EN), code-mixed (MIXED), and Spanish-only (ES) contexts as distinct registers rather than collapsing them into a single representation. This design directly operationalizes the sociolinguistic insight that code-mixed language constitutes a meaningful structured linguistic variety [39], enabling us to examine both casting (which characters are associated with which linguistic contexts) and framing (how the same character's semantic positioning shifts across contexts).

Through this approach we answer two main research questions (Fig. 1 illustrates these two questions):

- **RQ1 (Casting):** How do characters associated with Spanish linguistic contexts differ in their semantic positioning from characters associated with English contexts? Specifically, do Spanish-associated character

names occupy systematically different positions on sociocultural dimensions such as education, status, and tradition orientation?
- **RQ2 (Framing):** Does the same character's semantic positioning shift across linguistic contexts? When a character appears in English-only, code-mixed, and Spanish-only sentences, do their projections onto sociocultural axes differ systematically?

Our analysis yields three key contributions. First, we document a *contradiction in casting*: Spanish-associated characters are simultaneously positioned toward deficit framings (less mobile, less literate, colder) and agentic framings (helpers rather than helped, rooted rather than displaced), a pattern consistent with mixed stereotype content [21] that complicates simple notions of "biased" versus "unbiased" representation. Second, we identify systematic *framing shifts*: the same character's semantic positioning changes significantly across linguistic contexts, suggesting that global embeddings obscure meaningful variation that could affect downstream NLP applications processing multilingual text. Finally, we provide evidence that mixed language contexts constitute a *distinct semantic register*–not an intermediate blend of English and Spanish, but a space with its own characteristic associations. Methodologically, we propose a multi-lect framework that attempts to disentangle casting from framing and treats code-mixing as a meaningful variation rather than noise. This approach may be useful for auditing other multilingual corpora and understanding downstream effects on computational systems.

2 Methods

2.1 Dataset and Corpus Creation

We compiled a corpus of 70 bilingual young adult (YA) novels/books featuring Latine characters and themes, published predominantly from 2018 to 2019. The text spans multiple Latine backgrounds, including Mexican-American, Puerto Rican, Cuban-American, Central American, Argentine-American, and Dominican-American. The text in books includes: (i) predominantly English prose, (ii) Spanish language passages (usually dialogue, internal monologue, cultural expressions), and (iii) code-mixed text, where English and Spanish are interspersed within sentences (intrasentential Code-Mixing).

The text was digitized via Optical Character Recognition (OCR) using the Google Cloud Vision API. Post-OCR error detection and correction employed a custom pipeline combining rule-based methods with SymSpell [22], a spelling correction algorithm based on the symmetric-delete approach. Further validation against a bilingual dictionary and manual verification of a 1% random sample of sentences stratified by book (≈2,680 sentences) was performed too.

The final corpus comprised 268,078 sentences and 72,749 unique vocabulary tokens. Sentence segmentation was performed using pysbd [43] (Python Sentence Boundary Disambiguation) with custom pre-processing rules to preserve spanish-specific punctuation and diacritics (see Appendix A).

Why a single bilingual corpus? Our design requires analyzing English-only, Spanish-only, and code-mixed sentences within the same narratives rather than across separate monolingual corpora. Two reasons motivate this choice. First, our within-character paired analysis (RQ2, Section 2.2) compares the *same* character across linguistic contexts; this is only possible when a character appears in multiple lects within a single book, which separate English and Spanish corpora cannot support by construction. Second, separate monolingual corpora would confound linguistic context with author, genre, and publication context—making casting comparisons (RQ1) unreliable even on their own terms. Finally, our finding that code-mixed text constitutes a distinct semantic register would not have been observable without code-mixed sentences in the corpus; a pattern that is easily erased when code-mixed text is underrepresented in multilingual training data and evaluation benchmarks, as documented by recent evidence [19, 52].

2.2 Language Variety Labeling: English vs. Spanish vs. Code-mixed

Each sentence was classified into one of three linguistic contexts: `English`, `Spanish`, or `code-mixed`, utilizing a two-staged pipeline:

Stage 1: Language detection. We apply the Lingua [37] language detection library for each sentence in order to obtain confidence scores for English and Spanish.

Stage 2: Three-Lect Classification.

Standard Language Detectors perform poorly on code-mixed text, often mis-classifying sentences with a single Spanish word as code-mixed text. To address this, we supplemented Lingua's confidence scores with rule-based heuristics. A sentence was classified as code-mixed only if it contained evidence of *both* languages i.e. we required evidence of both languages' grammar, such as English articles and prepositions (the, to, with) co-occurring with Spanish equivalents (el, de, con), before classifying a sentence as code-mixed. Sentences were classified as Spanish only when Lingua showed high Spanish confidence with minimal English confidence. The resulting distribution was English (97.64%), code-mixed (2.04%), and Spanish (0.31%). This asymmetry reflects the empirical reality of bilingual YA literature, where Spanish typically appears in dialogue, internal monologue, and culturally marked scenes rather than sustained prose [11, 23]; we discuss its implications for statistical power in Section 4.4. Full classification rules and lexicon are provided in Appendix B.

2.3 Entity Extraction

Named entities from text were extracted using a bilingual named entity recognition (NER) pipeline based on sentence-language classification (Section 2.2):

- **EN sentences:** spaCy's [29] English model (`en_core_web_sm`), with Spanish model fallback only if no PERSON entities are detected.
- **ES sentences:** spaCy's Spanish model (`es_core_news_sm`), with English model fallback only if no PERSON entities are detected.
- **MIXED sentences:** Both models are applied (with duplicate detection to avoid double-counting).

Validation for extracted names was done using two validated name lists: (i) the NLTK names corpus [5] (7,533 English names) and the Spanish census from Spain's Instituto Nacional de Estadística [48] (16,691 names). Names were canonicalized by lowercasing and stripping diacritics for matching purposes (e.g., *María* → *maria*) while preserving original forms in the corpus. A denylist excluded Spanish function words frequently misclassified as names (*del*, *la*, *de*).

Names were filtered to require: (i) validation against at least one name list, (ii) a minimum of 10 corpus mentions, and (iii) a minimum of 3 characters. Spelling variants (e.g., *María*/*Maria*) were merged under canonical keys. Kinship terms (e.g., *Mamá*, *Tío*) were flagged for optional exclusion in downstream analyses. The final validated dataset comprised **619 unique character names** with **71,388 total mentions**. Out of these 619 names, 136 name appeared exclusively in the English name list, 80 appeared exclusively in the Spanish name list, and 403 appeared in both the lists.

2.4 Contextual Embedding Representation

We produce contextual token embeddings using **XLM-RoBERTa-base**[16] (XLM-R), a multilingual Transformer model pre-trained on large-scale CommonCrawl data across 100 languages, including English and Spanish. Our rationale for choosing XLM-R, a model designed for cross-lingual transfer, comes from prior work which provides evidence that semantically similar items across languages can occupy comparable regions of representation space while still being able to encode language-specific usage patterns [41] [13].

For every occurrence of a token, including previously extracted named entities, we extract hidden layer representations from XLM-R and form a single contextual embedding per token by aggregating the model's last

hidden layers. We follow conventional heuristics for extracting feature-based token/span embeddings from a pre-trained transformer, summing information from the final 4 hidden layers, producing a single vector per token occurrence [18].

Since our research question necessitates stable estimates of how tokens are framed across many local contexts, we aggregate the embeddings for each token occurrence via mean pooling based on scope. For **RQ1**, we pool across all occurrences to create global embeddings. For **RQ2**, we pool separately within each language bucket (English-only, Spanish-only, code-mixed sentences), resulting in up to three language-specific/lect-specific embeddings per token. This structure compares the same tokens across language contexts, so statistical differences reflect context rather than differences between tokens themselves.

2.5 Sociocultural Axes Construction

Our methodological approach is closely inspired by two lines of work. We constructed semantic axes inspired by the antonym pair method of Kozlowski et al. [32]. Their approach defines a semantic dimension (e.g., *illiterate*↔*literate*) as an axis in the embedding space, allowing a numerical quantification of where a character/name falls on this dimension, building on prior work demonstrating that social biases and cultural associations in upstream data is encoded as vector directions in word embeddings [8, 10, 24, 31]; of our 25 axes, 9 correspond in full (Affluence, Gender) or in part (Education, Status, Cultivation, Warmth, Likability, Praised–Criticized, Work Type) to dimensions defined in Kozlowski's line of work. Secondly, from Lucy et al. [36], we adopt the extension of Kozlowski's axis method to BERT-family contextual embeddings for the study of people: like Lucy et al. [36], we (i) use the averaged-pole difference as the axis vector, (ii) obtain multiple contextual embeddings per token (one per sentence occurrence) rather than a single type-level vector, (iii) mean-pool these occurrences, and (iv) project onto axes via cosine similarity. The key differences from Lucy et al. [36] are that we use XLM-RoBERTa (multilingual) rather than BERT-base, and we curate antonym pole sets rather than WordNet synsets.

Axis Construction. For each semantic axis, we define two pole word lists of equal size k: a negative pole $\{w_1^-, \ldots, w_k^-\}$ and a positive pole $\{w_1^+, \ldots, w_k^+\}$. These represent the positive and negative poles of a dimension. The axis vector ($\vec{a}$) is then computed as the difference between the averaged embeddings of each pole:

$$\vec{a} = \frac{1}{k}\sum_{i=1}^{k} \vec{w}_i^+ - \frac{1}{k}\sum_{i=1}^{k} \vec{w}_i^- \tag{1}$$

where $\vec{w}_i^+$ and $\vec{w}_i^-$ represent the normalized embeddings of the words from the positive and negative poles, respectively. This formulation is mathematically equivalent to averaging paired differences when pole sizes are equal. Kozlowski et al. [32] showed empirically that random re-pairing of synonyms across poles barely degraded performance, and that the grouped pairs method produced substantively similar results to antonym pairs. This indicated that pole set coherence matters more than one-to-one antonym structure in terms of axis quality. While Kozlowski et al. [32] used static word embeddings (word2vec), we apply the same geometric principle to contextual embeddings from XLM-RoBERTa, following recent work demonstrating that semantic axes remain valid for measuring cultural associations in transformer representations [33, 36]. We validate axis coherence by requiring (i) low cosine similarity between pole words ($\overline{\cos}(w^+, w^-) < 0.3$), and (ii) consistent directionality, such that individual pole words project in the expected direction. Pole words not attested in the corpus are excluded to ensure axes reflect corpus-relevant structure.

Name Projection. Each character name n is projected onto axis $\vec{a}$ via cosine similarity:

$$s_{n,a} = \cos(\vec{n}, \vec{a}) = \frac{\vec{n} \cdot \vec{a}}{\|\vec{n}\|\|\vec{a}\|} \tag{2}$$

leading us to a scalar score $s_{n,a} \in [-1, 1]$ indicating the name's positionality on that dimension. A positive value thus indicates an inclination towards the positive pole, and a negative value indicates proximity towards a negative pole, and values near 0 indicate neutrality on that particular sociocultural dimension.

We define multiple semantic axes relevant to the context of our Young Adult Books, spanning six categories:

- **Kozlowski Validated (2 axes)**: Core dimensions validated by Kozlowski et al. [32] were included as is, i.e.,*(1) affluence axis*: poor↔rich; *(2) gender axis*: man↔woman.
- **Hofstede Cultural Dimensions (6 axes)**: Cross-cultural value orientations from Hofstede's established framework [28], identifying societal value continua were adapted as *(1) individualism axis*: communal↔individual; *(2) power distance axis*: equal↔authority; *(3) Stereotypical gender traits axis*: nurturing↔competitive; *(4) uncertainty avoidance axis*: flexible↔rigid; *(5) tradition orientation axis*: traditional↔pragmatic; *(6) indulgence axis*: restrained↔indulgent.
- **Socioeconomic Status (5 axes)**: Dimensions that capture class, education, occupational prestige, mobility in social ladder or social hierarchies, i.e., key characteristics of social stratification that proxy cultural capital [9] and shape public perception of groups [21] were included as *(1) literacy skill axis*: illiterate↔literate; *(2) formal education axis*: uneducated↔educated; *(3) work type axis*: manual↔intellectual; *(4) status axis*: lowly↔high; *(5) social mobility axis*: stuck↔mobile [7, 46].
- **Belonging & Culture (5 axes)**: Dimensions capturing legal status, place attachments, and civilizational othering [12, 14], as well as rural-urban meanings [51] were included as, *(1) immigration status axis*: illegal↔legal; *(2) national origin axis*: foreign↔native; *(3) place belonging axis*: displaced↔rooted; *(4) cultivation axis*: uncivilized↔civilized; *(5) rural–urban axis*: rural↔urban.
- **Interpersonal/Relational Axes (3 axes)**: Dimensions corresponding to documented cultural values and constructs, including personalismo [17], and familismo [42] were included as: *(1) warmth*: cold↔warm; *(2) personalismo axis*: impersonal↔personal; *(3) familism axis*: individual↔family.
- **Affect (4 axes)**: Dimensions capturing characters' agency [49], social evaluations or emotions —central to how narratives position characters: *(1) character agency axis*: helped↔helper; *(2) praised–criticized axis*: criticized↔praised; *(3) fear axis*: brave↔afraid; *(4) likability axis*: unlikable↔likable.

Axes were developed collaboratively by two researchers through an iterative process: for each dimension, multiple candidate pole words were suggested, evaluated for semantic coherence, and then refined. This multi-researcher approach reduces individual bias in axis construction. To further validate the final pole word lists, we conducted an inter-rater reliability (IRR) study with two raters on a random 35% subsample of axes. Words-to axis-association reached substantial agreement ($\kappa = 0.67$). Pole assignment on words that the raters agreed on being associated to the axis achieved near-perfect agreement (Cohen's $\kappa = 0.94$) (see Appendix D for full results). The complete list of all axes, seed/pole words can be found in Appendix C.

These axes serve as the measurement instruments for both research questions: for **RQ1** (casting), we project each character's global embedding onto the axes to compare Spanish-associated and English-associated characters; for **RQ2** (framing), we project each character's lect-specific embeddings onto the same axes to measure within-character shifts across linguistic contexts.

2.6 Spanish Association score (ρ) and Character Grouping

To quantify every character's association with Spanish linguistic contexts in each book, we compute a Spanish Exposure Score/Proportion:

$$\rho_i = \frac{s_{i,\text{MIXED}} + s_{i,\text{ES}}}{s_{i,\text{total}}} \tag{3}$$

where $s_{i,\text{MIXED}}$ and $s_{i,\text{ES}}$ are the counts of unique sentences in which character i appears in code-mixed and Spanish linguistic contexts respectively, and $s_{i,\text{total}}$ is the total count of unique sentences containing/mentioning that character. The metric $\rho \in [0, 1]$ is continuous: $\rho = 0$ indicates the character appears exclusively in English-context sentences. Logically, $\rho = 1$ indicates an exclusive appearance in Spanish or code-mixed sentences. The denominator is the character's own total sentence count. Thus, ρ is inherently a within-character proportion

rather than a corpus-level frequency. Consequently, ρ is not affected by the overall scale imbalance between English, code-mixed, and Spanish text in the corpus.

It is important to clarify that ρ is not (i) a measure of the character's ethnicity or nationality in a narrative, (ii) the etymological origin of the name or (iii) any authorial labels. It is strictly a distributional measure of which linguistic context a character occupies within the corpus. For instance, characters with Spanish-originated names could still have a lower ρ if their predominant usage is in an English narrative context (for example, a Spanish character appearing in 10 sentences with 8 of them in English contexts yields $\rho = 0.20$, independent of corpus-wide EN/MIXED/ES distribution).

Threshold Selection (RQ1). To create a contrast groups for RQ1, we define the following thresholds:

EN-associated: $\rho \leq 0.01$, $n = 360$; **ES-associated:** $\rho \geq 0.15$, $n = 47$

We selected this threshold to prioritize statistical power for the ES-associated group while maintaining high distinctiveness (median $\rho_{\text{ES}} = 0.24$ vs. median $\rho_{\text{EN}} = 0$). Because Spanish-only sentences account for only 0.31% of the corpus (Section 2.2), the resulting sample sizes for ES-based comparisons are small; we return to this limitation in Section 4.4.

Within-Name Grouping (RQ2). We identify/bucket characters based on their appearance in multiple linguistic contexts. With ≥ 5 number of mentions requirement for reliability, name appearances per bucket were:

- EN vs. MIXED comparison: $n = 366$ names
- EN vs. ES comparison: $n = 39$ names
- MIXED vs. ES comparison: $n = 37$ names

A note on terminology: throughout the paper we use **Spanish-associated characters** to refer to the *subset of characters* whose overall appearance pattern skews toward Spanish linguistic contexts (high ρ), in contrast to characters **appearing in Spanish-only contexts**, which refers to *any character's sentences* that were classified as Spanish-only regardless of their overall ρ. The former is a property of the character (used in RQ1, casting); the latter is a property of individual mentions (used in RQ2, framing).

2.7 Statistical Testing and Multiple Comparisons

For **Research Question 1**, we employ Welch's t-test to compare the projections created utilizing our semantic axes between the EN-associated ($\rho \leq 0.01$, $n = 360$) and the ES-associated ($\rho \geq 0.15$, $n = 47$) groups. We choose Welch's t-test because it doesn't assume equal variances between groups. Further, we report Hedges' g as our standardized effect size measure because it applies a small-sample bias correction to Cohen's d [26]. We test each semantic axis independently.

For **Research Question 2**, we utilize paired t-tests in order to examine within-character shifts in semantic positioning across the three linguistic context labels (`EN`, `ES`, `MIXED`). For each character with sufficient mentions in relevant linguistic buckets (≥ 5 mentions per context), we compute separate embeddings for English, code-mixed, and Spanish contexts, and then test whether semantic axis projections differ systematically across groups. We conduct three pairwise comparisons: English vs. MIXED ($n = 366$ characters), English vs. Spanish ($n = 39$ characters), and MIXED vs. Spanish ($n = 37$ characters). The pairwise-design controls for individual character baseline differences, isolating the effect of linguistic context. We report Cohen's d_z, the mean differences divided by the standard deviation of differences.

We chose pairwise t-tests over repeated measures ANOVA for RQ2 because our data structure is imbalanced: not all characters appear in all three contexts with sufficient mentions for statistical significance.

To control for multiple testing, we apply Benjamini-Hochberg (BH) False Discovery (FDR) [4] correction separately with each research question. For **RQ1**, we correct across all 25 axes. For RQ2, we apply FDR within each pairwise comparison (25 axes per comparison), rather than all 3×25 tests simultaneously, as each comparison addresses a distinct sub-question. Finally, we set the FDR significance threshold to $q = 0.05$.

Table 1. RQ1: Semantic positioning of Spanish-associated (n=47, $\rho \geq 0.15$) vs. English-associated (n=360, $\rho \leq 0.01$) character names. Sorted by $|g|$.

Axis	Dimension	Δ Mean	g	95% CI	p	q
Social Mobility	stuck — mobile	-0.032	-0.52**	[-0.82, -0.21]	0.0001	0.003
Cultivation	uncivilized — civilized	-0.040	-0.47*	[-0.78, -0.17]	0.0010	0.012
Literacy Skill	illiterate — literate	-0.025	-0.47*	[-0.78, -0.17]	0.0031	0.019
Personalismo	impersonal — personal	-0.031	-0.46*	[-0.76, -0.15]	0.0093	0.033
Uncertainty Avoidance	flexible — rigid	+0.026	+0.45*	[+0.15, +0.76]	0.0024	0.019
Place Belonging	displaced — rooted	+0.024	+0.44*	[+0.14, +0.75]	0.0084	0.033
Warmth	cold — warm	-0.027	-0.43*	[-0.73, -0.12]	0.0107	0.033
Individualism	communal — individual	+0.023	+0.40*	[+0.09, +0.70]	0.0065	0.032
Character Agency	helped — helper	+0.021	+0.39*	[+0.08, +0.69]	0.0143	0.040

Sig: *** $q < 0.001$, ** $q < 0.01$, * $q \leq 0.05$. Positive g indicates ES-associated names toward the second pole.

3 Results

For **RQ1** we report effect sizes as Hedges' g (a small-sample-corrected standardized mean difference) with 95% confidence intervals. We interpret effect sizes following conventional thresholds: $|g| < 0.2$ negligible, $0.2 \leq |g| < 0.5$ small, $0.5 \leq |g| < 0.8$ medium, and $|g| \geq 0.8$ large [15, 45]. All stated q values reflect Benjamini-Hochberg correction for false discovery rates (FDR). For **RQ2**, we present Cohen's d_z (the mean difference divided by the SD of differences) as the effect size for within-subject differences/comparisons. We utilize the same interpretive thresholds as RQ1.

3.1 RQ1: Character Positioning by Language Association

We compare the semantic positioning of Spanish-associated ($n = 47$, $\rho \geq 0.15$) and English-associated ($n = 360$, $\rho \leq 0.01$) characters across the 25 aforementioned semantic axes. A positive Hedges' g value indicates that Spanish-associated characters are positioned toward pole B (the second term in each dimension); a negative g-value indicates the opposite: positioning towards pole A, relative to English-associated characters.

After FDR correction, nine axes show significant differences ($q \leq 0.05$) with effect sizes in the small to medium range ($0.39 \leq |g| \leq 0.52$) (see Table 1). Fig. 2 summarizes the aforementioned results: Panel A reports Hedges' g with 95% Confidence Intervals for the 9 significant axes ($q \leq 0.05$), while Panel B shows the distribution of effects across all the 25 tested axes, with the significant axes highlighted.

Deficit Framing on Mobility, Literacy, and Social Dimensions. Six axes show ES-associated characters positioned toward traits typically associated with deficit stereotypes. The largest effect emerges on *social mobility* ($g = -0.52$, $q = 0.003$): Spanish-associated characters were positioned as *stuck* rather than *mobile*. Similar patterns emerge for *cultivation* ($g = -0.47$, $q < .05$; positioned toward *uncivilized*), *literacy skill* ($g = -0.47$, $q < .05$; positioned toward *illiterate*), *personalismo* ($g = -0.46$, $q < .05$; positioned toward *impersonal*), *warmth* ($g = -0.43$, $q < .05$; positioned toward *cold*), and *uncertainty avoidance* ($g = +0.45$, $q < .05$; positioned toward *rigid* rather than *flexible*).

Agentic Framing on Belonging and Helper Roles. Two axes show ES-associated characters positioned toward traits reflecting agency and positive social roles. Spanish-associated characters appear more *rooted* than *displaced* on *place belonging* ($g = +0.44$, $q < .05$), and as *helpers* rather than *helped* ($g = +0.39$, $q < .05$).

Culturally Ambiguous Positioning. On Hofstede's *individualism* axis ($g = +0.40$, $q < .05$), Spanish-associated characters appear more *individual* than *communal*. The valence of this positioning is culturally dependent: while individualism is valorized in mainstream U.S. culture, collectivism and *familismo* are central

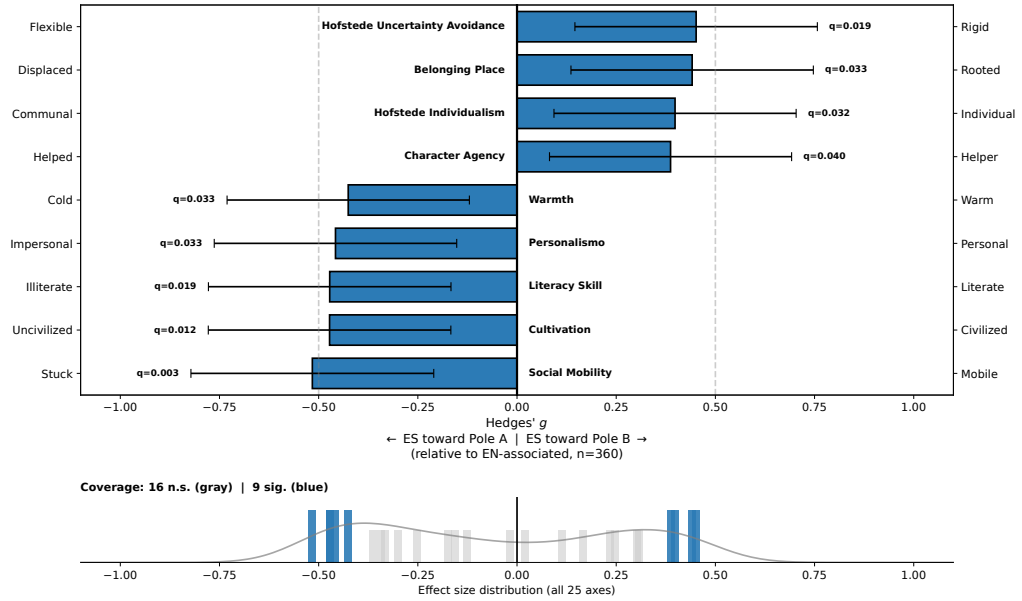

Fig. 2. RQ1: Semantic positioning by Spanish-association. **(A)** Forest plot showing effect sizes (Hedges' g) with 95% confidence intervals for nine significant axes ($q < 0.05$, blue). Positive g indicates Spanish-associated characters positioned toward pole B; negative g toward pole A. **(B)** Distribution of effect sizes across all 25 tested axes; blue bars indicate significant effects, gray bars indicate non-significant.

values in many Latine cultural frameworks [38], making this finding difficult to interpret as straightforwardly positive or negative.

Negligible Effects on Kozlowski's Axes. On Kozlowski's validated axes from the original paper, neither *gender* ($g = -0.02$, $q = 0.90$) nor *affluence* ($g = +0.30$, $q = 0.09$) showed significant differences. Spanish-associated characters in this corpus do not occupy systematically gendered or rich-poor semantic positions relative to English-associated characters. q values for all axes, including those not reaching significance, appear in Appendix E.

3.2 RQ2: Within-Character Semantic Shifts Across Linguistic Contexts

To examine whether the same character is represented differently across linguistic contexts, we computed character embeddings separately for each of the three lects: English-only (EN), code-mixed (MIXED), and Spanish-only (ES). Subsequently, we project these embeddings onto the 25 semantic axes defined in Section 2.5 and conduct paired t-tests comparing axis projections for the same characters across contexts. This resulted in three sets of pairwise comparisons: EN vs. MIXED ($n = 366$ characters that appear in both English and Mixed contexts), EN vs. ES ($n = 39$), and MIXED vs. ES ($n = 37$). These n values reflect a minimum threshold of 5 mentions per context, required to ensure stable per-lect embeddings. Cohen's d_z quantifies the direction and magnitude of within-character shifts along each semantic dimension as a function of language context. Thus, positive values indicate higher ratings along pole B (the second term in each dimension) in the second language context listed in the pair (e.g., in EN vs. MIXED positive d_z indicates higher ratings along pole B in MIXED), and negative values indicate higher ratings towards pole A in the second language context. We apply Benjamini-Hochberg FDR correction separately within each comparison set i.e., 25 tests per comparison reporting the q values.

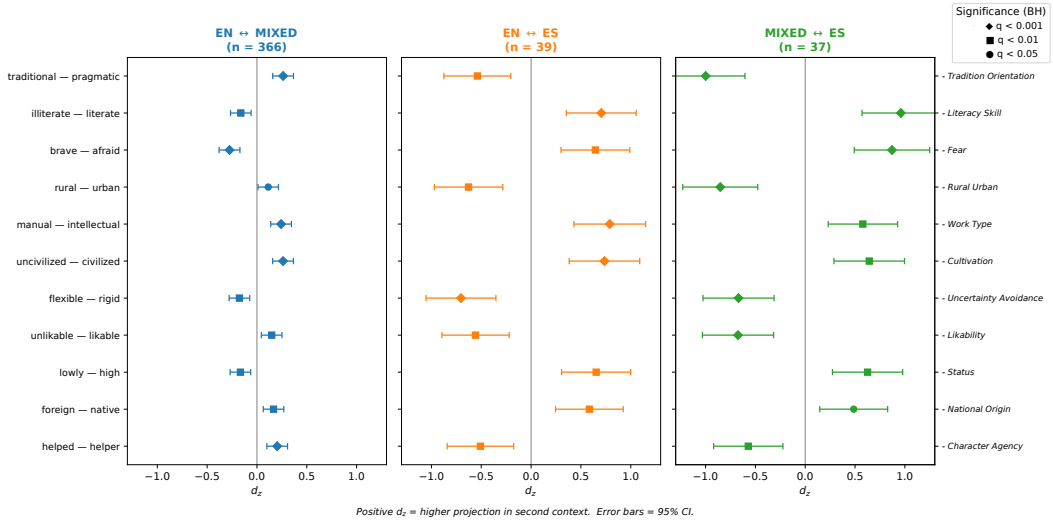

Fig. 3. RQ2: Within-character semantic shifts across linguistic contexts. Effect sizes (Cohen's d_z) with 95% confidence intervals for eleven axes significant ($q < 0.05$, BH-corrected) in all three pairwise comparisons. Marker shapes indicate significance level: $q < 0.001$, $q < 0.01$, $q < 0.05$. Positive d_z indicates higher projection in the second context.

Table 2. RQ2: Within-character semantic shifts across linguistic contexts. Showing axes significant ($q \leq 0.05$) in all three pairwise comparisons.

		EN ↔ ES		EN ↔ MIXED		MIXED ↔ ES	
Axis	Dimension	d_z	Sig	d_z	Sig	d_z	Sig
Tradition Orientation	traditional — pragmatic	-0.54	**	+0.26	***	-1.00	***
Literacy Skill	illiterate — literate	+0.70	***	-0.16	**	+0.96	***
Fear	brave — afraid	+0.65	**	-0.28	***	+0.87	***
Rural Urban	rural — urban	-0.63	**	+0.11	*	-0.85	***
Work Type	manual — intellectual	+0.79	***	+0.24	***	+0.58	**
Cultivation	uncivilized — civilized	+0.74	***	+0.26	***	+0.64	**
Uncertainty Avoidance	flexible — rigid	-0.70	***	-0.18	**	-0.67	***
Likability	unlikable — likable	-0.56	**	+0.15	**	-0.67	***
Status	lowly — high	+0.65	**	-0.17	**	+0.63	**
National Origin	foreign — native	+0.58	**	+0.17	**	+0.49	*
Character Agency	helped — helper	-0.51	**	+0.20	***	-0.57	**

Note: Sig: *** q < 0.001, ** q < 0.01, * $q \leq 0.05$. Positive d_z indicates higher projection in the second context. Sample sizes: EN↔ES (n=39), EN↔MIXED (n=366), MIXED↔ES (n=37).

Across the 25 tested axes, 23 showed significant within-character shifts for EN vs. MIXED, 15 for EN vs. ES, and 16 for MIXED vs. ES. Eleven axes reached significance in all three pairwise comparisons; we focus on these robust axes below, organizing results by the category framework from Section 2.5. Effect magnitudes were smaller for EN vs. MIXED (median $|d_z| = 0.23$) than for comparisons involving Spanish-only contexts (median $|d_z| = 0.48$ for EN vs. ES; 0.49 for MIXED vs. ES), likely reflecting the larger linguistic distance between

English-only and Spanish-only contexts (see Fig. 3). The complete set of results for all axes, including those not reaching significance, can be found in Appendix F.

Hofstede Cultural Dimensions. Two Hofstede axes show significant shifts across all three comparisons. On *tradition orientation*, same characters in ES contexts are positioned dramatically more toward *traditional* than in MIXED contexts ($d_z = -1.00$, the largest effect observed) or EN contexts ($d_z = -0.54$). On *uncertainty avoidance*, characters shift toward *flexible* in ES contexts relative to both EN ($d_z = -0.70$) and MIXED ($d_z = -0.67$) contexts. These patterns suggest Spanish-only contexts evoke heritage and traditional cultural framings while simultaneously positioning characters as more adaptable.

Socioeconomic Status. Three axes in this category, *literacy skill*, *work type*, and *status*, show consistent positive shifts in ES contexts. Characters appearing in ES contexts are positioned as more *literate* ($d_z = +0.70$ EN vs. ES; $d_z = +0.96$ MIXED vs. ES), more *intellectual* ($d_z = +0.79$ EN vs. ES; $d_z = +0.58$ MIXED vs. ES), and higher *status* ($d_z = +0.65$ EN vs. ES; $d_z = +0.63$ MIXED vs. ES) compared to the same characters appearing in other linguistic contexts. Notably, EN vs. MIXED comparisons show a mixed pattern: characters' positioning shifts toward *illiterate* ($d_z = -0.16$) and *lowly status* ($d_z = -0.17$) in MIXED relative to EN contexts, while shifting towards *intellectual work type* ($d_z = +0.24$).

Belonging & Culture. All three of the belonging & culture axes show significant results in all three pairwise comparisons. On *national origin*, characters appearing in ES contexts appear more *native* than *foreign* relative to both EN ($d_z = +0.58$) and MIXED ($d_z = +0.49$) contexts. On *cultivation* (uncivilized–civilized), characters shift toward *civilized* in ES contexts ($d_z = +0.74$ EN vs. ES; $d_z = +0.64$ MIXED vs. ES). The *rural–urban* axis shows that a character in ES contexts is positioned more toward *rural* as compared to the same character in EN ($d_z = -0.63$) or MIXED ($d_z = -0.85$) contexts. However, when a character appears in MIXED context relative to EN, they are positioned as slightly urban ($d_z = +0.11$).

Spanish Indexes Heritage and Rurality. The largest within-character effects in RQ2 emerge on axes indexing cultural orientation. On *tradition orientation*, characters shift dramatically toward the traditional pole in Spanish-only contexts ($d_z = -0.54$ for EN→ES; $d_z = -1.00$ for MIXED→ES). Similarly, on *rural-urban*, characters shift toward rural ($d_z = -0.63$ for EN vs. ES; $d_z = -0.85$ for MIXED vs. ES). This suggests that Spanish is used in narrative moments that are heritage-oriented/temporal-past oriented, and geographically rural-coded, rather than just functioning as a neutral alternative language across all contexts.

Affect. Three affect axes show significant shifts, revealing a complex pattern. On *fear* axis, characters in ES contexts are positioned as more *afraid* than in MIXED ($d_z = +0.87$) or EN ($d_z = +0.65$) contexts but, a character appearing in MIXED context is positioned as slightly braver ($d_z = -0.28$) than the same character appearing in EN context. On *likability*, characters become *unlikable* in ES contexts ($d_z = -0.67$ MIXED vs. ES; $d_z = -0.56$ EN vs. ES). On *character agency*, characters shift toward *helped* (needing assistance) in ES contexts ($d_z = -0.57$ MIXED vs. ES; $d_z = -0.51$ EN vs. ES) but they are positioned slightly more as helpers ($d_z = +0.20$) in MIXED context relative to EN contexts. This affect pattern presents a striking contrast with the socioeconomic findings: the same characters who are framed as more informed and of higher status in ES contexts are simultaneously framed as more afraid, less likable, and more in need of help. This within-character contradiction suggests that Spanish-only contexts may activate complex, multidimensional stereotype content rather than uniformly positive or negative framings.

Divergent Shift in Directions Across Comparisons. On 7 of the 11 axes significant in all three comparisons, the EN vs MIXED and EN vs ES comparisons show shifts in opposite directions. For example, on *literacy skill*, characters are positioned closer to illiterate in MIXED contexts relative to EN contexts ($d_z = -0.16$,$q < 0.01$), but closer to literate in ES contexts relative to EN contexts ($d_z = +0.70$, $q < 0.001$). Similar opposing patterns appear on *status*, *fear*, *likability*, *character agency*, *tradition orientation*, and *rural-urban* (see Table 2). On the remaining 4 axes (*cultivation*, *uncertainty avoidance*, *national origin*, *work type*), MIXED falls between EN and ES.

4 Discussion

4.1 The Contradictory Portrait

Our analysis shows that Spanish-associated characters occupy contradictory semantic positioning in bilingual young adult Spanish-English literature books through two complementary lenses: *casting* (i.e., how are characters placed in different linguistic contexts) and *framing* (i.e., how the same character's positioning shifts across linguistic contexts). Findings from RQ1, examining *casting* reveal that Spanish-associated names are simultaneously framed as having both deficit-oriented and agentic traits: more *stuck* than mobile, more *illiterate* than literate, more *cold* than warm, and more *rigid* than flexible; yet also as *helpers* rather than helped, and as *rooted* rather than displaced. Spanish-associated characters are thus not uniformly stereotyped but show a mixed pattern in semantic space, consistent with research documenting that group stereotypes are often ambivalent rather than uniformly negative [21].

A possible explanation for this contradiction could be that writers of these bilingual YA literature attempted to actively author against deficit stereotypes at the narrative level – casting Spanish-associated characters as helpers, protagonists, and rooted community members – while their word choices and collocations may still reflect broader cultural associations with low literacy, reduced mobility in social standing, and social coldness that pervade the linguistic environments in which both authors and readers are immersed. This suggests conscious authorial choices to subvert stereotypes may coexist with subtler semantic patterns that still reflect broader cultural associations.

If contradiction arose solely from casting, we would expect characters to receive consistent framing regardless of the linguistic context in which they appear. Instead, our results indicate that contradictions exist in framing as well. RQ2 reveals that in Spanish-only contexts, characters are positioned as more literate, intellectual, and of high-status, yet simultaneously more afraid and less likable. This contradictory portrait is thus not merely an artifact of which characters appear in Spanish contexts (RQ1), but also emerges in how the framing for individual characters changes when they appear in different language contexts. This further suggests that complexity is not reducible to authorial character selection alone. One possible explanation is that these trends reflect narrative conventions around Spanish in bilingual fiction, where the language marks both identity-affirming and vulnerability-laden scenes. For instance, a character might appear more *literate* in English contexts simply because English is used when she is at school, whereas Spanish appears in domestic scenes where education is not topical. When such domain-based language use occurs systematically across 70 books, it constitutes a corpus-level pattern with consequences for both NLP training and reader perceptions.

These findings speak to both NLP bias auditing and literary analysis. For literary scholars, these corpus-level patterns provide empirical grounding for what close reading might detect in individual texts: characters who are simultaneously elevated and diminished. For NLP practitioners, this underscores that representational unfairness is not always unidirectional: downstream models may encode such contradictory associations simultaneously, complicating simple notions of 'biased' and 'unbiased' representations. Future work could disaggregate by author background or individual novels to examine whether these contradictory patterns are consistent across the corpus or driven by specific texts in the books and/or author intent. Notably, this pattern may reflect authorial choices - that is, where authors choose to use Spanish in heritage-oriented or rural scenes, rather than properties of Spanish itself. Future work could also test whether this effect holds when we specifically examine dialogues vs narration within individual books, or even if a few novels heavily drive the pattern.

4.2 Code-Mixed Contexts Behave as a Distinct Language Register

Across our within-character analysis (RQ2), code-mixed (MIXED) contexts exhibit patterns inconsistent with serving as an intermediate point between English-only (EN) and Spanish-only (ES) language contexts. We refer to this phenomena as the *distinct register* effect: MIXED shifts are frequently divergent with respect to EN and

ES, rather than aligning consistently with either of their behaviors. **First**, we observe this when, out of the 11 axes significant across *all three* pairwise comparisons, 7 show *directional disagreement*: the EN→MIXED and EN→ES effects have opposite signs. Crucially, these are not peripheral dimensions—they include likability and status (fundamental social evaluation), literacy and fear (narrative stakes), character agency, and tradition and rural-urban (cultural positioning) dimensions. **Second,** all three pairwise comparisons are internally consistent. For each of the axes, the mixed→ES effect logically corroborates the other two comparisons: if EN→MIXED is positive and EN→ES is negative, then MIXED→ES must be negative, and it is. Across all 7 axes, EN falls between MIXED and ES (see Table 3). The two non-English registers pull character framing in opposite directions from the English baseline.

These findings challenge two common modeling defaults in multilingual NLP: First, many pipelines assume a single language label per instance, treating code-mixing as noise to filter or collapse. In multilingual settings, when code-mixing is modeled, it is often treated as an intermediate between concrete monolingual languages. However, our results are inconsistent with both assumptions: *mixed language contexts appear as a distinct register*, and across the 7 reversal axes, the English baseline is consistently intermediate, while mixed and Spanish pull in opposite directions. For accountable and valid measurement, collapsing English, Mixed, and Spanish into a single representation or forcing mixed into a monolingual bucket can change both the direction and interpretation of estimated associations, producing conclusions that may not hold within the contexts where entities actually appear. Second, this implies a concrete auditing guideline: *Mixed needs to be treated as a distinct register*, like other monolingual varieties, and needs to be audited separately. In short, disaggregating by register reveals associations and insights that are erased and filtered by interpolation assumptions which is imperative for more transparent and accountable evaluation.

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

Table 3. Ordering of the three lects on 11 robust axes. Each arrow shows the position order from pole A (left) to pole B (right). *Example:* Literacy Skill (MX→EN→ES) means the same character appears most *illiterate* in MIXED, intermediate in EN, and most *literate* in ES. Blue/orange rows (opposite directions, 7 axes) = MIXED and ES diverge from EN in opposite directions; gray rows (same direction, 4 axes) = MIXED falls between EN and ES. On 7 of 11 axes, EN is intermediate—inconsistent with MIXED as interpolation. MX = MIXED.

For qualitative analysis of bilingual literature, this result aligns with sociolinguistic accounts of code-switching as a deliberate stylistic resource i.e., authors deploy code-switching for specific narrative effects that differ from any monolingual modes [11, 23], and it cannot be reduced to a simple mixture of two languages [1]. The implications extend beyond literature too. Code-switching/mixing is socially patterned and central to the language practices of many multilingual communities [19]. For communities like Guaraní-Spanish mixing (Jopara) in Paraguay, Hinglish in India, Spanglish in the U.S., Code-Mixing is not an exception but a norm. Treating code-mixed text as an interpolation to collapse, or a noise to discard, risks erasing the linguistic reality of these populations from computational analysis [52]. If bias audits do not treat Code-Mixing as its own linguistic/lectal space, they risk systematic unfairness toward the communities whose language practices include it.

4.3 Broader Impact and Implications for Computational Systems

Literary corpora contribute to language model training through datasets like BookCorpus and other web-scraped collections. The patterns we find, such as contradictory casting and context-dependent framing, may resemble

patterns in similar upstream training data. Should such patterns exist in training corpora, they could potentially propagate into model representations, with possible implications for downstream tasks for bilingual users and Latine communities. We did not test this propagation directly in the paper, but our results suggest the possibility for it, warranting immediate attention. Future work could examine whether multilingual models exhibit similar 3-lect structures in their representations and whether model generated text reproduces the casting and framing effect we observe here. More broadly, our finding that code-mixed contexts exhibit characteristics of a distinct language space suggests value in treating code-mixed with the same analytic rigor as monolingual varieties in multilingual NLP. Therefore, we advocate for recognizing code-mixing as a norm rather than an exception, and treating it as a distinct register that preserves the sociolinguistic reality of multilingual communities.

Lastly, we note that our analysis documents corpus-level patterns across 70 novels, not the choices of individual authors; many of these authors are Latine writers, creating representation that was historically absent from Young Adult literature. The contradictory patterns we observe likely reflect broader cultural associations encoded in language rather than conscious authorial intent, and should not be read as criticism of authors writing for their communities. More generally, quantitative bias metrics risk oversimplification, underscoring the value of pairing computational analysis with qualitative interpretation.

4.4 Limitations and Future Direction

Our corpus of 70 YA novels represents a specific slice of bilingual literature, genre-specific to young adult fiction. Our scope may not generalize to other Spanish-English contexts, therefore, future work could extend this analysis to other publishing periods, or literary genres in varied bilingual contexts. Our three-lect classification involves heuristics with edge cases, and the Spanish-only category (0.31% of sentences) limits statistical power for comparisons involving ES contexts ($n = 39$ for EN vs. ES, $n = 37$ for MIXED vs. ES); larger or more balanced corpora could enable finer-grained register distinctions and more robust inferences. Most importantly, our observational design documents associations in embedding space but cannot establish causation. Future studies could pair computational audits with qualitative analysis, or model probing to examine authorial intent, authorial background (e.g., through Own Voices or insider-outsider framings), reader perception, or downstream propagation of our findings to models. Finally, methodological choices including model selection, embedding aggregation, and axis construction reflect rationalized design decisions. While these axes are theory-informed, future work could reuse and test these axes across datasets, languages, and modeling choices, providing additional evidence of their reliability and robustness.

5 Conclusion

Bilingual YA literature shapes how Latine youth perceive language, culture, and identity; understanding the semantic associations these texts encode becomes important for both literary analysis and NLP systems trained on such data. Using contextual embeddings projected onto semantic axes, we examined how 619 character names are positioned across 70 Spanish-English YA novels. Our findings show that Spanish-associated characters occupy a contradictory semantic space: simultaneously positioned toward deficit framings and agentic framings. This contradictory pattern extends to within-character framing: the same character's positioning shifts systematically across linguistic contexts, with Spanish-only contexts elevating competence while also heightening vulnerability. Crucially, code-mixed contexts do not fall between English and Spanish but constitute a distinct linguistic space. These results highlight the need to treat code-mixed as a legitimate linguistic register in multilingual NLP, not as interpolation to collapse. To the extent that literary corpora contribute to language model training, ensuring that bias audits account for linguistic context and language mixing will be important for fair and accurate representation of bilingual communities.

Acknowledgments

Support for this research was provided by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation. We acknowledge the contributions of Lauren Ciha, Nathan Han, Madison Lin, and Swarit Pakala, whose earlier exploratory work laid the groundwork for this study. We also thank Alina Guha, Kaycie Barron, Priyal Jain, and Shanthi Kuppa for their contributions to the inter-rater reliability validation.

Ethical Considerations

This study analyzes 70 bilingual YA novels. Our use is non-consumptive and transformative: we extract aggregate statistical patterns over the corpus and do not reproduce, or redistribute the source texts. The analysis does not displace the market for the original works, and our outputs cannot serve as a substitute for reading them.

As part of this fair-use posture, we do not name the individual authors, titles, or publishers in the corpus, and we have withheld the book list from this paper. Researchers with a scholarly interest in the corpus composition may contact the authors; access may be conditioned on appropriate institutional review.

The corpus was digitized via Google Cloud Vision: under its data usage terms, submitted content is not retained, used for model training, or shared with third parties. The digitized text was deleted from our pipeline immediately after embeddings were generated. We release our analysis code but not the corpus or any derivatives from which the source text could be reconstructed.

We report all findings at the corpus level. Our intent is to characterize broad patterns in bilingual literary representation and to offer a methodology others can extend, not to evaluate specific works or authors. We chose semantic axes and pole words to describe constructs neutrally and grounded each axis in established theoretical literature.

This study did not involve human subjects and was thus exempt from IRB evaluation.

Generative AI Usage Statement

The authors used generative AI tools during the preparation of this manuscript. Specifically, Claude (Anthropic) was used for: (1) improving the English language, grammar, and clarity of the text, and (2) debugging code in the analysis pipeline. The research design, methodology, analysis, interpretation of results, and scholarly contributions are entirely the work of the authors, who take full responsibility for the content of this paper.

References

[1] Peter Auer. 1998. *Code-Switching in Conversation: Language, Interaction and Identity.* Routledge, London.

[2] Jack Bandy and Nicholas Vincent. 2021. Addressing "Documentation Debt" in Machine Learning Research: A Retrospective Datasheet for BookCorpus. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks.* https://datasets-benchmarks-proceedings.neurips.cc/paper/2021/file/54229abfcfa5649e7003b83dd4755294-Paper-round1.pdf

[3] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency.* 610–623. doi:10.1145/3442188.3445922

[4] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 1 (1995), 289–300. http://www.jstor.org/stable/2346101

[5] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit.* O'Reilly Media, Inc. https://www.nltk.org/book/

[6] Rudine Sims Bishop. 1990. Mirrors, windows, and sliding glass doors. *Perspectives: Choosing and Using Books for the Classroom* 6, 3 (1990), ix–xi.

[7] Peter M. Blau and Otis Dudley Duncan. 1967. The American Occupational Structure. https://api.semanticscholar.org/CorpusID:144013887

[8] 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 (2016).

[9] Pierre Bourdieu. 2018. Distinction a social critique of the judgement of taste. In *Inequality.* Routledge, 287–318.

[10] Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. 2016. Semantics derived automatically from language corpora necessarily contain human biases. *CoRR* abs/1608.07187 (2016). arXiv:1608.07187 http://arxiv.org/abs/1608.07187

[11] Laura Callahan. 2004. *Spanish/English Codeswitching in a Written Corpus.* John Benjamins Publishing, Amsterdam.

[12] Leo R Chavez. 2025. *The Latino threat* (3 ed.). Stanford University Press, Palo Alto, CA.

[13] Rochelle Choenni and Ekaterina Shutova. 2020. Cross-neutralising: Probing for joint encoding of linguistic information in multilingual models. *arXiv preprint arXiv:2010.12825* (2020).

[14] James Clifford and Edward W Said. 1980. Orientalism. *Hist. Theory* 19, 2 (Feb. 1980), 204.

[15] Jacob Cohen. 1988. *Statistical power analysis for the behavioral sciences* (2 ed.). Routledge Member of the Taylor and Francis Group, New York, NY.

[16] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised Cross-lingual Representation Learning at Scale. *CoRR* abs/1911.02116 (2019). arXiv:1911.02116 http://arxiv.org/abs/1911.02116

[17] Rachel E Davis, Sunghee Lee, Timothy P Johnson, and Steven K Rothschild. 2019. Measuring the elusive construct of personalismo among Mexican American, Puerto Rican, and Cuban American adults. *Hisp. J. Behav. Sci.* 41, 1 (Feb. 2019), 103–121.

[18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR* abs/1810.04805 (2018). arXiv:1810.04805 http://arxiv.org/abs/1810.04805

[19] A. Seza Doğruöz, Sunayana Sitaram, Barbara E. Bullock, and Almeida Jacqueline Toribio. 2021. A Survey of Code-switching: Linguistic and Social Perspectives for Language Technologies. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, Online, 1654–1666. doi:10.18653/v1/2021.acl-long.131

[20] A. Seza Doğruöz, Sunayana Sitaram, Barbara E. Bullock, and Almeida Jacqueline Toribio. 2021. A survey of code-switching: Linguistic and social perspectives for language technologies. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*. 1654–1666.

[21] Susan T. Fiske, Amy J. C. Cuddy, Peter Glick, and Jun Xu. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology* 82, 6 (2002), 878–902. doi:10.1037/0022-3514.82.6.878

[22] Wolf Garbe. 2012. *SymSpell.* https://github.com/wolfgarbe/SymSpell

[23] Penelope Gardner-Chloros and Daniel Weston. 2015. Code-switching and multilingualism in literature. *Language and Literature* 24, 3 (2015), 182–193. doi:10.1177/0963947015585065

[24] Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences* 115, 16 (2018), E3635–E3644. arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1720347115 doi:10.1073/pnas.1720347115

[25] Nancy L. Hadaway, Terrell A. Young, and Barbara A. Ward. 2012. A Critical Analysis of Language Identity Issues in Young Adult Literature. *The ALAN Review* 39, 3 (2012). doi:10.21061/alan.v39i3.a.5

[26] Larry V. Hedges. 1981. Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *Journal of Educational Statistics* 6, 2 (1981), 107–128. http://www.jstor.org/stable/1164588

[27] Geert Hofstede. 1984. Cultural dimensions in management and planning. 1, 2 (1984), 81–99. doi:10.1007/BF01733682

[28] Geert Hofstede, Gert Jan Hofstede, and Michael Minkov. 2010. *Cultures and organizations: Software of the mind, third edition* (3 ed.). McGraw-Hill Professional, New York, NY.

[29] Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python. (2020). doi:10.5281/zenodo.1212303

[30] Dirk Hovy and Shrimai Prabhumoye. 2021. Five sources of bias in natural language processing. *Language and Linguistics Compass* 15, 8 (2021), e12432. doi:10.1111/lnc3.12432

[31] Austin C. Kozlowski, Matt Taddy, and James A. Evans. 2018. The Geometry of Culture: Analyzing Meaning through Word Embeddings. *CoRR* abs/1803.09288 (2018). arXiv:1803.09288 http://arxiv.org/abs/1803.09288

[32] 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 (Sept. 2019), 905–949. doi:10.1177/0003122419877135

[33] Keita Kurita, Nidhi Vyas, Ayush Pareek, Alan W Black, and Yulia Tsvetkov. 2019. Measuring Bias in Contextualized Word Representations. arXiv:1906.07337 [cs.CL] https://arxiv.org/abs/1906.07337

[34] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174. doi:10.2307/2529310

[35] Cammie Jo Lawton. 2024. *Instigators and Partners: Young Adult Literature's Role in Readers' Identity Formation.* Doctoral dissertation. University of Tennessee. https://trace.tennessee.edu/utk_graddiss/10135/

[36] Li Lucy, Divya Tadimeti, and David Bamman. 2022. Discovering Differences in the Representation of People using Contextualized Semantic Axes. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, Yoav Goldberg, Zornitsa Kozareva, and

Yue Zhang (Eds.). Association for Computational Linguistics, Abu Dhabi, United Arab Emirates, 3477–3494. doi:10.18653/v1/2022.emnlp-main.228
[37] Peter M. stahl. [n. d.]. *lingua*. https://github.com/pemistahl/lingua GitHub repository. Accessed 2025-12-21.
[38] Gerardo Marín and Barbara VanOss Marín. 1991. Research with Hispanic Populations. *Applied Social Research Methods Series* 23 (1991).
[39] Pieter Muysken. 2000. *Bilingual speech*. Cambridge University Press, Cambridge, England.
[40] Carol Myers-Scotton. 1993. *Social Motivations for Codeswitching: Evidence from Africa*. Clarendon Press, Oxford.
[41] Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How Multilingual is Multilingual BERT?. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Anna Korhonen, David Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, Florence, Italy, 4996–5001. doi:10.18653/v1/P19-1493
[42] Fabio Sabogal, Gerardo Marín, Regina Otero-Sabogal, Barbara Vanoss Marín, and Eliseo J Perez-Stable. 1987. Hispanic familism and acculturation: What changes and what doesn't? *Hisp. J. Behav. Sci.* 9, 4 (Dec. 1987), 397–412.
[43] Nipun Sadvilkar and Mark Neumann. 2020. PySBD: Pragmatic Sentence Boundary Disambiguation. In *Proceedings of Second Workshop for NLP Open Source Software (NLP-OSS)*. Association for Computational Linguistics, Online, 110–114. https://www.aclweb.org/anthology/2020.nlposs-1.15
[44] Rachel S. Savitz, Leila Roberts, and David Stockwell. 2022. The Impact of Analyzing Young Adult Literature for Racial Identity / Social Justice Orientation with Interdisciplinary Students. *Journal of College Reading and Learning* 52, 4 (2022), 264–289. doi:10.1080/10790195.2022.2044933
[45] Shlomo S Sawilowsky. 2009. New effect size rules of thumb. *J. Mod. Appl. Stat. Methods* 8, 2 (Nov. 2009), 597–599.
[46] Pitirim Aleksandrovich Sorokin. 1959. *Social and cultural mobility*. Glencoe, Ill : Free Press.
[47] Karolina Stanczak and Isabelle Augenstein. 2021. A survey on gender bias in natural language processing. *arXiv preprint arXiv:2112.14168* (2021).
[48] J. Valhondo. 2016. Dataset of Spanish Names and Surnames. https://github.com/jvalhondo/spanish-names-surnames Data sourced from Spain's Instituto Nacional de Estadística (INE), reference date 01/01/2015.
[49] Margaret Vaughn, Dixie Massey, Adrienne Vitullo, Jessica Masterson, and Xuejiao Li. 2022. Critical depictions of agency in Pura belpré awarded texts. *Child. Lit. Educ.* 53, 3 (Sept. 2022), 343–359.
[50] Jill West. 2010. *Gender Bias and Stereotypes in Young Adult Literature: A Content Analysis of Novels for Middle School Students*. Master's thesis. University of North Carolina at Chapel Hill. doi:10.17615/9mjt-p939
[51] Raymond Williams. 2011. *The country and the city*. Spokesman Books, Nottingham, England.
[52] Genta Indra Winata, Alham Fikri Aji, Zheng-Xin Yong, and Thamar Solorio. 2023. The decades progress on code-switching research in NLP: A systematic survey on trends and challenges. In *Findings of the Association for Computational Linguistics: ACL 2023*.
[53] Yukun Zhu, Ryan Kiros, Rich Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Aligning Books and Movies: Towards Story-Like Visual Explanations by Watching Movies and Reading Books. In *Proceedings of the IEEE International Conference on Computer Vision*. 19–27.

A Text Preprocessing Pipeline

Following OCR digitization, we conducted four steps to ensure we protect bilingual text integrity as we process the raw output:

- **Dictionary-validated de-hyphenation**: joins hyphenated line breaks only when the result is a valid word in either an English or Spanish dictionary (via SymSpell [22]).
- **Character normalization and filtering**: standardizes typographic variants (e.g., smart quotes) while removing OCR artifacts, preserving Spanish diacritics (á, é, í, ó, ú, ñ, ü) and inverted punctuation (¿, ¡) via an approved character whitelist.
- **Safe single-character filtering**: removes single-character noise while protecting Spanish connectives and vowels (*a, y, o, e, i, u*) via a dedicated whitelist.
- **Sentence segmentation and quality filtering**: using pysbd [43], with Spanish-specific punctuation preserved by the above whitelist, we filtered sentences to those containing more than 3 words.

Notably, we ensured that no accent removal, lowercasing, stopword removal, or stemming was applied at any stage.

B Three-Lect Classification Rules

Algorithm 1 presents our three-lect sentence classification procedure. To handle OCR artifacts and spelling variants, tokens are canonicalized by stripping diacritics for matching (e.g., *María* → *maria*); original text is preserved for embeddings. Table 4 defines the lexicons and the threshold parameters used.

Table 4. Lexicons and Threshold Parameters for Three-Lect Classification

Lexicon	Size	Examples
$\mathcal{S}_{\text{en}}$	68	*the, and, is*
$\mathcal{S}_{\text{es}}^{\text{strict}}$	87	*el, la, de*
$\mathcal{S}_{\text{es}}^{\text{loose}}$	16	*me, lo, le*
$\mathcal{W}_{\text{disc}}$	50	*abuela, mijo*

Parameter	Description	Value
$\phi_{\text{es}}^{\text{high}}$	Spanish conf. for ES	0.85
$\phi_{\text{en}}^{\text{low}}$	Max English for ES	0.15
$\phi_{\text{es}}^{\text{mid}}$	Spanish for MIXED	0.35
$\phi_{\text{en}}^{\text{mid}}$	English for MIXED	0.35
τ_{en}	Min EN stopwords	2
τ_{es}	Min ES stopwords	2
τ_{len}	Min tokens	6

Algorithm 1: Three-Lect Sentence Classification

Data: sentence s, Lingua confidence scores $(\phi_{\text{en}}, \phi_{\text{es}})$
Result: label $\in$ {EN, ES, MIXED}
$T \leftarrow$ canonical tokens of s (diacritics stripped);
$n_{\text{en}} \leftarrow |T \cap \mathcal{S}_{\text{en}}|$;
$n_{\text{es}} \leftarrow |T \cap \mathcal{S}_{\text{es}}^{\text{strict}}|$;
$n_{\text{es}}^{\text{any}} \leftarrow |T \cap (\mathcal{S}_{\text{es}}^{\text{strict}} \cup \mathcal{S}_{\text{es}}^{\text{loose}})|$;
$n_{\text{disc}} \leftarrow |T \cap \mathcal{W}_{\text{disc}}|$;

```
// Rule 1: Spanish-dominant
if φ_es ≥ 0.85 and φ_en ≤ 0.15 and n_es ≥ 2 and n_en ≤ 1 then
    return ES;

// Rule 2: Code-switching via stopword co-occurrence
if n_en ≥ 2 and n_es ≥ 2 then
    return MIXED;

// Rule 3: Code-switching via Lingua + Spanish evidence
if φ_es ≥ 0.35 and φ_en ≥ 0.35 and (n_es^any ≥ 1 or n_disc ≥ 1) then
    return MIXED;

// Rule 4: Code-switching via discourse marker
if n_disc ≥ 1 and n_en ≥ 2 and |T| ≥ 6 then
    return MIXED;

// Default: English-dominant
return EN;
```

Results. EN: 261,764 (97.64%); MIXED: 5,470 (2.04%); ES: 844 (0.31%). Total: 268,078 sentences.

Table 5. Example Sentences by Classification Category

Label	Example Sentence
ES	"¿Por qué no me dijiste antes? Siempre es lo mismo contigo."
ES	"Pero yo quiero ir con ella porque es mi prima."
MIXED	"My abuela always said that family comes first."
MIXED	"She whispered que lo sentía, but by then I was already out the door."
MIXED	"I told her que no, but she just laughed and said ya veremos."
EN	"I went to the store yesterday."
EN	"She looked out the window and sighed."

C Semantic Axes and Pole Words

Following the methodology of Kozlowski et al. [32], we construct semantic axes as the mean difference vector between pole word sets. Table 6 presents all 25 axes organized by category.

Table 6. Semantic axes with pole words.[†]

Axis	Pole A (–)	Pole B (+)
Kozlowski Validated (2 axes)		
Affluence	poor, poorer, poorest, poverty, impoverished, inexpensive, cheap, needy	rich, richer, richest, affluence, affluent, expensive, luxury, opulent
Gender	man, men, he, him, his, boy, boys, male, masculine	woman, women, she, her, hers, girl, girls, female, feminine
Hofstede Cultural Dimensions (6 axes)		
Individualism	communal, group, community, collective, together, we, our	individual, independent, self, autonomous, alone, personal, private
Power Distance	equal, democratic, flat, informal, peer	authority, obedient, deferential, formal, superior
Stereotypical Gender Traits	nurturing, caring, cooperative, modest, tender, gentle	competitive, assertive, ambitious, tough, aggressive, dominant
Uncertainty Avoidance	flexible, tolerant, relaxed, accepting, open, adaptable	rigid, strict, anxious, structured, orderly, precise
Tradition Orientation	traditional, past, heritage, conventional, conservative	pragmatic, future, modern, progressive, innovative
Indulgence	restrained, disciplined, controlled, strict, serious, duty	indulgent, fun, playful, free, enjoying, pleasure
Socioeconomic Status (5 axes)		
Literacy Skill	illiterate, ignorant, unread	literate, learned, scholarly

Continued on next page

Table 6 continued from previous page

Axis	Pole A (–)	Pole B (+)
Formal Education	uneducated, untrained, unschooled, uncredentialed, unqualified, dropout, amateur	educated, trained, schooled, credentialed, qualified, graduate, certified
Work Type	manual, physical, labor, menial	professional, intellectual, mental, skilled
Status	lowly, humble, inferior, subordinate, minor	high, superior, prestigious, esteemed, distinguished
Social Mobility	stuck, trapped, static, stagnant, fixed	mobile, rising, advancing, climbing, progressing
Belonging & Culture (5 axes)		
Immigration Status	illegal, unauthorized, undocumented, illegitimate	legal, authorized, documented, legitimate
National Origin	foreign, alien, immigrant, outsider	native, patriotic, domestic, national
Place Belonging	displaced, homeless, rootless, wandering, lost	rooted, home, belonging, settled, established
Cultivation	uncivilized, barbaric, savage, primitive, wild	civilized, cultured, refined, sophisticated, polished
Rural–Urban	rural, country, farm, village, agricultural	urban, city, metropolitan, downtown, suburban
Interpersonal Relational Values (3 axes)		
Warmth	cold, distant, aloof, unfriendly, hostile	warm, friendly, welcoming, kind, loving
Personalismo	impersonal, formal, bureaucratic, distant, cold	personal, warm, intimate, friendly, close
Familism	individual, independent, autonomous, solo, alone	family, familial, kinship, clan, relatives
Affect (4 axes)		
Character Agency	helped, saved, rescued, assisted, supported	helper, savior, rescuer, protector, supporter
Praised–Criticized	criticized, blamed, attacked, condemned, faulted	praised, celebrated, lauded, commended, honored
Fear	brave, courageous, fearless, bold, confident	afraid, scared, fearful, frightened, terrified
Likability	unlikable, unpleasant, disagreeable, annoying, irritating	likable, pleasant, agreeable, charming, delightful

† Affluence and Gender axes use pole words from Kozlowski et al. [31].

D Inter-Rater Reliability Results for Pole Word Selection

To establish the reliability of the word lists used to construct each semantic axis, we conducted an inter-rater reliability (IRR) study following the general approach of Kozlowski et al. (2019). Two raters independently

evaluated a subset of 8 of the 23 axes included in the full analysis (35% of all axes; excluding the 2 axes validated by Kozlowski et al. [32]), selected at random. The axes evaluated were: Familism, Immigration Status, Literacy Skill, Personalismo, Place Belonging, Uncertainty Avoidance, Warmth, and Work Type. For each axis, raters were presented with the axis's own words alongside an equal number of foil words drawn randomly from a different axes among the remaining 15 (23 - 8 = 15). This design tested not only whether raters agreed on pole assignment but also whether they could correctly distinguish axis-relevant words from semantically plausible but irrelevant foils. Raters were not told which words were foils but only provided the definition of each axis they were judging. For each word, raters completed two sequential judgments: first, whether the word was related to/represented the given axis (Yes/No), and if they answered "Yes", which pole of the axis the word belonged to (Positive/Negative). Reliability was computed as Cohen's Kappa (κ) between the two raters separately for each judgment task. For Task 1 (Yes/No validity), overall inter-rater agreement was substantial ($\kappa = 0.672$, $agreement = 83.8\%$). Performance was strong across most axes, with Work Type reaching almost perfect agreement ($\kappa = 1.000$) and Immigration Status approaching it ($\kappa = 0.889$). The remaining axes fell in the moderate-to-substantial range ($\kappa = 0.500$–0.727), indicating generally consistent ability to distinguish valid from foil words. Rater 2 showed slightly higher overall accuracy against the answer key (90.6%) compared to Rater 1 (86.9%). For Task 2 (pole assignment), reliability was markedly higher. Scored only on valid words where both raters agreed the word belonged to the axis, overall kappa was almost perfect ($\kappa = 0.944$, $agreement = 97.1\%$). Six of the eight axes achieved $\kappa = 1.000$, indicating perfect pole agreement. Work Type showed substantial agreement ($\kappa = 0.600$).

Taken together, these results support the reliability of the semantic axis word lists used in this study. The near-perfect pole assignment agreement across axes indicates that the positive and negative ends of each dimension are conceptually clear and consistently interpreted. The moderate-to-substantial Task 1 agreement reflects expected difficulty in distinguishing semantically adjacent foil words from valid axis words (a more demanding task) and is consistent with acceptable IRR standards for this type of judgment [34].

E Full RQ1 Results

Table 7 presents Welch's t-test results comparing Spanish-associated ($\rho \geq 0.15$, $n = 47$) and English-associated ($\rho \leq 0.01$, $n = 360$) characters across all 25 semantic axes. Effect sizes are reported as Hedges' g with 95% confidence intervals. All q-values reflect Benjamini-Hochberg FDR correction.

Table 7. RQ1: Full results for all 25 semantic axes. Positive g indicates ES-associated characters positioned toward pole B; negative g toward pole A.

Axis	Dimension	g	95% CI	q	Sig
Significant after FDR correction ($q \leq 0.05$)					
Social Mobility	stuck—mobile	−0.52	[−0.82, −0.21]	0.003	**
Cultivation	uncivilized—civilized	−0.47	[−0.78, −0.17]	0.012	*
Literacy Skill	illiterate—literate	−0.47	[−0.78, −0.17]	0.019	*
Personalismo	impersonal—personal	−0.46	[−0.76, −0.15]	0.033	*
Uncertainty Avoidance	flexible—rigid	+0.45	[+0.15, +0.76]	0.019	*
Place Belonging	displaced—rooted	+0.44	[+0.14, +0.75]	0.033	*
Warmth	cold—warm	−0.43	[−0.73, −0.12]	0.033	*
Individualism	communal—individual	+0.40	[+0.09, +0.70]	0.032	*

Continued on next page

Table 7 continued

Axis	Dimension	g	95% CI	q	Sig
Character Agency	helped—helper	+0.39	[+0.08, +0.69]	0.040	*
Not significant after FDR correction ($q \geq 0.05$)					
Immigration Status	illegal—legal	−0.36	[−0.67, −0.06]	0.090	
Familism	individual—family	−0.34	[−0.65, −0.04]	0.055	
Formal Education	uneducated—educated	−0.33	[−0.64, −0.03]	0.093	
Likability	unlikable—likable	+0.31	[+0.00, +0.61]	0.091	
Kozlowski Affluence	poor—rich	+0.30	[−0.00, +0.61]	0.093	
Work Type	manual—intellectual	−0.30	[−0.61, +0.00]	0.072	
Fear	brave—afraid	−0.25	[−0.56, +0.05]	0.109	
Power Distance	equal—authority	+0.25	[−0.06, +0.55]	0.204	
Praised–Criticized	criticized—praised	+0.23	[−0.07, +0.54]	0.177	
National Origin	foreign—native	−0.17	[−0.48, +0.13]	0.356	
Tradition Orientation	traditional—pragmatic	+0.17	[−0.14, +0.47]	0.356	
Indulgence	restrained—indulgent	−0.16	[−0.46, +0.15]	0.356	
Status	lowly—high	−0.13	[−0.43, +0.18]	0.422	
Stereotypical Gender Traits	nurturing—competitive	+0.11	[−0.19, +0.42]	0.434	
Rural–Urban	rural—urban	+0.02	[−0.28, +0.32]	0.898	
Kozlowski Gender	man—woman	−0.02	[−0.32, +0.29]	0.898	

Note: ** $q < 0.01$, * $q \leq 0.05$. ES: $n = 47$; EN: $n = 360$. Sorted by $|g|$ within significance groups.

F Full RQ2 Results

Table 8 presents paired t-test results for within-character semantic shifts across all three pairwise comparisons. Effect sizes are reported as Cohen's d_z. All q-values reflect Benjamini-Hochberg FDR correction applied separately within each comparison.

Table 8. RQ2: Within-character semantic shifts across linguistic contexts for all 25 axes. Sorted by number of significant comparisons, then by maximum $|d_z|$.

		EN→MIXED		EN→ES		MIXED→ES	
Axis	**Dimension**	d_z	q	d_z	q	d_z	q
Tradition Orientation	traditional—pragmatic	+0.26	< .001	−0.54	.004	−1.00	< .001
Literacy Skill	illiterate—literate	−0.16	.003	+0.70	< .001	+0.96	< .001
Fear	brave—afraid	−0.28	< .001	+0.65	.001	+0.87	< .001
Rural–Urban	rural—urban	+0.11	.034	−0.63	.001	−0.85	< .001
Work Type	manual—intellectual	+0.24	< .001	+0.79	< .001	+0.58	.003

Continued on next page

Table 8 continued

		EN→MIXED		EN→ES		MIXED→ES	
Axis	**Dimension**	d_z	q	d_z	q	d_z	q
Cultivation	uncivilized—civilized	+0.26	< .001	+0.74	< .001	+0.64	.001
Uncertainty Avoidance	flexible—rigid	−0.18	.001	−0.70	< .001	−0.67	< .001
Likability	unlikable—likable	+0.15	.006	−0.56	.004	−0.67	< .001
Status	lowly—high	−0.17	.002	+0.65	.001	+0.63	.001
National Origin	foreign—native	+0.17	.002	+0.58	.002	+0.49	.010
Character Agency	helped—helper	+0.20	< .001	−0.51	.007	−0.57	.003
Individualism	communal—individual	−0.01	.826	−0.45	.014	−0.78	< .001
Social Mobility	stuck—mobile	+0.67	< .001	+0.48	.009	+0.05	.800
Familism	individual—family	−0.44	< .001	+0.24	.185	+0.60	.002
Warmth	cold—warm	+0.60	< .001	+0.40	.028	−0.05	.800
Personalismo	impersonal—personal	+0.50	< .001	+0.48	.009	+0.15	.458
Stereotypical Gender Traits	nurturing—competitive	−0.23	< .001	+0.22	.229	+0.44	.019
Immigration Status	illegal—legal	−0.40	< .001	+0.20	.250	+0.42	.024
Kozlowski Gender	man—woman	−0.22	< .001	+0.10	.578	+0.37	.049
Praised–Criticized	criticized—praised	−0.56	< .001	−0.19	.279	+0.24	.223
Formal Education	uneducated—educated	+0.47	< .001	+0.34	.064	+0.09	.684
Power Distance	equal—authority	−0.27	< .001	−0.04	.861	+0.07	.748
Indulgence	restrained—indulgent	+0.12	.021	+0.26	.174	+0.22	.251
Place Belonging	displaced—rooted	+0.14	.008	−0.24	.185	−0.25	.196
Kozlowski Affluence	poor—rich	+0.07	.218	−0.01	.966	+0.01	.973