Written vowel variation in Imbabura Media Lengua

Jesse Stewart
University of Saskatchewan

This study investigates written vowel variation and stability in Imbabura Media Lengua, a contact language formed from Spanish lexicon and Kichwa morphosyntax. Through analyzing transcribed conversations by two native speakers, this research identifies patterns in orthographic choices, specifically focusing on written variability between mid and high vowels. Generalized linear mixed-effects models show that vowel choice is largely stable and source-aligned, though variation arises predictably in two domains, (1) root-final vowels preceding suffixes and (2) in lexical "reborrowings" — words originally borrowed into one source language from the other and subsequently borrowed again into Media Lengua. The minimal orthographic variation aligns closely with perceptual research, suggesting that how a vowel is perceived plays a greater role than how it is produced in determining spelling preferences. These results emphasize the roles of etymological origin and auditory perception in shaping Media Lengua's nascent orthography.

Keywords: Media Lengua, orthographic variation, language contact, nascent orthography, vowel saliency

1. Introduction

Media Lengua (glottolog: MEDI1245) is a Quechuan language (glottolog: QUEC1387), often described as a mixed language with Spanish origin lexicon and Kichwa (Quichua)[1] origin morphosyntax (Gómez Rendón 2005; Lipski 2017; Muysken 1997; Stewart accepted). It is spoken by an estimated 3,000 people in several rural communities located sparsely throughout the Ecuadorian Andes

1. Kichwa (also spelled Quichua) is the generic name given to the northern Quechuan languages primarily spoken in Ecuador. The variety of Kichwa discussed here is Imbabura Kichwa (glottolog: IMBA1240) unless otherwise mentioned.

https://doi.org/10.1075/wll.00096.ste
Written Language & Literacy 28:2 (2025), pp. 171–226. ISSN 1387-6732 | E-ISSN 1570-6001

(Stewart et al. 2023). The language exhibits an unusually high degree of lexical borrowing from Spanish (~90%), primarily transferred through relexification. Through this process, the phonological form of nearly every Kichwa word is replaced by its Spanish counterpart (Muysken 1981). Example (1) provides a sample of Media Lengua as it is spoken in the province of Imbabura. It contains Imbabura Kichwa (IK), Unified Kichwa[2] (UK), and Spanish (Sp) translations for comparison with bolded elements of Spanish origin.

(1) ***Aurita**ca **uno**gullata **tene**ne **vacona**guta **vende**rcani **ya toro**tapish.*
aurita=ka **unǫ**-gu-ʒata **tẹnẹ**-ni **bakuna**-gu-ta **bẹndẹ**-ɾka-ni **ja**
now=TOP one-DIM-TOT have-1 calf-DIM-ACC vender-PST-1 so
tǫɾǫ-ta=piʃ
bull-ACC=CONJ
IK *Cunanga shucgullata charini **vaquilla**gutami jaturcani ña **toro**dapish.*
kunan=ga ʃuk-gu-ʒata ʧaɾi-ni **bakiʒa**-gu-ta=mi xatu-ɾka-ni ña
now=TOP one-DIM-TOT have-1 heifer-DIM-ACC sell-PST-1 so
tuɾu-da=piʃ
bull-ACC=CONJ
UK *Kunanka shukkullatak charini warmi wakrakutami katurkani ña kari warkatapish.*
kunan=ka ʃuk-ku-ʒata ʧaɾi-ni waɾmi wakɾa-ku-ta=mi katu-ɾka-ni ña
now=TOP one-DIM-TOT have-1 female cow-DIM-ACC=VAL sell-PST-1 so
kaɾi wakɾa-ta=piʃ
male cow-ACC=CONJ
Sp *Ahora por lo menos solo tengo una, ya vendí la vaconita y el toro también.*
Eng 'At least I only have one now; I sold the calf and the bull as well.'
(Stewart & Prado Ayala 2025 C.S11.MI)

One area of Media Lengua research that has received considerable attention involves its vowel system, which combines Spanish and Kichwa vowels in complex ways (See Table 1 & § 2.4). This paper examines variation in the written use of Media Lengua high and mid vowels along with vowel sequences in conversational data from the Media Lengua Collection housed in The Archive of the Indigenous Languages of Latin America (AILLA), deposited by Stewart and Prado Ayala (2025).

Analyzing the transcriptions provides insight into orthographic practices, structural patterns, the degree of variability and stability, influences from both

2. Throughout this paper, "Unified Kichwa" refers to a standardized Kichwa system with reforms to orthography, lexical usage, and structure. Its development began in the late 20th century for use in education and literacy planning across Ecuador. See Section 2.2 for more details.

Table 1. Media Lengua vowels

Spanish origin	Kichwa origin	Vowel sequences
⟨i⟩ /i/	⟨i⟩ /i/	⟨ia⟩ /ia/, ⟨ie⟩ /ie/, ⟨io⟩ /io/
⟨e⟩ /e/	–	⟨ea⟩ /ea/, ⟨eo⟩ /eo/
⟨u⟩ /u/	⟨u⟩ /u/	⟨ua|wa⟩* /ua/
⟨o⟩ /o/	–	⟨oi|oy⟩ /oi/
⟨a⟩ /a/	⟨a⟩ /a/	⟨ai|ay⟩ /ai/, ⟨ae⟩ /ae/, ⟨au⟩ /au/

* The regular expression *vertical bar* represents *or* within parentheses (*(x|y|z)*) to avoid confusion with narrow transcription with allophonic variations indicated using brackets (*[]*).

source languages, and the potential impact of production and perception on written vowel choice. This study accounts for word frequency by modeling *Word* as a random effect in generalized linear mixed-effects models (see § 4.2). Example (2) illustrates the type of variation analyzed.

(2) a. *Eseguanmari com**i**nchi qu**i**gohuanta com**i**shunyari.*
'We eat with this; if not, what would we eat with?'
b. *Qu**e**tata com**e**gringue?*
'What are you going to eat?'
c. *Y**u**ca zapatos nuevotami qu**i**r**i**ni.*
'I want new shoes.'
d. *Y**o**ca mio maridopa trabajopemi mejorachun qu**e**r**e**ne.*
'I hope that my husband's job improves.'

(Stewart 2025 C.So6.Ma)

In (2a), the words *com**i**nchi* 'we eat', *com**i**shunyari* 'we might eat', *qu**i**gohuanta* 'with what' contain ⟨i⟩ vowels in the bolded positions. In contrast, (2b) shows the same root words with ⟨e⟩ vowels in the same bolded positions. Similarly, in (2c), the words *y**u**ca* 'I' and *qu**i**r**i**ni* 'I want' contain ⟨u⟩ and ⟨i⟩ vowels in the bolded positions, whereas (2d) shows the same slot words with ⟨o⟩ and ⟨e⟩ vowels in the same bolded positions, respectively.

Previous phonetic research shows that Media Lengua speakers produce Spanish origin mid vowels with substantial overlap with high vowels in acoustic space, unlike Spanish, where the categories are clearly separated (Onosson & Stewart 2021b; Stewart 2014) (see § 2.4.1). Despite this overlap, perceptual studies demonstrate that listeners reliably distinguish high and mid vowel contrasts in minimal pairs (Stewart 2018b), suggesting that perceptual distinctions are maintained even when acoustic separation is reduced (see § 2.4.2). Follow-up work further shows that vowel identification is conditioned not only by first and second formant fre-

quencies (F1xF2), but also by factors such as syllable position, lexical stress, and speaker familiarity (Onosson & Stewart 2021a).

Despite this body of work, the extent to which these patterns are reflected in written forms remains largely unexplored. This is a significant gap, particularly given the potential for developing a written system for Media Lengua that reflects speaker preferences and ongoing efforts to develop writing systems for Kichwa. As it currently stands, the orthographic practices of Media Lengua speakers are not governed by a standard system. Instead, they often reflect a combination of phonetic spellings[3] from Spanish, Kichwa phonological expectations, and individual choices. This makes written Media Lengua an ideal test case for exploring how speakers of a mixed language with a nascent orthography, and often with limited formal literacy practices, negotiate phonological representation in writing.

The present study investigates whether written vowel variation in Media Lengua corresponds to acoustic and perceptual distinctions and/or whether it follows alternative patterns shaped by standard source language orthographic conventions, etymology, or morphology. Specifically, the following questions are addressed:

1. Are written vowel forms guided by perceptual categories?
   If so, mid and high vowels in Media Lengua would mostly align with their counterparts in Standard Spanish and Unified Kichwa (see § 2.4.2).
2. Does acoustic overlap between mid and high vowels influence spelling?
   If so, we would expect written vowel choices to be variable or inconsistent (see § 2.4.1).
3. Does the substrate phonology (Kichwa) have greater influence than the lexifier (Spanish) in shaping vowel spelling?
   If so, we would expect Spanish origin mid vowels to be systematically replaced by high vowels, reflecting Kichwa's vowel system.
4. Do contextual cues (lexical origin, morphology etc.) inform orthographic practices?

At a broad level, this study shows that written vowel usage in Media Lengua is predominantly stable and aligned with source language expectations, with variation largely restricted to root-final vowels, lexical 'reborrowings' (see § 5.2), and place names. These domains reveal how morphology and etymology shape the emergence of orthographic tendencies in Media Lengua, a point taken up in detail in the Results (§ 4) and Discussion (§ 5.2).

3. I use 'phonetic spelling' to refer to the practice of writing words based on how they sound, rather than following the conventional spelling rules of a language that might deviate from how they sound.

Although the primary goal is to describe how mid and high vowel contrasts are encoded in Media Lengua writing, the patterns observed have direct implications for existing Unified Kichwa orthographic standards. In particular, the results provide an empirical basis for reconsidering how Unified Kichwa orthography accommodates local practices in Imbabura Kichwa and Media Lengua, especially in the domains of borrowings and place names.

Section 2 provides background information pertaining to orthographic practices in contact languages (§ 2.1), Unified Kichwa's orthography (§ 2.2), Media Lengua's nascent orthography (§ 2.3). Section 2.4 describes Media Lengua's phonology focusing specifically on monophthong vowel production (§ 2.4.1), monophthong vowel perception (§ 2.4.2), and the production of vowel sequences (§ 2.4.3). Each section explores how these studies might predict written vowel choice in Media Lengua.

2. Background

2.1 Orthographic practices in minority and contact languages

The relationship between spoken language and writing systems is bidirectional, where pronunciation informs orthographic conventions, and established orthographies can influence pronunciation. Writing systems are fundamentally grounded in the phonological systems of their corresponding spoken languages, yet deviations and adaptations often occur due to linguistic, social, or historical factors.

In general, the closer the match between the written and spoken forms, the easier it is to learn a language's orthography. For example, Finnish uses a highly transparent orthography where each letter corresponds closely to a specific phoneme. By contrast, English orthography reflects layers of historical sound changes, borrowings, and inconsistencies, leading to complexities like silent letters in *knight* or vowel shifts in *read* (present vs. past tense).

For traditionally oral languages, like those found in the Quechuan family, first attempts at writing systems often carry additional layers of complexity, as they can be heavily influenced by colonial languages. In the case of Quechuan languages, Roman-based orthographies have traditionally been applied, often with substantial compromise. For instance, Cuzco Quechua orthographies typically do not reflect distinctions like aspirated or ejective stops, leaving important phonological details unmarked (See the 2002 translation of "The Little Prince": Saint-Exupéry 1946; and the first Cuzco Quechua grammar from the 1500s Santo Tomás 1560). Similarly, early Navajo orthographies struggled with consistent tonal repre-

sentation (see Bahr 2004; and Young and Morgan 1972) and transliterations of the Cherokee syllabary gloss over distinctions important for communication.

When orthographies are introduced or standardized, they often reinforce particular ways of speaking. For example, in Hawaiian, the inclusion of the glottal stop (ʻ) in written forms has helped provide a more accurate representation of the language. Compare the 1862 translation of "Bluebeard" (*Umiumi Uliuli* (Perrault 1862 trans. JW) with no glottal stop grapheme with modern-day orthography (Pukui and Elbert 1986). Similarly, Ojibwe orthographic systems that mark nasalized vowels, vowel length, and fortis and lenis consonants preserve important contrasts that might otherwise result in incorrect homographs or pronunciation when read aloud (compare the Baraga system with the double vowel system and the Ojibwe syllabary). However, in other cases, for instance, Ecuadorian Kichwa, the standardized writing system has raised concerns of dialectal leveling (Gualapuro Gualapuro Forthcoming) and favoring more prestigious dialects (Grzech, Schwarz, and Ennis 2019; Limerick 2018, 2020).

Orthographic choices also reflect practical and cultural priorities. Hawaiian immersion schools, for instance, use macrons to indicate long vowels, helping students pronounce words more accurately (Slaughter and Watson-Gegeo 1988). For Ojibwe, the adoption of both syllabic and Roman scripts accommodates speakers in different regions and contexts, promoting literacy and accessibility (Hutchinson 1988; Mitchell 1988; Rhodes 1985). Over the last 50 years, activists and scholars working to reclaim Ecuadorian Kichwa have shifted from a Spanish-based system to a phonetic-based system while maintaining the Roman script (Cerda and Malaver 2005). There have been at least 5 major revisions to this system (see Ministerio de Educación, 2009).

Other complex cases involve the orthographies of contact languages; languages that emerge from the interaction and blending of two distinct linguistic systems. These cases present unique challenges and opportunities for orthographic development. For example, Haitian Creole, which evolved from French and Fongbe, saw its first attempt at an orthographic standard in 1925. This system was later revised in 1940 resulting in the McConnell-Laubach orthography and once again in 1953 by the *Office National d'Alphabétisation et d'Action Communautaire* before its official standardization in 1980 (Schieffelin and Doucet 1994). Prior to this, Haitian Creole texts, such as the 1757 poem *Lisette quitté la plaine* (de la Mahautiére 1757) were written using French-based conventions that often failed to capture the phonological distinctions of Haitian Creole.

Another example is Michif, a language evolved from Cree and French elements, which still lacks a unified spelling standard primarily due to differences in local dialects that complicate transcription. Two primary systems have been developed in an attempt to provide a standard; the Turtle Mountain spelling,

created in the 1970s (Laverdure et al. 1983) and Rita Flamand's phonetic system (Flamand 2002; also see Papen 2005). Papen (2005) states that prior to Flamand's system, and his subsequent revisions, various proposed systems relied on English, French or Cree and thus fail to fully capture Michif's phonology.

Unlike Haitian Creole or Michif, Media Lengua has a nascent orthography while its source languages (Spanish and Kichwa) both have established systems. The written form of Media Lengua offers an opportunity to study how its hybrid phonological system is represented in the minds of speakers who are writing the language for the first time. Therefore, it is helpful to first understand the Unified Kichwa writing system (§ 2.2).

2.2 Kichwa orthography

Before turning to Media Lengua orthography, it is important to briefly outline the Unified Kichwa writing system, since its development has shaped perceptions of what "correct" spelling should look like in Ecuadorian Kichwa, and most Media Lengua speakers are familiar, at least passively, with its conventions.

Unified Kichwa (also known as Standard Kichwa, Standardized Kichwa, *kichwa unificado* in Spanish, or *shukyachishka kichwa* in Kichwa) is a standardized language established through bilingual-education and literacy planning initiatives beginning in the 1970s and later led by Kichwa-speaking planners in the Intercultural Bilingual Education program (EBI) (see Ministerio de Educación, 2009 for a complete overview). Among its many reforms was the creation of a single writing system for all Ecuadorian Kichwa varieties, replacing earlier Spanish-based spellings. Standardization efforts, aimed at promoting linguistic prestige, also focused on reducing Spanish influence through the creation of neologisms and the elimination of common loanwords; a normative stance that contrasts with long-standing community practices in all varieties of Kichwa.

The unified system has undergone multiple revisions, and many speakers report that it does not consistently represent regional phonological distinctions, particularly those found in Amazonian varieties of Kichwa (Limerick, 2018) and even Andean varieties such as Imbabura Kichwa. This has led to uncertainty about correct spelling and hesitancy around written use in community contexts. Therefore, in this study, Unified Kichwa and Standard Spanish are treated only as external references rather than prescriptive targets.

As such, the standard orthographies function solely as reference baselines for evaluating written vowel variation in Media Lengua. For example, the Kichwa-origin word *wachu* 'furrow' is commonly written as *huacho* in Media Lengua (Stewart et al. 2020:1319), making it possible to assess how often the vowel appears as ⟨u⟩ versus ⟨o⟩. Similarly, the Spanish-origin word *balde* 'bucket' fre-

quently surfaces as *baldi* in Media Lengua (p. 319), allowing comparison of how often ⟨e⟩ vs. ⟨i⟩ is retained relative to the Standard Spanish form.

2.3 Media Lengua orthography

Media Lengua is primarily an oral language with a few written exceptions published within the last 13 years. These include a storybook compiled in 2013 (Stewart 2013), a cookbook published in 2021 (Prado Ayala et al. 2021), a dictionary published in 2020 (Stewart et al. 2020), and the Media Lengua Collection housed in the AILLA (Stewart and Prado Ayala 2025). When Media Lengua is written, it is based on phonetic Spanish orthography rather than the Unified Kichwa spelling system described in Section 2.2. In the case of the storybook *Stories and Traditions from Pijal*, the texts were transcribed entirely by the two native-speaker annotators in ELAN, with my later involvement limited to adding punctuation and confirming sentence boundaries. During this project, the annotators asked about the Unified Kichwa graphemes ⟨k⟩ vs. ⟨qu|c⟩, ⟨y⟩ vs. ⟨i⟩, and ⟨w⟩ vs. ⟨hu|gu⟩; after a brief discussion, they expressed a preference for using these forms, and I incorporated them by search-and-replace. Importantly, no discussion was had regarding vowel spelling or other orthographic choices, which remained entirely their own. Although speakers reported little difficulty reading with ⟨k⟩, ⟨y⟩, and ⟨w⟩, subsequent work has shown that community members overwhelmingly prefer to write in the more familiar Spanish-based system. As a result, all subsequent Media Lengua publications have used Spanish-based orthographic conventions, with ⟨k⟩, ⟨y⟩, and ⟨w⟩ appearing only sporadically. It is also worth noting that speakers independently debated on how to transcribe /z/, resulting in ⟨ts⟩.

When work began on the Media Lengua Collection, which was subsequently used to compile the Media Lengua dictionary, the same two native speakers were hired to transcribe and translate nearly 24 hours of recordings. While substantial spelling variations were observed in their transcriptions and translations, the vast majority involved *heterographs*. This term refers to two or more graphemes that represent a single phoneme, for instance, ⟨b⟩ vs. ⟨v⟩ for /b/ ([β]) (e.g., *abentana* vs. *aventana* both pronounced as /abẹ̣ntana/ 'to fan a fire'), ⟨h⟩ vs. ⟨Ø⟩ for /Ø/ (e.g., *helados* vs. *elados* both pronounced /ẹ̣ladọ̣s/ 'ice-cream'), ⟨hu(e|a)⟩ vs. ⟨gu(e|a)⟩ for /w(ẹ̣|a)/ (e.g., *hueco, gueco* 'hole' both pronounced /wẹ̣kọ̣/), ⟨c(a|o|u)⟩ vs. ⟨k⟩ vs. ⟨qu(i|e)⟩ for /k/ (e.g., *acavana, akabana* 'finish', both pronounced /akabana/ and *aqui* vs. *aki* 'here', both pronounced /aki/), ⟨s⟩ vs. ⟨z⟩ (e.g., *arros* vs. *arroz* 'rice' both pronounced as /azọ̣s/) etc. (all instances are present in the Media Lengua dictionary (Stewart et al. 2020)).

There was also spelling variation observed in Spanish origin words with variable modern vs. colonial-influenced pronunciations e.g., *abuela* /abuẹ̣la/ vs.

aguela /agu̯ela/ 'grandma' (Lapesa 1981: 468); *jabas* /xabas/ vs. *habas* vs. *abas* /abas/ 'fava beans' (Narbona Jiménez et al. 2022: 89); and words with stable colonial-influenced pronunciations such as *suedra* /su̯ɛdra/ 'mother-in-law' (Standard Spanish *suegra* /suegɾa/) (Navarro 1956: 421); *gomyta* /go̯mita/ (Standard Spanish *vomita* /bomita/ 'vomit') (Corominas & Pascual 1983 vol 5:842); and *limfiana* /limfiana/ 'to clean' (Standard Spanish *limpiar* /limpiaɾ/) (Corominas & Pascual 1984 vol 3:658–9).

Analyzing the degree of spelling variation in the Media Lengua dictionary (which does not account for frequency) reveals that of the 2,030 headwords with direct Spanish cognates, 22% showed variation from Standard Spanish based solely on consonant spelling (e.g., *trensa* vs. *trenza* 'braid'), while only 12% showed variation based solely on vowel spelling (e.g., *trasti* vs. *traste* 'dishes'). The degree of vowel divergence from Standard Spanish is far lower than the 47% and 64% mid-high vowel overlaps observed in acoustic space (see Figure 1 in § 2.4.1), suggesting that vowel spelling is not directly determined by acoustic proximity. Accordingly, the answer to the second research question in § 1, *Does acoustic overlap between mid and high vowels influence spelling?*, is likely no.

To interpret the 12% rate of vowel variation, it is useful to compare it with the patterns found in the consonants. As mentioned, the 22% consonant variation observed in the dictionary primarily reflects heterographs originating from Standard Spanish orthography. Because these variants correspond to the same sound, a relatively high level of written alternation is expected, even if the underlying phonology is fully stable. This provides a baseline for interpreting vowel behavior. If high and mid vowels in Media Lengua were simply alternative spellings of the same phoneme (i.e., heterographs), vowel variation would be expected to be similar to that of consonant heterographs (~22% variance). Conversely, if high and mid vowels were fully contrastive, written alternation should be essentially zero, since consonant *graphones* (graphemes with a one-to-one phoneme mapping) show virtually no variation in the dictionary (e.g., ⟨t⟩ /t/ is never written in place of ⟨d⟩ /d/ and ⟨sh⟩ /ʃ/ is never written in place of ⟨ch⟩ /ʧ/).

Instead, the observed rate (~12%) from the dictionary falls neatly between the 22% and 0% reference points. This suggests that the high and mid vowel distinctions are contrastive but carry relatively low functional load (i.e., they distinguish only a small set of lexical items). A low functional load along with the overlapping categories in acoustic space shows that, while the contrast exists, its lexical impact is limited. This might also suggest why descriptive accounts of Media Lengua vowel production vary (see Gómez Rendón 2005, 2008; Muysken 1997) and theoretical accounts (see van Gijn 2009; Muysken 2013) point to a non-stratified system.

The problem with this analysis is that the Media Lengua dictionary does not report the frequency of different spelling variations. Therefore, while the headword shows spelling X and the orthographic variation entry shows alternative spellings Y and Z, no information is provided that shows how stable a given spelling of a word is. Therefore, a novel approach is adopted here to account for written vowel variation that includes frequency trends, structural patterns, information about stability, and whether etymology plays a role in written vowel choice (see §3.3).

While this overview establishes how vowel spellings vary in written Media Lengua, understanding these patterns also requires an understanding of the underlying phonological system. The following section provides a description of Media Lengua phonology, focusing on the vowel inventory and contact-related influences that inform how vowels may be represented in writing.

2.4 Media Lengua phonology

Imbabura Media Lengua's phonotactic structure and phoneme inventory are primarily Imbabura Kichwa but contain several additions from Spanish. In terms of Kichwa mapping, Spanish /r/ (⟨r-⟩, & ⟨-rr-⟩) and /ʎ/ (⟨ll⟩) shift to Kichwa /ɻ/ and /ʒ/, respectively (Stewart 2020) and Spanish /f/ shifts to Kichwa /ɸ/ (Stewart, under-review). For instance, Media Lengua *carro* 'car, bus' and *rrogana* 'beg' are produced as /kaɻo̯/ and /ɻo̯gana/ and Media Lengua ***ll**ubia* 'rain' and *anillo* 'ring' are produced as /ʒubia/ and /aniʒo̯/. Spanish origin Intervocalic /s/ (⟨-s-⟩, ⟨-z-⟩, & ⟨-c(i|e)⟩) frequently conserves voicing from colonial Spanish (spelled as ⟨ts⟩ in Media Lengua) as seen in examples like ***ats**ena* [aˈʦe̯na] 'to do' from Colonial Spanish [xazeɾ]* (Modern Spanish *hacer* /aseɾ/ 'to do') (Muysken 1997 p. 381; Lapesa 1981: 567; Stewart 2011 p. 70). Additionally, Media Lengua retains /ʃ/ from Kichwa words instead of shifting to [ʧ], a phoneme that also appears in Ecuadorian Spanish in Kichwa borrowings (e.g., Media Lengua *jari**sh**ina* /xaɾiʃina/ 'poorly done', Imbabura Kichwa *jari**sh**ina*,[4] Unified Kichwa *kari**sh**ina* /kaɾiʃina/, Ecuadorian Spanish *cari**sh**ina* /kaɾiʃina/) (Stewart 2011: 29).

Prosodically, Media Lengua aligns with Imbabura Kichwa on nearly every level. Lexical stress is primarily penultimate, even if the original Spanish word

4. Imbabura Kichwa has both /k/ and /x/ as phonemes. The velar fricative /x/ functions in two ways: (i) as the northern reflex of aspirated /kʰ/ found in central and southern varieties (e.g., central/ south /kʰuru/ → Imbabura /xuru/ 'worm'), and (ii) as an independent phoneme in items such as /xatun/ 'large', attested across dialects. In local writing practices, /x/ is typically represented using Spanish ⟨j⟩ (e.g., ⟨*jatun*⟩ 'large') due to entrenched literacy norms, whereas Unified Kichwa represents /x/ with ⟨h⟩ in more formal contexts (e.g., *hatun* 'large').

had non-penultimate stress. (e.g., Media Lengua *abi**ta**cion* [abi'**ta**siọn] 'room' vs. Spanish *habita**ción*** [abita'**sion**] or Media Lengua *ce**du**la* [se'**du**la] 'ID card' vs. Spanish ***cé**dula* ['**se**dula]) (Stewart et al. 2020: 16, 560). Moreover, Media Lengua's intonational phonology is essentially Kichwa with a bitonal pitch accent (L+H*) on the penultimate syllable of nearly every prosodic word. Boundary tones are also shared with Kichwa with no evidence of Spanish-like pitch contours, and only minor instances of innovation (Stewart 2015a).

In contrast, Media Lengua rarely modifies consonant clusters absent in Kichwa (e.g., the /ɸl/ (⟨fl⟩) cluster in ***fl**aco* /**ɸl**akọ/ 'skinny', from Spanish ***fl**aco* /**fl**ako/ 'skinny', does not undergo epenthesis or elision) (Lipski 2020; Stewart accepted). Media Lengua has also adopted the voiced stop series from Spanish (/b, d, g/, ⟨b, d, g⟩) both productively and perceptually, and minimal/ near-minimal pairs are common (e.g., ***b**año* /**b**aɲọ/ 'bathroom' vs. ***p**año* /**p**aɲọ/ 'cloth'), whereas in the native Kichwa lexicon, voiced stops (barring some examples of /g/) only appear allophonically in native Kichwa vocabulary (Stewart 2015b, 2018a). However, it must be noted that voiced stops appear extensively in Kichwa in Spanish borrowings and speakers clearly contrast them productively and perceptually (Stewart 2015b, 2018a).

2.4.1 *Media Lengua vowel production*

The vowel system in Media Lengua is notably complex in how Spanish and Kichwa vowels are organized in acoustic space. This has attracted significant attention in the literature (van Gijn 2009; Gómez Rendón 2005, 2008; Hadley 2026; Muysken 1997; Onosson & Stewart 2021b, 2021a, 2023; Stewart 2011, 2014, 2018b; Stewart & Onosson 2023). Early descriptions (see e.g., Gómez Rendón 2005; Muysken 1997), largely based on impressionistic observation, offer differing descriptions as to how Spanish mid vowels are incorporated into Media Lengua.

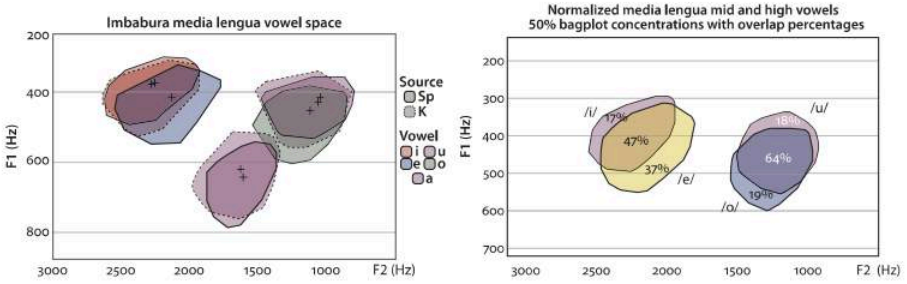
Gómez Rendón (2005) suggests that the mid vowels [e] and [o] appear in Media Lengua primarily in relexified roots and interjections, but that their presence is inconsistent; some roots retain the Spanish mid vowel entirely (e.g. *compra-* 'buy', *maneja-* 'drive'), while others show partial lowering or raising across vowels (e.g. *vendi-* 'sell', *ofreci-* 'offer'). He further argues the patterns are governed not by regular phonological processes, but by non-phonetic factors, resulting in a degree of phonetic variation exceeding that found in neighboring Kichwa varieties. To capture this variability, Gómez Rendón proposes a gradient model of relexification, ranging from pronunciations that fully preserve Spanish phonetic properties, through intermediate forms, to outcomes that align entirely with the Kichwa vowel system.

Muysken (1997) similarly characterizes Media Lengua as involving extensive phonological adaptation of Spanish forms to Kichwa, noting that Spanish mid

vowels are frequently replaced by high vowels. At the same time, he observes that [e] and [o] are more likely to be retained in names, interjections, stressed syllables, and certain high-frequency lexical items, while other high-frequency verbs consistently surface with high vowels. Together, these accounts point to a system in which mid vowel retention is lexically and contextually conditioned rather than categorically phonological.

From an acoustic perspective, Media Lengua mid vowels are substantially raised in acoustic space, leading to partial overlap (Figure 1A). To account for this, a raised diacritic below mid vowels ([e̝] and [o̝]) is used in both phonemic and phonetic transcriptions in this study. High vowels in Media Lengua can also exhibit a wide range of F1 frequencies, often produced in regions typically associated with mid vowels in a prototypical five-vowel system (Lipski 2015; Stewart 2011, 2014). The resulting system closely resembles that of Imbabura Kichwa (see Guion 2003; Lipski 2015; Stewart 2011, 2014), although Media Lengua maintains a somewhat greater separation between mid and high vowels than is typically found in Spanish borrowings into Kichwa.

To better understand how the overlapping categories might influence spelling, Stewart's (2014) Media Lengua vowel data were reanalyzed using normalized F1 and F2 values. This normalization accounted for individual speaker variation using random effect values from his mixed effects models. Bagplots were generated to visualize vowel clustering (Figure 1B), and histogram pixel counts were used to calculate overlap percentages. This plot shows that most vowels within the bagplots overlap; 47% for the front vowels and 64% for the back vowels.



A. B.

Figure 1. (A) Imbabura Media Lengua vowels plotted on the F1xF2 dimension, based on Onosson and Stewart (2021b), show overlapping mid and high vowels; (B) normalized Imbabura Media Lengua mid and high vowel bagplots with overlap percentages

2.4.2 *Media Lengua vowel perception*

Building on this apparent stratified system, Stewart (2018b) investigated whether Media Lengua listeners could aurally distinguish between mid and high vowels within these overlapping spaces. This investigation used a two-alternative forced choice (2AFC) identification task with 10 participants using minimal pairs: *piso-peso* ['pi.so-'pe.so] 'floor-weight'; *pipa-pepa* ['pi.pa-'pe.pa] 'pipe-seed'; *lona-luna* ['lo.na-'lu.na] 'tarp-moon'; *poma-puma* ['po.ma-'pu.ma] 'jug-puma'. For each token, the target vowel was altered by gradually shifting the first three formants (Hz), duration (ms), and pitch (fo) in equal increments across a ten-step continuum from a canonical high vowel to a canonical mid vowel. This design made it possible to locate where listeners shifted from identifying a vowel as 'high' to identifying it as 'mid', if the contrast were perceptually relevant to the listener. This allowed for the direct comparison between perceptual boundaries and the orthographic patterns analyzed here.

Results indicated that Media Lengua listeners reliably identified significant differences between canonical mid vowels (Figure 2, step 1) and canonical high vowels (Figure 2, step 10) across both the front and back series. These results suggest that Media Lengua listeners effectively navigate the acoustic overlap between mid and high vowels and consistently identify differences between them.

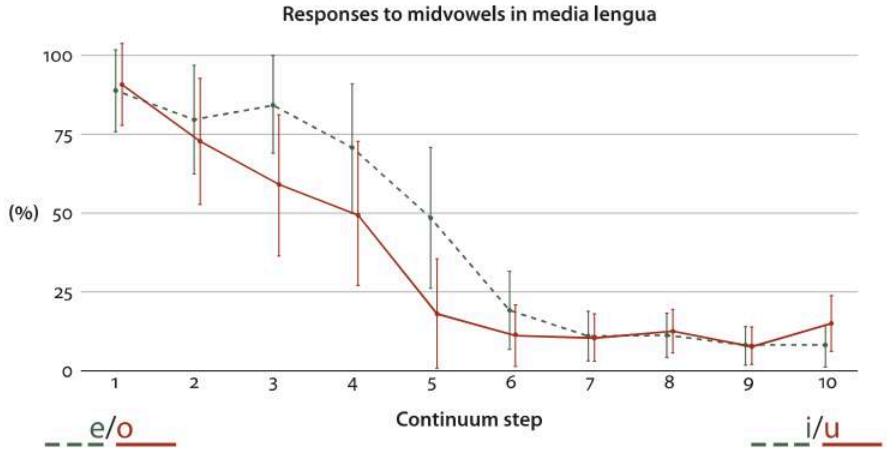


Figure 2. 2AFC identification task results for Media Lengua listeners averaged across the 10-step continua for the front (dashed) and back (solid) vowel series, based on Stewart (2018b)

2.4.3 *Media Lengua vowel sequences*

Onosson and Stewart (2021b) advanced the analysis of Media Lengua vowels by documenting how vowel sequences from both Spanish and Kichwa function in Media Lengua. Their study shows that nearly all Spanish origin and Kichwa origin vowel sequences appear in Media Lengua but have been adapted to fit a Kichwa-like vowel space. Their analysis included both Spanish-only vowel sequences (/ae, ea, oa, ei, ie, oe, ue, ao, eo, io, oi, eu, uo/) and the sequences /ia, ua, ai, ui, au, iu/, which appear in both Kichwa and Spanish. As expected, due to the overlapping mid and high vowel categories, vowel sequences often resemble monophthongs. However, changes in the formant trajectories, though small, distinguish them from true monophthongs in most cases (e.g., the F2 in /e̯i/ increases in frequency between the target vowels in *reina* [ze̯ina] 'laugh' whereas the F2 in /ie̯/ decreases in frequency between the target vowels in *acienda* [asie̯nda] 'ranch, farm'). The differences in trajectories between Spanish and Media Lengua are illustrated in Figure 3 for /e̯i/ (A) and /ie̯/ (B). These results provide further support that Media Lengua speakers successfully navigate the overlapping mid and high vowel categories and make subtle changes to vowel quality to produce consistent vowel sequences within a syllable nucleus.

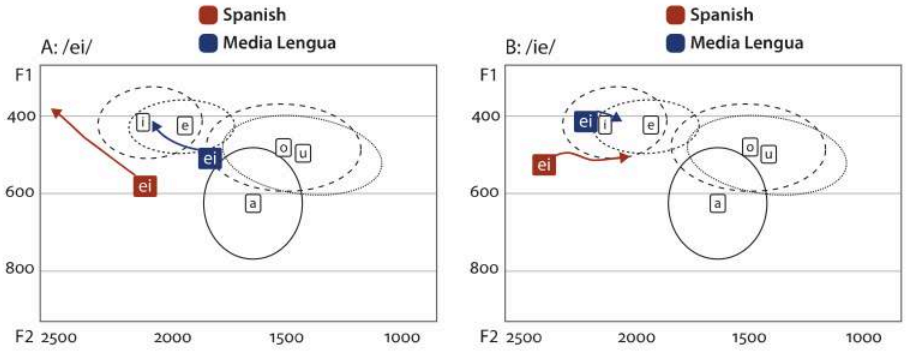


Figure 3. Vowel sequence 20%–80% trajectories of /ei/ (left) and /ie/ (right) overlaid on the Imbabura Media Lengua vowel space; background ellipses indicate one standard deviation. This figure is based on Onosson and Stewart (2021b)

3. Methods

3.1 Annotators

Two trilingual, native speakers of Imbabura Media Lengua and Imbabura Kichwa with early L2 Spanish, from the community of Pijal, Imbabura, Ecuador, tran-

scribed roughly equal portions of the Media Lengua collection used in this study (described in §3.2). They used ELAN's annotation mode to transcribe recorded conversations, narratives, and elicitations, while providing translations in both Imbabura Kichwa and in their local variety of Spanish. While both completed high school later in life, they still used variable phonetic spelling for both Kichwa and Spanish translations, likely a result of their lifelong language practices. Annotator 01 was born in 1971, and Annotator 02 was born in 1968 and both learned Spanish upon entering elementary school at age 5. This suggests that while their spelling practices are likely shaped by Spanish, the degree and types of variation observed in the dataset (see §3.3) indicate that these choices are also perceptually grounded in Media Lengua phonology. As detailed in Section 2.3, both transcribers consistently represent graphones in their spelling (e.g., ⟨t⟩ /t/ vs. ⟨d⟩ /d/) but show high degrees of variation with heterographs (e.g., ⟨s⟩ /s/ vs. ⟨z⟩ /s/).

All analyzed forms reflect the spellings produced directly by the transcribers. No guidance or correction was provided, and transcribers were instructed simply to write what they heard. As mentioned in Section 2.3, the only prior discussion regarding grapheme choice involved ⟨k⟩, ⟨w⟩, and ⟨y⟩. However, in later written practice, these preferences have fallen out of favor (see Stewart 2013 vs. Stewart et al. 2020 and Prado Ayala et al. 2021). It should also be reiterated that Standard Spanish and Unified Kichwa spellings are used only as external baselines for identifying orthographic divergence in the analysis; they were not used to adjust or normalize the dataset in any way.

Finally, although both annotators are fluent native speakers of Media Lengua and Imbabura Kichwa, Spanish is now dominant in their households and in most public interactions. While Media Lengua is falling out of use in favor of Spanish, it remains active, particularly in encounters with community members in passing who are known to speak the language and during communal work activities (*mingas*). Younger generations are now primarily Spanish-speaking, contributing to a broader shift away from both Media Lengua and Kichwa in everyday communication. Importantly, Media Lengua is virtually never written; as noted in Section 2.3, there are only three texts produced in the language, all initiated by the author, and aside from occasional Facebook posts (typically directed toward the author), there is no evidence of routine written use within the community.

3.2 Database

The dataset used in this project comes by way of the Media Lengua Collection housed in the AILLA, which represents the first and currently only large-scale repository of the Media Lengua language. The archive contains approximately 24 hours of audio and video material, primarily consisting of conversations, nar-

rations, and elicitations, with more than 68,000 individual words documented from over 40 speakers. The collection also includes detailed morpheme divisions and interlinear glosses for linguistic analysis.

The subset of the archive analyzed in this study comes from 12 hours, 17 minutes of conversational recordings annotated in 34 ELAN Annotated Format (EAF) files, representing 43 speakers (36 women, 7 men) from Pijal. Within this subset, the dataset contains 24,393 annotated word tokens and 1,641 unique lexical items.

3.2.1 *Dataset preparation*

The EAF files from the dataset mentioned in Section 3.2 were converted into a tab-delimited format for ease of analysis. A line-by-line comparison was conducted, where each word token was manually analyzed against its Standard Spanish or Unified Kichwa counterpart to determine if the vowels used in Media Lengua corresponded to the standard orthography of its source languages. In most cases, the pronunciation of vowels is reflected in the orthography of both Spanish and Kichwa, however this is not always the case. For instance, ⟨y⟩ and ⟨w⟩ in Unified Kichwa can be treated as both a consonant and vowel. This pattern reflects the fact that, in addition to its consonantal uses, Unified Kichwa does not permit adjacent written vowels. The graphemes ⟨y⟩ and ⟨w⟩ are therefore used to separate vowel sequences. For this reason, particular attention was given to the phonological role of these segments in the dataset e.g., whether they functioned as onsets or codas, or appear in complex nuclei. For example, the ⟨w⟩ in ***wasi*** /wasi/ 'house' and *pawana* /pawana/ 'jump', and the ⟨y⟩ in ***yana*** /jana/ 'black' and ***kuyi*** /kuji/ 'guinea pig' were treated as consonants, while the ⟨w⟩ in ***tawri*** /tauɾi/ 'lupin bean' and the ⟨y⟩ in ***kay*** /kai/ 'this' were treated as part of vowel sequences.

From Spanish, the sequences, *qu*+(e|i)(*qué* 'what'), *gu*+(e|i) (*guerra* 'war'), *gu*+[ao] (*guapo* 'handsome'), *hu*+v (*huerto* 'garden') contain a muted ⟨u⟩ in the first two examples while the latter two are typically produced as [g^{w}] or [w], respectively (e.g., [ke], [gera], [(g^{w}|)wapo], [weɾto], respectively). To this, vowel sequences were also treated as distinct vowel categories. For instance, ⟨i⟩ in ⟨ie⟩ was not analyzed as independent, but rather the ⟨ie⟩ cluster was analyzed separately from single written vowels. From Standard Spanish, the vowel categories under analysis included ⟨i⟩, ⟨e⟩, ⟨u⟩, ⟨o⟩, ⟨ae⟩, ⟨ai|ay⟩, ⟨au⟩, ⟨ea⟩, ⟨ei|ey⟩, ⟨eo⟩, ⟨eu⟩, ⟨ia⟩, ⟨ie⟩, ⟨io⟩, ⟨iu⟩, ⟨oi|oy⟩, ⟨ua⟩, ⟨ue⟩, ⟨ui⟩, and ⟨uo⟩, and from Unified Kichwa the vowel categories included ⟨i⟩, ⟨u⟩, ⟨uy⟩, ⟨ay⟩, ⟨aw⟩, and ⟨wa⟩.

The origin of each word was recorded (*Spanish* or *Kichwa*), along with the annotator's identification (*01* & *02*). The base form (for non-verbs), infinitive form (for verbs), or the suffix containing the vowel in question was also recorded. For instance, in cases where inflected forms such as *come-ni* 'I eat' and *come-nchi* 'we eat' were analyzed for ⟨e⟩ in the Spanish origin root, they were recorded as

come-na 'to eat' in the infinitive form. However, if ⟨i⟩ were under analysis in *-**ni*** 'first-person singular' or *-nch**i*** 'first-person plural', ⟨i⟩ would be recorded since it is part of the Kichwa suffixing morphology. The position of the vowel within the word was also noted: *initial* (e.g., ⟨o⟩ in ***o**lla* 'pot'), *root-final* (e.g., ⟨o⟩ in *perr**o**-ca* 'dog-TOP' or ⟨e⟩ in *come-na* 'eat'), *final* (e.g., *bald**e*** 'bucket' with no subsequent morphology), *root-internal,* not at a boundary, (e.g., ⟨o⟩ in *c**o**me-na* 'eat'), and as part of a *suffix* in Media Lengua's morphological structure (e.g., ⟨u⟩ or ⟨i⟩ in *come-naj**u**-nch**i*** 'eat-RECP-1P'). Table 2 provides a sample of the dataset used for analysis.[5]

Table 2. Sample of the Imbabura Media Lengua dataset illustrating examples of words analyzed for vowel variation

MediaLengua	Word	ML_Vowel	Standard_Vowel	SD	Origin	Annotator	Position
quimicuta	quimico	u	o	Different	Sp	2	Root_final
quimicuta	quimico	u	o	Different	Sp	2	Root_final
quimicutaka	quimico	i	i	Same	Sp	2	Root_internal
quimicutaka	quimico	u	o	Different	Sp	2	Root_final
qumico	quimico	i	i	Same	Sp	2	Root_internal
qumico	quimico	o	o	Same	Sp	2	Final
aborrido	aborrido	o	o	Same	Sp	1	Final
aborridoca	aborrido	o	o	Same	Sp	1	Root_final
abrircanchi	abrina	i	i	Same	Sp	1	Root_final
abrircanchi	1P	i	i	Same	K	1	Suffix
aguelacuna	PL	u	u	Same	K	1	Suffix
aguelapa	aguela	ue	ue	Same	Sp	1	Root_internal
aguelo	aguelo	o	o	Same	Sp	1	Final
aguelo	aguelo	ue	ue	Same	Sp	1	Root_internal
aguelocuna	aguelo	o	o	Same	Sp	1	Root_final
aguelocuna	aguelo	o	o	Same	Sp	1	Root_final
aguelocuna	aguelo	ue	ue	Same	Sp	1	Root_internal
aguelocuna	aguelo	ue	ue	Same	Sp	1	Root_internal
aguelocunaca	PL	u	u	Same	K	1	Suffix
aguelocunaca	PL	u	u	Same	K	1	Suffix

5. The entire dataset and R code for modelling can be found at: https://github.com/JesseStewart-LING/JesseStewart-LING.github.io/tree/main/publications/WVVML

3.2.2 *Frequency and count information*

While the dataset contains a substantial number of words (24,393 items from 1641 unique words), there is considerable variation in word frequency. Some words occur over 1000 times (e.g., *ese* 'this/ that/ the' with 1118 instances), while others appear only once (e.g., *brindarina* 'toast/cheers'). The mean frequency of word repeats is 15, with a median of four. Of the 1641 unique words in the dataset, 271 have a single occurrence. Although these words were included in the statistical models (§ 4.2), they are not discussed in detail due to their limited frequency.

3.3 Quantifying variation

To analyze the stability and variability of written vowels in Media Lengua, separate generalized linear mixed-effects models with a logit link (i.e., binary logistic GLMMs) were run for each vowel and vowel sequence under analysis. Models were created using the GLMER function from the *lme4* package (Bates et al. 2015).

Because repeated tokens from the same lexical item are not strictly independent (e.g., the spelling of one word may prime the same spelling for subsequent words), *Word* was modelled as a random intercept to capture item-level clustering. This allows each lexical item to have its own baseline intercept for spelling behavior, so that fixed-effect inference is not confounded by idiosyncratic properties of particular words. All models converged without singular fits (*isSingular* = FALSE). Goodness-of-fit was evaluated using AIC/BIC, residual inspection, and likelihood-ratio ANOVA tests comparing models that included *Word* as a random effect to models that did not (i.e., *glmer* vs. *glm* formulations). For all monophthong models, inclusion of the *Word* random intercept significantly improved model fit ($\Delta\chi^2 = 14–173$, $p < 0.001$), indicating that item-specific orthographic stability is a meaningful source of variance that should be explicitly modelled (Baayen, 2008; Barr et al. 2013). In contrast, for vowel sequence models (e.g., ⟨ie⟩, ⟨ia⟩, ⟨ea⟩), ANOVA tests were non-significant, which was expected given the comparatively small token counts and low variability observed in these categories (see § 4.1).

To determine the stability of vowel usage, Best Linear Unbiased Predictors (BLUPs) were extracted using the *ranef* function in R v.4.4.1 (R Development Core Team 2025) and added to the fixed intercept. These values then underwent inverse-logit transformation using *plogis*, and were converted to percentages (plogis*100). Words with values ≥ 99% were classified as stable/non-variable with respect to the vowel in their standard source language spelling, whereas values closer to 0% indicated stable use of a different vowel from Standard Spanish or Unified Kichwa conventions.

Each model contained every word with the target vowel or vowel sequence, regardless of its frequency in the dataset. However, to focus on words with sufficient data, those with fewer than three instances were excluded from the figures. This approach allows us to concentrate on words with higher counts, which are more likely to reveal patterns of written vowel variation.

The initial model-building phase began with the full set of theoretically motivated predictors for each vowel category identified in Section 4.1. These included *Origin*, to determine whether variation may be etymologically motivated; *root-final*, to test whether there is more variation in written vowels that appear at the end of lexical roots just before bound Kichwa origin suffixation begins; and *Annotator* to assess whether there was variation in how the two annotators transcribed specific written vowels. A minimal-adequate-model framework using backward elimination was employed to select the fixed effects used in the models (Baayen, 2008, pp. 186, 205; Johnson, 2008, p. 90; Upton, 2017, p. 90). Predictors that did not show evidence of contributing to the models were removed one-by-one based on the closest z-value[6] to zero. The only random effect tested was *Word*.

In all models, the dependent binary variable *Same-Different* was used. This indicated whether the written vowel under analysis in each word was the 'same' or 'different' from its original Standard Spanish or Unified Kichwa spelling.

For example, if the vowel under analysis were ⟨e⟩ in the Spanish origin word *come-ni* ('I eat', from Spanish *comer* 'to eat'), it would be marked as 'same'; if it were written as ⟨i⟩ in *comi-ni*, it would be marked as 'different.' For Kichwa, if the vowel under analysis were ⟨u⟩, as in the Kichwa origin morpheme *-shun* '1.PL.FUT' in *comi-shun* 'we will eat', it would be marked as 'same'; if it were written as ⟨o⟩ as in *comi-shon*, it would be marked as 'different', since the Unified Kichwa word contains ⟨u⟩ in this position (*miku-shun* 'we will eat').

4. Results

All vowels/ vowel sequences are examined in this section: ⟨i⟩, ⟨e⟩, ⟨u⟩, ⟨o⟩, ⟨ie⟩, ⟨ia⟩, ⟨ea⟩, ⟨ai⟩, ⟨io⟩, ⟨ue⟩, and ⟨ua⟩. However, ⟨ai⟩, ⟨io⟩, ⟨ue⟩, and ⟨ua⟩ were not frequent enough in the dataset to warrant statistical modelling, and are therefore discussed descriptively in Section 4.2.6. Section 4.1 presents descriptive statistics based on the raw data, while Section 4.2 presents inferential analyses that model orthographic variation, with particular attention to word-level effects.

6. A z-value is the statistic used in GLM/GLMM output to calculate the p-value; larger absolute z-values correspond to smaller p-values.

4.1 Descriptive statistics

The following sections describe general orthographic variation by vowel (§ 4.1.1), annotator influence on written vowel choice (§ 4.1.2), positional effects on written vowel choice (§ 4.1.3), and differences in written vowel choice based on a word's origin (§ 4.1.4). It is important to note that these data are based on every word in the dataset, therefore, words with high token repeats could potentially skew the actual representation of the data, which is accounted for in Section 4.2.

4.1.1 *Orthographic variation by vowel*

The analysis of the raw data in Figure 4 shows that, for monophthongs, the vowel ⟨o⟩ demonstrates the highest alignment with its original Spanish spelling, with 95% of words retaining the original source vowel. The vowels ⟨i⟩ and ⟨e⟩ also display high levels of alignment, with 91% and 89% of words, respectively, using Standard Spanish (for ⟨i⟩ & ⟨e⟩) and Unified Kichwa (for ⟨i⟩) spelling. The vowel ⟨u⟩ exhibits the most variation among monophthongs, with just 82% of words spelled with the same vowels as their Standard Spanish or Unified Kichwa counterparts. All vowel sequences show over 90% consistency with their original spelling, except for ⟨ea⟩, which only retains Spanish spelling in 26% of the cases.

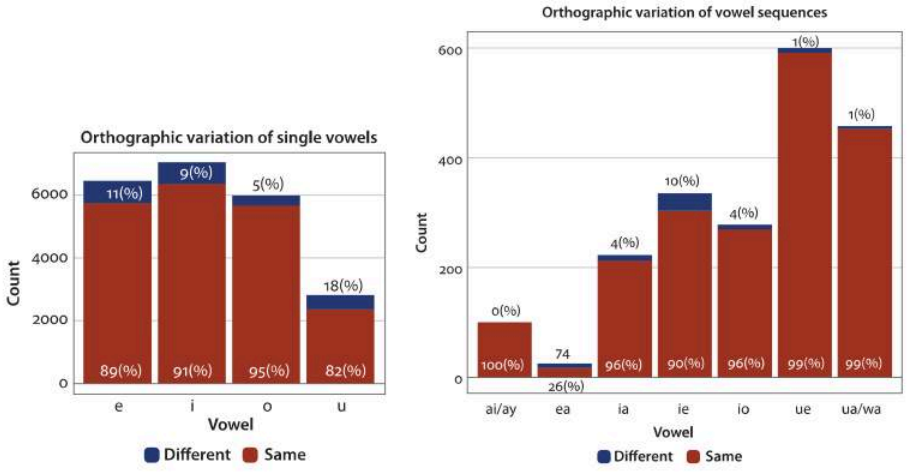


Figure 4. Orthographic variation of single written vowels (Panel A) and written vowel sequences (Panel B) in Imbabura Media Lengua, highlighting the proportion of vowels that are spelled the same as, or differently from, their counterparts in Standard Spanish and Unified Kichwa

4.1.2 *Orthographic variation by annotator*

For monophthongs (Figure 5A), Annotator 01 shows slightly higher alignment with 92% of vowels spelled the same as their Standard Spanish and Unified Kichwa counterparts. Annotator 02 shows a similar pattern but with slightly less consistency, with 89% of vowels matching the standard source language orthographic norms. This suggests a small but noticeable difference in spelling practices between the two annotators for monophthongs. For vowel sequences (Figure 5B), both annotators demonstrate high alignment with Annotator 01 showing 96% while Annotator 02 shows 95% consistency.

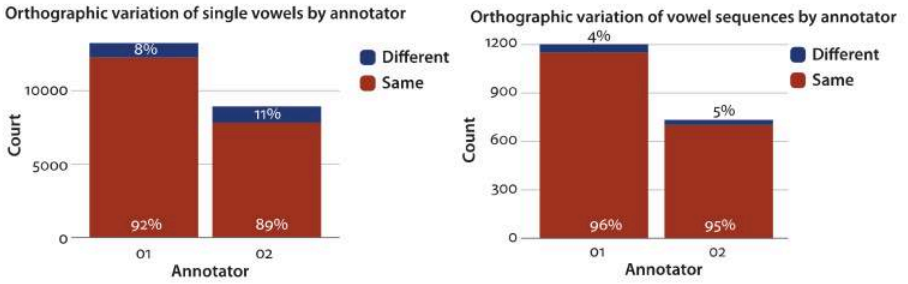


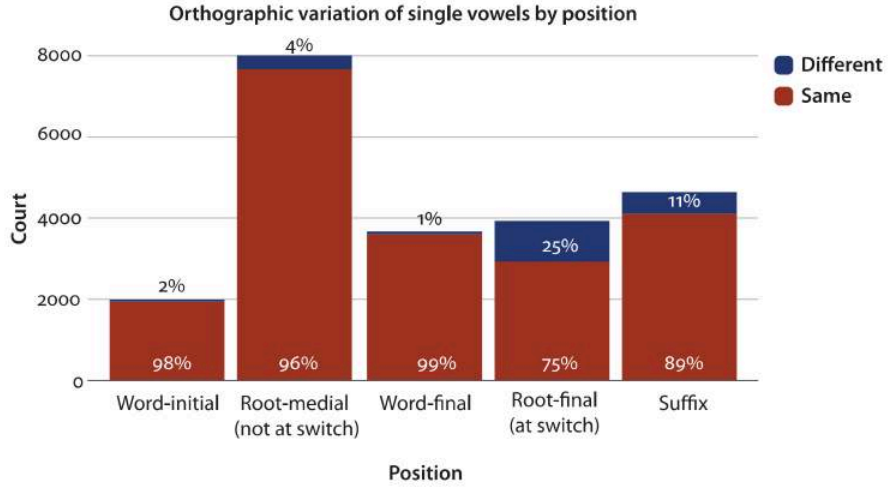
Figure 5. Orthographic variation of single written vowels (A) and written vowel sequences (B) by annotator (coded as 01 & 02), displaying the proportion of vowels spelled the same as, or differently from, their Standard Spanish or Unified Kichwa counterparts

4.1.3 *Orthographic variation by position*

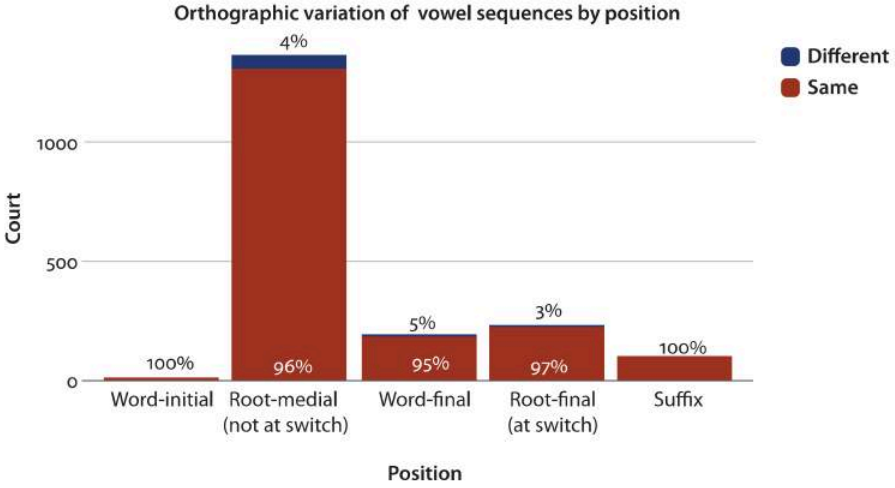
For monophthongs (Figure 6A), word-initial and word-final positions show the highest alignment, with 98% and 99% of written vowels retaining the standard source language spelling respectively. Root-medial position shows 96% alignment, indicating that most vowels retain their original spelling in this context as well. However, greater variation (25%) from the standard source language spelling is observed in root-final position before bound morphology. Additionally, vowels in a suffix show an 11% shift from Unified Kichwa spelling. These results suggest that written vowels at morphological boundaries (typically Spanish in origin) and in suffixes (nearly always Kichwa in origin) are more likely to be represented with vowel spellings that do not align with source language orthographic norms.

For diphthongs (Figure 6B), minimal variation is observed across different positions. The 4% variation observed occurs in root-internal positions, and in

Media Lengua, ⟨ea⟩ is only found in this position (see ⟨ea⟩ variation in Figure 6B). This consistency across positions suggests that vowel sequences are generally stable in Media Lengua, even at morphological boundaries.



A.



B.

Figure 6. Orthographic variation of single written vowels (A) and written vowel sequences (B), displaying the proportion of vowels spelled the same as, or differently from, their Standard Spanish or Unified Kichwa counterparts

4.1.4 *Orthographic variation by origin*

For single written vowels (Figure 7A), words of Spanish origin show a high degree of alignment, with 92% of vowels retaining their original spelling and 8% showing variation. Words and suffixes of Kichwa origin exhibit slightly more variability, with 87% of vowels following Unified Kichwa norms. This suggests that single vowels of Kichwa origin are more prone to orthographic variation than those of Spanish origin. For vowel sequences (Figure 7B), the data show high consistency across both origins. Vowel sequences of Kichwa origin retain a spelling that reflects standard pronunciation 100% of the time (e.g., /a̲i/ appears as either ⟨ai⟩ or ⟨ay⟩, but never as e.g., ⟨ae⟩). Vowel sequences of Spanish origin also demonstrate strong consistency, with 97% using a spelling that reflects standard pronunciation. This suggests that written vowel sequences generally coincide with graphemes used to represent the pronunciation in the standard languages, with only minimal variation observed, particularly in those of Spanish origin.

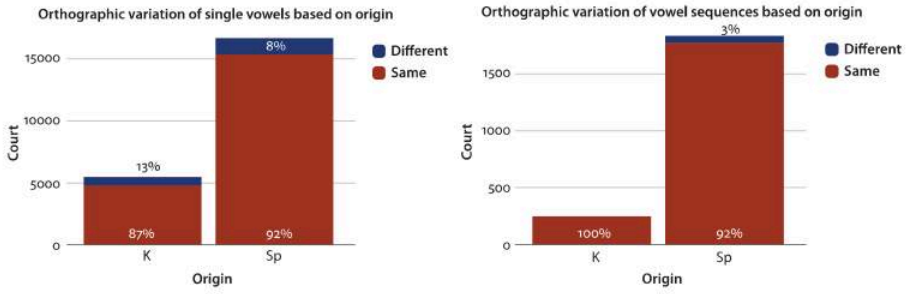


Figure 7. Orthographic variation of single written vowels (A) and vowel sequences (B) in the data based on their origin, either Kichwa (K) or Spanish (Sp)

4.1.5 *Summary of orthographic variation*

The raw data analysis of orthographic variation across these four factors reveals several patterns. First, there is high consistency in the spelling of monophthongs and vowel sequences, with most vowels matching their source language standardized spellings. However, certain vowels like ⟨u⟩ and the vowel sequence ⟨ea⟩ show either greater variability or stable non-standard vowel usage.

Differences between annotators are minimal, suggesting consistent spelling practices overall, though some minor variation exists in written single vowels. Positional analysis shows that vowels are most stable at word-initial and word-final positions, with increased variability at morphological boundaries (root-final

position). Finally, vowels of Kichwa origin tend to exhibit slightly more variability than those of Spanish origin, particularly for single vowels, while vowel sequences remain largely stable across both origins.

4.2 Inferential analysis

This section provides inferential statistical analyses to identify the factors that significantly influence orthographic variation in Media Lengua. Each analysis was conducted using generalized linear mixed-effects models (GLMMs). All models initially included the four predictors introduced in §3.2.1; non-significant factors were subsequently removed following the backward elimination approach described in §3.3 to arrive at the final models reported here. *Word* was included as a random intercept to capture item-specific deviations from the overall pattern. The statistical focus of the analyses is therefore on the extent to which individual words diverge from population-level tendencies, as well as on the fixed effects that systematically condition variation across tokens.

Figure 8 through 12 show item-specific (lexical and bound morphology) predicted probabilities derived from the models. These predictions are obtained by combining the fixed intercept with the word-level random intercepts (BLUPs), yielding baseline probabilities for each word. If significant, predictors that vary within a lexical item (e.g. root-final position), condition-specific predictions incorporate the corresponding fixed-effect coefficient in addition to the word-level random intercept. Predictors that are constant within a word (e.g. lexical origin) are already conditioned into the random intercept and are therefore not reapplied.

Words shown in the figures exhibit more than 1% variation from the source language standards and occur at least three times in the dataset. Root words and morphemes are spelled using their standard source language orthography, with the exception of Spanish accents and compound words. Spanish origin verbs are presented with the Media Lengua infinitive suffix *-na* to aid in identifying them as verbs vs. non-verbs. Each point reflects the predicted probability of a vowel realization for a given word or bound morpheme, with token frequency represented by point size and lexical origin indicated by color. Vertical reference lines mark stability zones; ≤20% indicates stability with a non-standard source language vowel; ≥80% indicates stability with a source-language standard vowel; and words between these zones show considerable variability between standard and non-standard source language vowels.

Words marked with an asterisk contain the same origin vowel more than once (e.g., *tenena* 'have' or *vivina* 'live', from Spanish *tener* and *vivir*, respectively). In such cases, a vowel may be stable in one position but variable in another, or both

vowels may be variable. This is a known limitation of the graphing method, and such cases are discussed in the accompanying text.

4.2.1 ⟨*i*⟩ *variation*

This model examines the probability of ⟨i⟩ being realized as ⟨i⟩ versus another vowel (specifically ⟨e⟩). The only fixed effect included in the model was *Root-final position*. The model output (Table 3) indicates that the intercept value for ⟨i⟩ is 7.59, which represents the log-odds of Standard Spanish and Unified Kichwa ⟨i⟩ being realized as ⟨i⟩ in Media Lengua in the dataset under baseline conditions. This highly positive intercept suggests a strong tendency for the Media Lengua ⟨i⟩ to match ⟨i⟩ in its source language orthographies rather than changing to ⟨e⟩.

Table 3. Model summary of ⟨i⟩ variation, showing the intercept and the effect of vowel position at the morphological boundary (root-final position) on the likelihood of⟨i⟩ being realized as ⟨i⟩ versus another vowel

⟨i⟩*	Estimate	Std. Error	z value	Pr(>|z|)
Intercept: Same/ Different	7.59	0.71	10.72	< 2e-16
Position: Root-final	−0.99	0.24	−4.16	3.14e-05

* Column headers left to right: ⟨*v*⟩ is vowel under analysis; *Estimate* indicates the model coefficient (log-odds) for each predictor; *Std. Error* shows the uncertainty of that estimate; *z-value* is the ratio of the estimate to its standard error; and *Pr(>|z|)* reports the corresponding two-tailed p-value testing whether the coefficient differs from zero.

The coefficient estimate for *root-final* is −0.99, indicating that when ⟨i⟩ occurs just before bound morphology (*impedi-rca-ni* 'prevent-PST-1'), the log-odds of it being realized as ⟨i⟩ decrease. Since ⟨e⟩ is the only alternative to ⟨i⟩ in the dataset, this corresponds to a slightly increased likelihood of ⟨e⟩ appearing in root-final position (*impede-rca-ni*).

Figure 8 illustrates predicted variation in the written realization of ⟨i⟩ in Media Lengua words and bound morphemes. Items displayed represent approximately 8% of all ⟨i⟩ tokens (46 out of 586). Several items show high stability despite diverging from their source language orthography. These include the Spanish origin word *cintura* 'waist', which appears almost categorically with ⟨e⟩, spelled *senturas* in nearly all instances. Similarly, *Kayampi*, a place name from the pre-Columbian Kara language with no living speakers, comes into Media Lengua via Kichwa, *Kayampi* and Spanish, *Cayambe*, and shows a strong preference for ⟨e⟩, reflecting Standard Spanish orthography.

Several high-frequency Kichwa origin bound morphemes show minor variation between ⟨i⟩ and ⟨e⟩, despite the absence of /e/ phoneme in Unified Kichwa's

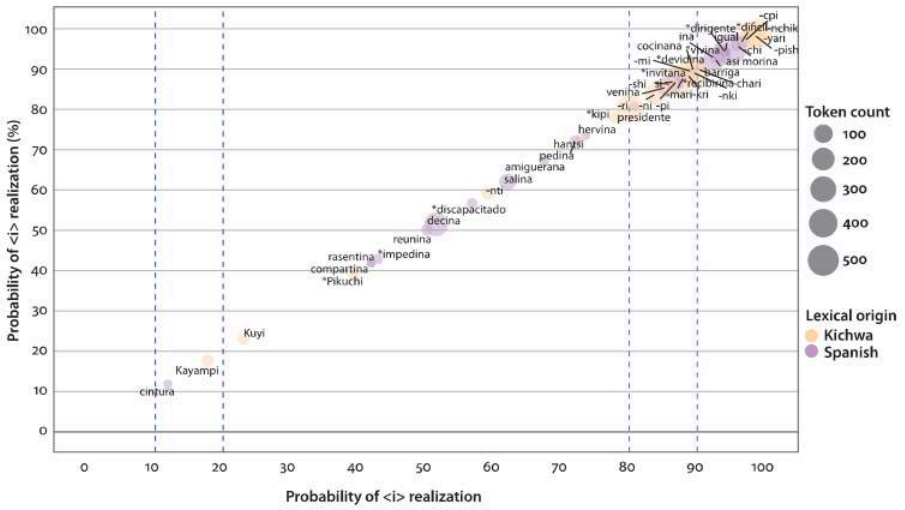


Figure 8. Bubble plot illustrating ⟨i⟩ variation in Imbabura Media Lengua from the random effects output, predicting the percentage similarity to Unified Kichwa and Standard Spanish. Each bubble represents a word, color-coded by origin (Kichwa or Spanish) and sized according to its frequency in the dataset. The plot includes only words with less than 99% stability and at least three tokens

vowel system and ⟨e⟩ in Unified Kichwa orthography. These include *-ri* 'reflexive', *-ni* 'first person', *-mi* 'validative', *-mari* 'affirmative', *-kri* 'future', *-chi* 'causative', and *-pi* 'locative'. Most of these morphemes cluster around or above 80% ⟨i⟩ realization. One notable exception is *-nti* 'comitative', which shows the highest degree of variability among Kichwa morphemes, with approximately 43% of its tokens appearing with ⟨e⟩ (*-nte*).

Several Spanish origin lexical items also show high stability with ⟨i⟩, clustering above 80%. These include verbs such as *cocinana* 'cook', *venina* 'come', and *morina* 'die' (from Spanish *cocinar, venir,* and *morir*, respectively), all of which largely preserve Spanish ⟨i⟩ despite some variability. Three additional words containing two ⟨i⟩s also cluster in this range suggesting that both vowels are stable. These include *vivina* 'live', *invitana* 'invite', and *dirigente* 'manager' (from Spanish *vivir, invitar,* and *dirigente*, respectively). The word *asi* 'like that' (from Spanish *así*) is the most frequent word displayed in Figure 8 (544 tokens) with only 7% of the instances containing ⟨e⟩.

Figure 8 also shows that the greatest variability between ⟨i⟩ and ⟨e⟩ occurs in a set of words, many of them verbs, that vary in root-final position, consistent with the model's predictions. These include *decina* 'say', *hervina* 'boil', *salina* 'leave', *pedina* 'ask', *compartina* 'share', and *resentina* 'resent' from Spanish *decir, hervir,*

salir, pedir, compartir, and resentir, respectively. Among these, *decina* is of particular interest given its high token count (225) and high variability with approximately 48% of instances appearing with ⟨e⟩.

Additional items falling within the variable 20–80% range include words with two Standard Spanish origin ⟨i⟩s. These include *impedina* 'prevent' and *discapacitado* 'disabled' (from Spanish *impedir* and *discapacitado*), both of which show stable shift of the first ⟨i⟩ to ⟨e⟩ while retaining ⟨i⟩ in the second position. A similar pattern is observed in the place name *Pikuchi* (from Kara via Kichwa *Pikuchi*, Spanish *Peguche*), where the first vowel consistently appears as ⟨e⟩, while the second alternates between ⟨i⟩ and ⟨e⟩. The Kichwa origin word *kipi* 'knapsack' also occupies this intermediate range, with two of the eight instances appearing as *quipe* rather than *quipi*.

Finally, *kuyi* 'guinea pig' presents an interesting case. The alternation between ⟨i⟩ and ⟨e⟩ likely reflects the phonemic proximity of ⟨y⟩ (/j/) and ⟨i⟩ (/i/), where lowering the vowel may increase syllabic salience. Notably, Spanish also borrows this word from Kichwa, but as *kuy* (/kui/), also avoiding the ⟨yi⟩ sequence.

Overall, the results suggest that written ⟨i⟩ variation in Media Lengua is neither random nor completely uniform. Instead, it reflects a combination of lexical origin, morphological position, orthographic influence from Spanish, and item-specific patterns, with the greatest variability concentrated in root-final verbs.

4.2.2 ⟨*e*⟩ *variation*

This model examines the probability of the ⟨e⟩ being realized as ⟨e⟩ versus other vowels (specifically ⟨i⟩). Both *Annotator* and *Root-final position* were included as fixed effects. The model output (Table 4) indicates that the intercept value is 6.77, which represents the log-odds of the vowel ⟨e⟩ being realized as ⟨e⟩ in the dataset under baseline conditions. This highly positive intercept suggests a strong tendency for the vowel ⟨e⟩ to remain ⟨e⟩ in most contexts.

Table 4. Model summary of ⟨e⟩ variation showing the intercept and the effects of annotator and root-final position on the likelihood of ⟨e⟩ being realized as ⟨e⟩ versus another vowel

⟨e⟩	Estimate	Std. Error	z value	Pr(>|z|)
Intercept: Same/ Different	6.77	0.72	9.41	< 2e-16
Annotator: 02	−0.34	0.13	−2.56	0.0106
Position: Root-final	−3.25	0.24	−13.81	< 2e-16

The coefficient estimate for *Annotator* (#02) is −0.34 log-odds, indicating that *Annotator* negatively affects the likelihood of ⟨e⟩ being realized as ⟨e⟩. Specifi-

cally, the log-odds decrease when the second annotator transcribes the data, suggesting very minor variability (⟨1%) introduced through this annotator's vowel choice. This negative effect implies that annotator-specific decisions may influence vowel realization, but not substantially. Since ⟨i⟩ is the only alternative to ⟨e⟩ in the dataset, this corresponds to a slightly increased likelihood of ⟨i⟩ being written by annotator #2.

The coefficient estimate for *Root-final position* is −3.25 log-odds, indicating a strong negative effect on the likelihood of ⟨e⟩ being realized as ⟨e⟩ in root-final position. In other words, when a vowel occurs at the end of a lexical root and subsequently followed by bound morphology, the log-odds of selecting ⟨e⟩ drop substantially. This corresponds to an increased likelihood of ⟨i⟩ appearing in this position. Unlike the minor annotator effect, the magnitude of this coefficient suggests that root-final position is a substantial factor on vowel spelling, reflecting a predictable positional pattern rather than idiosyncratic transcription behavior.

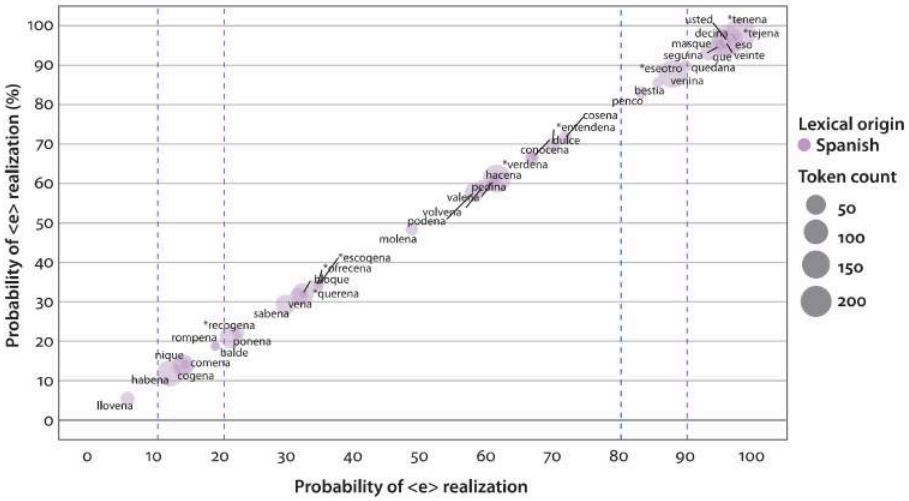


Figure 9. Bubble plot illustrating ⟨e⟩ variation in Media Lengua, showing the percentage similarity to Standard Spanish. Each bubble represents a word, sized according to its frequency in the dataset. No Kichwa origin words appear in this figure as ⟨e⟩ is not present in the in the Unified Kichwa writing system or phoneme inventory

Figure 9 illustrates predicted variation in the written realization of ⟨e⟩ in Media Lengua words. Words displayed represent approximately 6% of all ⟨e⟩ tokens (40 of 628). This subset includes Spanish origin lexemes only, as Unified Kichwa does not contain the vowel ⟨e⟩ as either a phoneme (/e/) or as part of its standardized orthography. Therefore, the words plotted span from non-standard, Kichwa-influenced ⟨i⟩ (left), to standard, Spanish-influenced ⟨e⟩ (right).

At the lower end of the scale (≤20%) are words predicted to be consistently spelled with ⟨i⟩ rather than ⟨e⟩. These include verbs such as *llovena* 'rain', *habena* 'have', *cogena* 'take', *comena* 'eat', and *rompena* 'break'. The verbs *recogena* 'gather' (with two ⟨e⟩s) and *ponena* 'put' fall close to the 20% boundary, also suggesting a strong preference for ⟨i⟩ over ⟨e⟩. Non-verbs in this range include the Spanish origin univerbation *nique* 'nothing', which frequently appears as *niqui*, and *balde* 'bucket', which is consistently spelled as *baldi*.

Words showing high ⟨e⟩ stability (≥80% or higher) form a cluster toward the upper end of the plot. These include verbs such as *decina* 'say', *quedana* 'stay', *seguina* 'follow', and *venina* 'come', all of which contain root-internal ⟨e⟩ rather than root-final vowels in line with the model predictions. Also included in this group are *tenena* 'have' and *tejena* 'sew', both showing stable ⟨e⟩ realization in multiple vowel positions. Non-verbs with stable ⟨e⟩ include *penco* 'agave', *bestia* 'idiot', *que* 'what', *eso* 'that', the second person formal pronoun *usted*, and compounds such as *masque* 'whatever' and *eseotro* 'the other'.

Between 20% and 80% lies a group of words showing a high degree of variability. Many of these are verbs with alternation in root-final position, as predicted by the model. These include *sabena* 'know', *vena* 'see', *molena* 'grind', *podena* 'can', *volvena* 'come back', *valena* 'useful for', *hacena* 'do', *conocena* 'know', and *cosena* 'sew'. The only verb in this range that shows root-internal variation is *pedina* 'ask', frequently appearing as *pidina*. Words containing multiple instances of ⟨e⟩ also appear in this intermediate zone, including *querena* 'want', *ofrecena* 'offer', *escogena* 'select', *verdena* 'make green', and *entendena* 'understand', the latter containing three potential ⟨e⟩ sites. Non-verbs with variable ⟨e⟩ include *bloque* 'cement block', which often appears as *bloqui*, and dulce 'sweet', frequently surfacing as *dulci*.

4.2.3 ⟨*u*⟩ *variation*

This model examines the probability of the vowel ⟨u⟩ being realized as ⟨u⟩ versus another vowel in the dataset (specifically ⟨o⟩). Both *Annotator* and *Origin* were included as fixed effects. The model output (Table 5) shows that the intercept value for ⟨u⟩ variation is 6.64, representing the log-odds of the vowel ⟨u⟩ being realized as ⟨u⟩ in the dataset under baseline conditions. This highly positive intercept suggests a strong tendency for the vowel ⟨u⟩ to remain ⟨u⟩ rather than changing to ⟨o⟩ in most contexts.

The coefficient estimate for Annotator #2 is −0.41 log-odds, indicating that the second annotator is slightly less likely to write ⟨u⟩ as ⟨u⟩ compared to annotator #1. This effect is statistically significant but small, suggesting minor annotator-specific variation in vowel transcription, again corresponding to a slightly increased likelihood of the alternative spelling with ⟨o⟩.

Table 5. Model summary of ⟨u⟩ vowel variation, showing the intercept and the effects of annotator and origin on the likelihood of ⟨u⟩ being realized as ⟨u⟩ versus another vowel

⟨u⟩	Estimate	Std. Error	z value	Pr(>|z|)
Intercept: Same/ Different	6.64	1	6.65	1.65e-11
Annotator: 02	−0.41	0.14	−2.98	0.00287
Origin: Spanish	2.91	1	2.91	0.00351

In contrast, the effect of Origin (Spanish) is both positive and substantially larger (Estimate = 2.91 log-odds), indicating that vowels occurring in Spanish origin lexical items are more likely to be written as ⟨u⟩. This effect suggests that etymological source plays a meaningful role in orthographic choice for /u/, with Spanish origin words showing greater stability in high vowel spelling. Together, these results indicate that while annotator effects exist, orthographic patterns for /u/ are driven primarily by lexical origin rather than transcriptional idiosyncrasy.

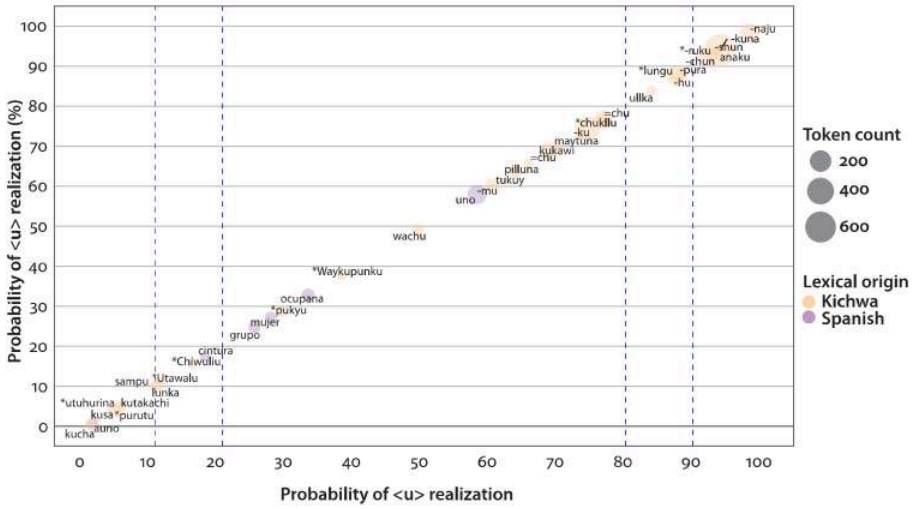


Figure 10. Bubble plot illustrating ⟨u⟩ variation in Media Lengua from the random effects output, showing the percentage similarity to Unified Kichwa and Standard Spanish. Each bubble represents a word, color-coded by origin (Kichwa or Spanish) and sized according to its frequency in the dataset. The plot includes only words with less than 99% stability and at least three tokens

Figure 10 illustrates predicted variation in the written realization of ⟨u⟩ in Media Lengua words and bound morphemes. Items displayed represent approximately 12% of all ⟨u⟩ tokens (37 out of 304). Unlike the patterns observed for

⟨i⟩ and ⟨e⟩, variation in ⟨u⟩ is not driven by root-final position, which did not reach significance in the model. Instead, the distribution of ⟨u⟩ and ⟨o⟩ spellings is largely conditioned by lexical origin; Spanish origin words with underlying /u/ are spelled more consistently with ⟨u⟩, whereas Kichwa origin words and bound morphemes show substantially more alternation between ⟨u⟩ and ⟨o⟩. This is evidenced by the domination of Kichwa items in Figure 10, many of which favor ⟨o⟩ despite the absence of /o/ in Unified Kichwa.

Some of the stable non-standard spellings among Kichwa origin items include the lexical items *kucha* 'lake', *lunku* 'youth', *purutu* 'bean', *utuhurina* 'make holes', and *kusa* 'husband'. Several of these forms are also borrowed into Ecuadorian Spanish with ⟨o⟩ (e.g. *cocha*, *longo*, *poroto*, and *sambo,* respectively), suggesting that Spanish orthographic conventions influence their written realization in Media Lengua. *Utuhurina* contains three /u/s in Unified Kichwa but shows near-categorical ⟨o⟩ usage across the word in this dataset and does not appear to be borrowed into Spanish (though it might be found rural varieties). Speakers often jokingly say that there is no difference between (Kichwa) *kusa* 'husband' and (Spanish) *cosa* 'thing', which may suggest that the ⟨u⟩ in *kusa* is pronounced closer to /o/ than /u/.

The only two Spanish origin words that sit in the low range are *auno* 'not yet', which frequently appears as *aunu,* and *cintura* 'waist', which often appears as *sentoras.* Several place names similarly favor Spanish-style ⟨o⟩; *Kutakachi* (Sp *Cotacachi*), *Utawalu* (Sp *Otavalo*), and *Chiwuliyu* (Sp *Chibuleo*), all of which appear clustered near the ≤20% boundary.

The remaining few Spanish origin words in Figure 10 show variable ⟨u/o⟩ usage rather than categorical behavior. These include *uno* 'one', *mujer* 'woman', *grupo* 'group', and *ocupana* 'occupy', all of which fall within the intermediate range of the plot.

Kichwa origin bound morphemes make up a large portion of Figure 10. High-frequency morphemes such as *-mu* 'translocative', *=chu* 'negative/polar question', and *-ku* 'diminutive' show notable variability between ⟨u⟩ and ⟨o⟩, with *-mu* showing the greatest variation (44% ⟨o⟩). Other morphemes, including *-hu* 'progressive', *-chun* 'different-subject purposive', *-nahu* 'reciprocal', *-shun* 'future 1PL', and *-kuna* 'plural', show higher stability with ⟨u⟩, though minor variation remains. The morpheme *-ruku*, containing two instances of ⟨u⟩, is highly stable with ⟨u⟩ in both medial and final position.

Additional Kichwa origin lexical items showing variable ⟨u/o⟩ usage include *wachu* 'furrow', *Waykupunku* (place name; Sp *Huaycopungo*), *pukyu* 'slope', *chukllu* 'corn', *kukawi* 'potluck food', *maytuna* 'wrap', *anaku* 'skirt', and *ullka* 'ball of wool'. Nearly all these words are borrowed into Spanish with ⟨o⟩ including *huacho*

'furrow', *choclo* 'corn', *anaco* 'skirt', *cocayo* 'shared snack', *mayto* 'set of diapers', and *anaco* 'skirt'.

4.2.4 ⟨o⟩ *variation*

This model examines the probability of the ⟨o⟩ being realized as ⟨o⟩ versus other vowels (specifically ⟨u⟩). The only fixed effect included in the model was *Root-final position*. The model output (Table 6) indicates that the intercept value for ⟨o⟩ variation is 10.47, which represents the log-odds of the vowel ⟨o⟩ being realized as ⟨o⟩ in the dataset under baseline conditions. This highly positive intercept suggests a strong tendency for the vowel ⟨o⟩ to remain ⟨o⟩ in most contexts.

Table 6. Model summary of ⟨o⟩ variation, showing the intercept and the effect of vowel position at the morphological boundary (root-final position) on the likelihood of⟨o⟩ being realized as ⟨o⟩ versus another vowel

⟨o⟩	Estimate	Std. Error	z value	Pr(>|z|)
Intercept: Same/ Different	10.47	0.77	13.6	⟨2e-16
Position: Root-final	−3.37	0.39	−8.56	⟨2e-16

The coefficient for *Root-final position* is −3.37 log-odds, indicating that when the vowel occurs at this position, the log-odds of ⟨o⟩ being realized as ⟨o⟩ decrease. This negative effect suggests that ⟨o⟩ at this position significantly influences vowel realization, making it more likely for ⟨o⟩ to change to ⟨u⟩ in these contexts, albeit only slightly (⟨1%). Since ⟨u⟩ is the only alternative to ⟨o⟩ in the dataset, this corresponds to a slightly increased likelihood of ⟨u⟩ appearing in root-final position.

Figure 11 illustrates predicted variation in the written realization of ⟨o⟩ in Media Lengua words. Words displayed represent approximately 6% of all ⟨o⟩ tokens (40 of 701). This subset includes only Spanish origin lexemes, as Unified Kichwa does not contain the vowel ⟨o⟩ as either a phoneme (/o/) or as part of its standard orthography. Therefore, the words plotted span from non-standard, Kichwa-influenced ⟨u⟩ (left), to standard, Spanish-influenced ⟨o⟩ (right).

Notable words consistently spelled with ⟨u⟩, rather than ⟨o⟩, include *potrero* 'pasture', *plastico* 'plastic' with *blanco* 'white' and the verb *llovena* 'rain' (from Spanish *potrero, plástico, blanco,* and *llover,* respectively). The word *potrero* contains two instances of ⟨o⟩, which are both consistently realized as ⟨u⟩; additionally, the ⟨re⟩ is elided in every instance, showing that in this dataset, the word is stably spelled as *putru* (incidentally, similar to the word *potro* 'colt' in Spanish).

There are several high frequency and mid frequency words that consistently varied between ⟨o⟩ and ⟨u⟩; most notably, *ahora* 'now/ today', *ahorita* 'right now', and the verb *morina* 'die'. This trend prominently appears with words that contain

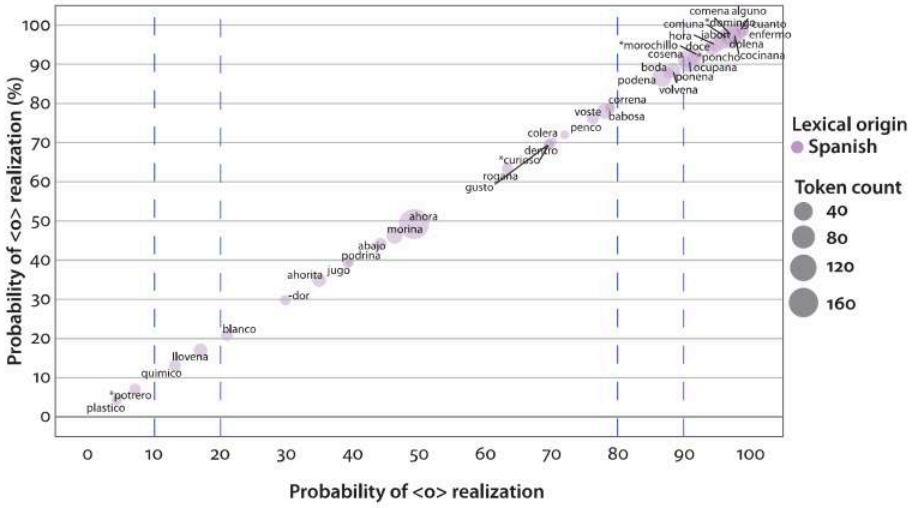


Figure 11. Bubble plot illustrating ⟨o⟩ variation in Media Lengua showing the percentage similarity to Standard Spanish. Each bubble represents a word, sized according to its frequency in the dataset. No Kichwa origin words appear in this figure as ⟨o⟩ is not present in the Unified Kichwa writing system or phoneme inventory

⟨o⟩ at the end of the root as the model suggests; *abajo* 'below', *curioso* 'curious', and *dentro* 'inside'. Words with substantial root-internal ⟨o/u⟩ variation include, *podrina* 'rot', *rogana* 'beg', *colera* 'angry', *babosa* 'slime', *correna* 'run', and the Spanish origin agentive morpheme *-dor*.

There is a notable cluster of words above 80% that were mostly spelled with ⟨o⟩ conforming to Standard Spanish expectations. These include *ponena* 'put', *comena* 'eat', *podena* 'can', *comuna* 'community', *cocinana* 'cook', *dolena* 'hurt', *volvena* 'return', *hora* 'hour', *ocupana* 'occupy', *jabon* 'soap', *boda* 'type of soup'. Finally, there were two words that contained two instances of ⟨o⟩. These included, *domingo* 'Sunday' (raw counts show ⟨o⟩ in medial position 100% of the time and ⟨o⟩ in final position 77% of the time) and *poncho* 'poncho' (raw counts show ⟨o⟩ in medial position 88% of the time and ⟨o⟩ in final position 75% of the time). Both trends suggest yet again that root-final position is more variable than root-internal.

4.2.5 ⟨*ie*⟩, ⟨*ia*⟩, *and* ⟨*ea*⟩ *variation*

This section examines the probability of vowel sequences ⟨ia⟩, ⟨ea⟩, and ⟨ie⟩ realized as such versus other vowel sequences. While each vowel sequence was modeled separately, variation was limited, therefore, all sequences appear in Figure 12.

The intercepts represent the likelihood of the vowel sequence being realized as itself under baseline conditions.

For the ⟨ie⟩ sequence, no fixed predictors improved model fit, and the final model therefore includes only an intercept, with *Word* modelled as a random effect to account for item-specific deviations. The intercept value (4.64; Table 7) represents the log-odds of the vowel sequence ⟨ie⟩ being realized as ⟨ie⟩ under baseline conditions. This strongly positive intercept indicates that the sequence ⟨ie⟩ is highly stable in the dataset, with little evidence of systematic variation beyond word-level differences.

Table 7. Model summary of ⟨ie⟩ variation, showing the effect of the intercept on the likelihood of ⟨ie⟩ being realized as ⟨ie⟩ versus another vowel sequence

⟨ie⟩	Estimate	Std. Error	z value	Pr(>|z|)
Intercept: Same/ Different	4.64	1.15	4.05	5.22e-05

For the ⟨ia⟩ sequence, no fixed effects were retained in the final model, and *Word* was included as a random effect to account for item-specific deviations. The model output (Table 8) shows a large positive intercept (9.11), representing the log-odds of the sequence ⟨ia⟩ being realized as ⟨ia⟩ across the dataset. This indicates a strong overall tendency for ⟨ia⟩ to remain stable in writing, with little evidence of systematic variation beyond word-specific differences.

Table 8. Model summary of ⟨ia⟩ variation, showing the effect of the intercept on the likelihood of ⟨ia⟩ being realized as ⟨ia⟩ versus another vowel sequence

⟨ia⟩	Estimate	Std. Error	z value	Pr(>|z|)
Intercept: Same/ Different	9.11	2.33	3.91	9.24e-05

For the ⟨ea⟩ sequence, no fixed effects were retained in the final model, with *Word* included as a random effect to account for item-specific variation. The model intercept (Table 9) is negative (−6.70 log-odds), suggesting a low overall likelihood of ⟨ea⟩ being realized as ⟨ea⟩ in the dataset; however, this effect did not reach conventional levels of statistical significance ($p = 0.052$). This indicates that while ⟨ea⟩ is rare in writing, the limited number of tokens prevents a strong statistical conclusion beyond this general tendency.

Table 9. Model summary of ⟨ea⟩ variation, showing the effect of the intercept on the likelihood of ⟨ea⟩ being realized as ⟨ea⟩ versus another vowel sequence

⟨ea⟩	Estimate	Std. Error	z value	Pr(>|z|)
Intercept: Same/ Different	−6.79	3.49	−1.94	0.052

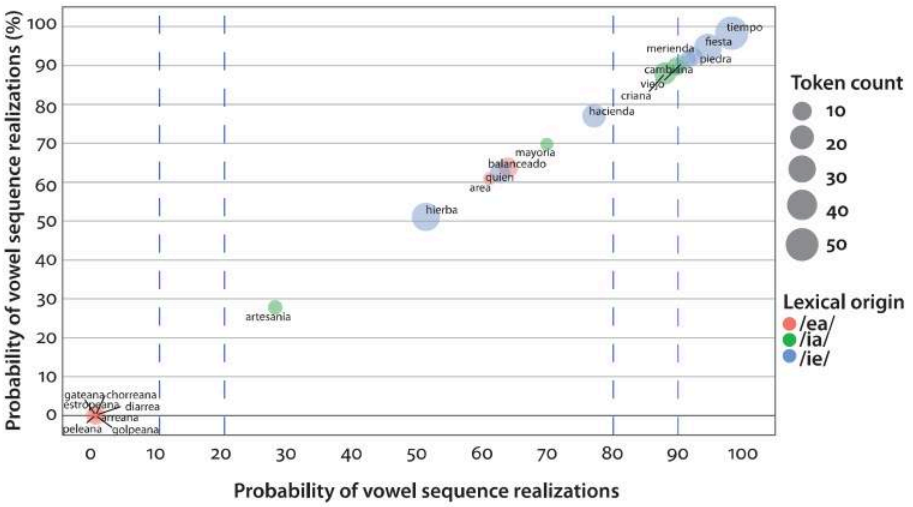


Figure 12. Bubble plot illustrating the variation of vowel sequences ⟨ie⟩, ⟨ia⟩, and ⟨ea⟩ in Media Lengua words of Spanish origin. Each bubble represents a word and is color-coded by vowel sequence (⟨ie⟩ (light blue), ⟨ia⟩ (light green), or ⟨ea⟩ (light pink)) and sized according to its frequency in the dataset. All vowel sequences are derived from a unique model

Figure 12 illustrates predicted vowel sequence variation for words with less than 99% variation and containing at least three tokens in the dataset. These items represent 17% of the total instances of the vowel sequences ⟨ie⟩, ⟨ia⟩, and ⟨ea⟩ (21 tokens of a possible 129). The vowel sequence subset of this analysis is based solely on Spanish origin terms, as no Unified Kichwa words contain ⟨ie⟩ and ⟨ea⟩ vowel sequences in the standard spelling system or as part of the phonemic inventory. The vowel sequence ⟨ia⟩, in Kichwa words like *ufiana* (Unified Kichwa *upyana*) 'drink', was not analyzed due to the lack of Kichwa origin words containing this sequence in Media Lengua. Therefore, the 17 tokens represent a range of 0–99% similarity to Standard Spanish only.

The model analysis suggests that the sequence ⟨ie⟩ is highly stable with just three words of the 61 analyzed showing variability. These include *hierba* 'herb', *quien* 'who', and *hacienda* 'ranch' (from Spanish *hierba, quién*, & *hacienda*, respectively), which appear with variable monophthongs such as *guerba, guirba,* and *hirba* for *hierba*; *quen* and *quin* for *quien*; and *asinda* and *azenda* for *hacienda*. Other words that contain ⟨ie⟩ show less variability. These include *merienda* 'dinner', *fiesta* 'party', *tiempo* 'time', *piedra* 'rock', and *viejo* 'old'.

The ⟨ia⟩ sequence is also stable, with only minor variation observed in, *mayoria* 'majority' (from Spanish *mayoría*) with yod-insertion in some tokens written as *mayoriya*. The other two words, *criana* 'breed' and *cambiana* 'change' (from

Spanish *criar* & *cambiar*, respectively) also show variation appearing as *creana* once for *criana;* and *cambina* and *cambihana* once each for *cambiana.*

The sequence ⟨ea⟩, while less common, and not having reached significance, appears stable as ⟨ia⟩ in *golpeana* 'hit', *gateana* 'crawl', and *arreana* 'herd' (from Spanish *golpear, gatear,* & *arrear*, respectively) but variable in both *balanceado* 'animal feed' and *area* 'area' with the former appearing as *balanciado* in two instances and the latter appearing as *aria* in one instance. Words with two or less instances included *batea* 'wood reciprocal', spelled as *batia, chorreana* 'spill', spelled as *chorriana*, *diarrea* 'diarrhea', spelled as *diarrheya*, *estropeana* 'destroy', spelled as *estropiana, menear* 'stir', spelled as *meniana*, and *retaceana* 'cut into pieces' as *rretaciana*, suggesting that both annotators overwhelmingly avoided ⟨ea⟩ in preference of ⟨ia⟩.

4.2.6 *Other vowel sequences*

Other vowel sequences were not analyzed inferentially as there were insufficient data to provide meaningful results or because variation was minimal. For instance, ⟨ai|ay⟩, ⟨au|aw⟩, ⟨oi⟩ showed no variation from Standard Spanish or Unified Kichwa. Of the 73 words containing ⟨io⟩ in this dataset, only *mio* 'my' (from Spanish *mío*) showed variation with ten of a total of 83, containing yod-insertion; two of these spelled as *meyo* and eight spelled as *miyo*. The sequence ⟨ue⟩ appeared in 52 words with variation occurring in two of eight instances of *almuerzo* 'lunch' (from Spanish *almuerzo*) spelled as *almorzo* and *almurzo*. Variation also occurred in the pronoun *nuestro* 'we/our' (from Spanish *nuestro*) appearing as *noestro* out of seven total instances, and once in the word *dueño* 'owner' (from Spanish *dueño*) spelled as *duino*.

Of the 70 words containing ⟨ua|wa⟩, only *cuarenta* 'forty' (Spanish origin) and *kinwa* 'quinoa' (Kichwa origin) showed variation. For *cuarenta* the same variation was observed in two instances of eight total tokens, appearing as *curenta*. For *kinwa,* ⟨u⟩ was inserted after the ⟨n⟩ in two of four total tokens, appearing as *quinuwa*; the other two instances reflected the pronunciation of the standard (⟨*quinua*⟩ /kinwa/).

The sequence ⟨eo⟩, with a limited number of tokens (30) showed some variation in the word *peon* 'pion/worker' (from Spanish *peón*) with one of two instances spelled as *peun*. In every instance (4 analyzed) of the word *acordeon* 'accordion' (from Spanish *acordeón*) appeared as *accordion*. Additionally, every instance of the univerbation *abeoras* 'when/ when there was/ once upon a time' (*abe* from Spanish *haber* 'have' + *oras* 'hours' from Spanish *horas*) appeared as *abeuras*.

Similarly, for ⟨eu⟩, the univerbation *deunaves* 'right away' (*de* from Spanish *de* 'of' + *una* from Spanish *una* 'one' + *ves* from Spanish *vez* 'time'), appeared

as *deonaves* (twice), *dionaves* (4 times), *diunaves* (9 times) out of 25 instances. The only other ⟨eu⟩ variation occurred in *reumatismo* 'rheumatism' (from Spanish *reumatismo*), spelled once as *rioma* (a reduced form). For ⟨iu⟩, only the spelling of the *Chibuleo* nationality appears as a variant of Unified Kichwa *Chiwuliyu*; however, *Chibuleo* is the official Standard Spanish spelling, and it is not of Kichwa origin. Finally, ⟨uo⟩ shows variation in the word *cuota* 'quota' (from Spanish *cuota*) occurring as *cota* in 12 of 14 instances.

5. Discussion

The results of this study provide new insights into the Media Lengua vowel system through an analysis of written vowel usage in transcribed conversations. Using random effects produced from generalized linear mixed-effects models (GLMMs), the stability and variability of written vowels was quantified, identifying degrees of difference from Media Lengua's standard source language orthographies. Overall, the results show a strong tendency for Media Lengua to retain source vowels from both Standard Spanish and Unified Kichwa, though adherence to these source orthographies is not strict. Instead, patterns of vowel variation and stable use of non-standard vowels emerge, particularly in root-final position, place names, and cases of "reborrowings" (see § 5.2), suggesting a nascent orthographic system that diverges in systematic ways from both source languages.

This lack of strict adherence is unsurprising given the sociolinguistic context in which Media Lengua is written. Unlike other contact languages (e.g., Michif and Haitian Creole), where early stages of orthographic development were imposed by external standards, which often glossed over the contact language's phonology, Media Lengua has no record of such prescriptivism. Therefore, while the annotators obviously have Standard Spanish, and to some degree Unified Kichwa, at their disposal, there was no pressure to adhere to these systems. This allowed them to make relatively unconstrained choices that reflected how they perceive their language.

Section 5.1 examines cases of source language-oriented vowel stability, where written vowels consistently align with Standard Spanish or Unified Kichwa expectations. Section 5.2 turns to vowel variability and non-standard vowel stability, organizing the discussion around recurring lexical and morphological patterns. These include root-final effects, reborrowings and place names, and Kichwa origin words, each of which may show either stable non-standard vowels or variability. Section 5.3 considers the broader implications of these patterns for Media Lengua and Unified Kichwa orthography. Finally, Section 5.4 outlines methodological and data-related limitations and identifies directions for future research.

5.1 Source language-oriented vowel stability

A key finding in this study is the overall stability of vowel usage, with most words written by the two transcribers in Media Lengua retaining the vowels prescribed in Standard Spanish or Unified Kichwa. For instance, Media Lengua ⟨e⟩ and ⟨o⟩ largely follow Standard Spanish patterns, while high vowels, ⟨i⟩ and ⟨u⟩, show consistent alignment with Standard Spanish in Spanish origin words and with Unified Kichwa in Kichwa origin words and bound morphemes. This pattern extends to vowel sequences where ⟨ie⟩ and ⟨ia⟩ show even greater conservation with Standard Spanish. The one exception is ⟨ea⟩, which nearly always defaulted to ⟨ia⟩. However, this deviation is easily explained as many rural varieties of Spanish, even beyond Ecuador, frequently use /ia/ rather than the standard /ea/ (see e.g., Floyd 1995: 154).

Of the 1641 unique items analyzed with at least three repeats in the dataset, 170 (10%) either exhibited variable vowel usage under 99% or consistent non-standard vowel usage. Limiting the analysis further to words with stable non-standard vowel spellings and words with substantial vowel variation (≤80%), reduces this number to 105 words (6%). This suggests that relatively few vowels written by the transcribers deviated substantially from Standard Spanish or Unified Kichwa norms. These findings help answer two questions posed in § 1: *Does acoustic overlap between mid and high vowels influence spelling* (resulting in written vowel choices to be variable)? And, *Does the substrate phonology (Kichwa) have greater influence than the lexifier (Spanish) in shaping vowel spelling* (resulting in a preference of high vowels over mid vowels)? The answer to both appears to be no, with some exceptions detailed in Section 5.2.

It could be assumed that Spanish education contributes to the conservation of Spanish origin vowels in Media Lengua. However, this assumption would also imply that Spanish education stabilizes consonant usage, which is unlikely. As detailed in Section 2.3, consonant heterographs (two or more graphemes that represent a single phoneme) in the Media Lengua dictionary showed substantially more written variation than mid and high vowels (22% vs. 12% respectively).

When frequency is taken into account, the present results suggest that vowel variation is even more constrained than the dictionary figures alone would imply (the abovementioned 10% or 6% for highly variable and stable non-standard vowels) whereas consonant variation would be expected to remain comparatively high (extrapolated to be approximately 18%[7]). Furthermore, the transcribers consistently used corresponding Standard Spanish or Unified Kichwa graphones.

7. ((22% [for dictionary consonants] * 10% [vowels in this study]) / 12% [for dictionary vowels])

This suggests that the transcribers were more conservative with vowel usage than with consonant heterographs (e.g., ⟨b-⟩ vs. ⟨v-⟩ for /b/) but showed greater variation than with consonant graphones. This may imply that vowels occupy an intermediate status in the transcribers' orthographic system; they are not fully neutralized, but neither are they treated as maximally contrastive, resulting in quantifiable degrees of variation.

This orthographic behavior aligns with the perceptual patterns described in §2.4.2. As shown by Stewart (2018b), Media Lengua listeners reliably distinguish mid and high vowels at rates around 90%, though not at the ceiling levels observed for native Spanish listeners (~99%). These results reinforce the same idea reached from the orthographic data; mid and high vowel contrasts are perceptually robust enough to be maintained, yet weak enough that some degree of variation occurs, though without impeding comprehension. This convergence between perceptual and orthographic evidence is consistent with earlier proposals (Stewart 2015b, 2018b, 2018a, accepted) that these contrasts carry a low functional load in Media Lengua.

Regarding the possible effects of overlapping categories observed in the F1xF2 dimension, as production results suggest (see §2.4.1), it appears that the ability to perceive categorical differences mitigates the influence of overlapping formant frequencies on written vowel choice. This observation aligns with the work of Onosson and Stewart (2021a), which demonstrates that vowel identification in Media Lengua is shaped by multiple factors, both social and linguistic, that go beyond formant frequencies. These observations, again, directly respond to the second research question posed in §1: *Does acoustic overlap between mid and high vowels influence spelling* (resulting in written vowel choices to be variable)? Based on the results, the answer appears to be no, at least not to the same degree; perception and structural/contextual cues play a stronger role than acoustic proximity in shaping orthographic choices.

5.2 Vowel variability and non-standard vowel stability

Of the 6% of words that show strong vowel variability (≤80%) or non-standard vowel stability, most can be grouped into two distinct categories: (1) vowels appearing in root-final position, and (2) "reborrowings".[8] The use of non-standard vowels is particularly prominent in root-final position where Spanish mid vowels often shift to high vowels (*querena* → *querina* from Spanish *querer* 'want'), while

8. "Reborrowings" refer to Kichwa words borrowed into Spanish and later reintroduced into Media Lengua, or Spanish words borrowed into Kichwa and subsequently adopted back into Media Lengua. This category also includes place names.

the reverse pattern is comparatively rare (e.g., *salina* → *salena* from Spanish *salir* 'leave').

This asymmetry is clearly reflected in the statistical models. Root-final position shows a strong and significant effect for ⟨e⟩ and ⟨o⟩, a much weaker effect for ⟨i⟩, and no detectable effect for ⟨u⟩. Taken together, these results indicate that orthographic variability is not driven by symmetric alternation between mid and high vowels. Instead, they point to a directional tendency in which Standard Spanish mid vowels (⟨e⟩ and ⟨o⟩) are more likely to shift toward high vowels (⟨i⟩ and ⟨u⟩) in root-final position, rather than the reverse.

These results provide a clear answer to the fourth research question posed in §1: *Do contextual cues inform orthographic practices in Media Lengua?* Root-final position, a morphologically defined context, emerges as a strong predictor of orthographic behavior for mid vowels, while having a limited impact on high vowels. This demonstrates that orthographic practices in Media Lengua are informed by contextual cues such as morphological position and lexical origin, rather than by unconstrained vowel alternation.

This pattern is most evident in verbs, which undergo regular morphological restructuring in the transition from Spanish to Media Lengua. Specifically, the Spanish infinitive suffix *-r* is replaced by the Kichwa infinitive marker *-na*, leaving the pre-*r* vowel from Spanish in root-final position at a morphological boundary. However, root-final vowel variation is not limited to verbs. They also occur in non-verbal items such as *ondi-pi* 'where-LOC' (from Spanish *dónde* 'where') and *parti-man* 'place-DIR' (from Spanish *parte* 'place').

Table 10 provides a list of words identified as having root-final stable vowel usage that consistently deviates from each source language's standard orthography. Each table in this section contains spelling variations from the dataset, beginning with the most common and progressing to the least common (*Variations*). Additionally, the table includes the word's origin (*Or*), the vowel used in the source language's standard orthography (*St*), the Media Lengua vowels used in the dataset (*ML V*), the actual variability from the dataset (*Actual %*), the predicted variability from the models (*Pred %*), and the number of tokens analyzed for each word (*Count*). Words marked with an asterisk (*) contain two or more instances of the vowel under analysis. For these cases, the predicted variability represents an average of the vowels, while the actual variability is broken down by position: (RF = root-final, RI = root-internal, I = word-initial).

Table 10. Stable usage of root-final vowels that differ from the source language's standard orthography (≤20%) with at least three instances in the dataset

Non-standard Vowel-stable — Root-final								
Word	**Variations**	**Gloss**	**Cognate***	**Or**	**St**	**ML V**	**Actual %**	**Pred %**	**Count**
llovena	*lluvina, llovina*	'rain'	*llover*	Sp	e	i	0%	5.0%	13
rompena	*rompina, ronpina*	'break'	*romper*	Sp	e	i	0%	19%	3
balde	*baldi*	'bucket'	*balde*	Sp	e	i	0%	19%	3
habena	*abina, habina, avina, abena, habena, avena*	'have'	*haber*	Sp	e	i, e	14%	12%	122
comena	*comina, comena, cumina*	'eat'	*comer*	Sp	e	i, e	13%	14%	68
cogena	*cogina, cojina, cogena*	'take'	*coger*	Sp	e	i, e	12%	14%	65
nique	*niqui, nique*	'nothing'	*ni que*	Sp	e	i, e	14%	14%	7
plastico	*plasticu, plastiku*	'plastic'	*plástico*	Sp	o	u	0%	4%	4
gusto	*gushtu, gushto, gusto*	'liking'	*gusto*	Sp	o	u,o	50%	70%	4
arreana	*arriana*	'herd'	*arrear*	Sp	ea	ia	0%	0%	3
golpeana	*golpiana*	'hit'	*golpear*	Sp	ea	ia	0%	0%	3
gateana	*gatiana*	'crawl'	*gatear*	Sp	ea	ia	0%	0%	3
**potro (potrero)*	*putru*	'pasture'	*patrero*	Sp	o	u	0% (RF)	7%	6
					o	u	0% (RI)		
**verdena*	*verdina*	'make green'	*(hacer) verde*	Sp	e	e	100% (RI)	67% (avg)	6
					e	i	0% (RF)		
**recogena*	*rrecogina, recogina*	'collect'	*recoger*	Sp	e	e	100% (RI)	50% (avg)	8
					e	i	0% (RF)		

* We use the term 'cognate' to emphasize that Media Lengua lexical items share historical origin with their Spanish and/or Kichwa sources, rather than resulting from synchronic borrowing processes. As Media Lengua is considered a separate language from both Spanish and Kichwa, 'cognate' better reflects etymological relationships without implying ongoing loanword incorporation.

Table 11 lists words with root-final variable vowel usage, displaying variability between approximately 20% and 80%. The word with the highest count was *decina* 'say', with 225 instances and a variability of nearly 50%, indicating that the transcribers showed no strong preference between using ⟨i⟩ or ⟨e⟩ for the root-final vowel. In contrast, the root-internal Spanish origin ⟨e⟩ exhibited minimal variation, with a strong preference for ⟨e⟩ in 95% of instances, despite the dataset containing at least 11 identified spelling variations for this word.

Table 11. Variable usage of root-final vowels (approximately between 20% and 80%) with at least three instances in the dataset

Vowel-variable — Root-final vowel								
Word	**Variations**	**Gloss**	**Cognate**	**Or**	**St**	**ML V**	**Actual %**	**Pred %**	**Count**
ponena	*ponina, ponena, punina*	'put'	*poner*	Sp	e	i, e	23%	21%	75
sabena	*sabina, sabena, savina*	'know'	*saber*	Sp	e	i, e	27%	29%	49
vena	*vina, vena*	'see'	*ver*	Sp	e	i, e	37%	31%	41
molena	*molina, molena*	'grind'	*moler*	Sp	e	i, e	43%	48%	7
bloque	*bloque, bloqui*	'cement block'	*bloque*	Sp	e	e, i	50%	32%	4
dulce	*dulce, dulci, dulse*	'sweet'	*dulce*	Sp	e	e, i	67%	67%	3
cosena	*cosena, cotsena, cosina, cusina*	'sew'	*coser*	Sp	e	e, i	67%	72%	9
podena	*podena, podina, pudina*	'can'	*poder*	Sp	e	e, i	57%	58%	44
hacena	*asena, asina, atsena, atsina, hacena, hasena, hasina, azina, hazina*	'do/make'	*hacer*	Sp	e	e, i	59%	61%	149
valena	*valena, valina, balena*	'have value'	*valer*	Sp	e	e, i	55%	61%	11
que	*que, qui, ki*	'what/that'	*que*	Sp	e	e, i	76%	96%	152
conocena	*conosena, conocena, conosina*	'know'	*conocer*	Sp	e	e, i	64%	67%	11
volvena	*bolvena, bolvina, volvena, volvina, vulvina, bolbena, volbena, volmuna*	'return'	*volver*	Sp	e	e, i	53%	59%	15
resentina	*resentena, resentina*	'resent'	*resentir*	Sp	i	e, i	33%	42%	3
compartina	*compartena, comparstena, conparstina*	'share'	*compartir*	Sp	i	e, i	33%	42%	3
salina	*salina, salena*	'leave'	*salir*	Sp	i	i, e	62%	61%	61

Table 11. *(continued)*

Vowel-variable — Root-final vowel									
Word	**Variations**	**Gloss**	**Cognate**	**Or**	**St**	**ML V**	**Actual %**	**Pred %**	**Count**
decina	*desina, desena, detsina, detsena, tsena, tsina, dicina, detzina, decena, ditsina, ditsina*	'say'	*decir*	Sp	i	i, e	51%	52%	225
blanco	*blanku, blancu, blanco*	'white'	*blanco*	Sp	o	u, o	29%	21%	7
curioso	*curioso, curiusu*	'curious'	*curioso*	Sp	o	u, o	50%	70%	4
abajo	*abajo, abaju*	'below'	*abajo*	Sp	o	o, u	57%	44%	7
dentro	*dentro, dentru*	'in/inside'	*dentro*	Sp	o	o, u	66%	70%	3
area	*area, aria*	'area'	*area*	Sp	ea	ea, ia	67%	61%	3
mayoria	*mayoria, mayoriya*	'majority'	*mayoría*	Sp	ia	ia, iya	67%	70%	3
artesania	*artesania, artesaniya*	'handicraft'	*artesanía*	Sp	ia	ia, iya	25%	28%	4
cambiana	*cambiana, cambihana*	'change'	*cambiar*	Sp	ia	ia, iha	89%	90%	9
criana	*criana, creana*	'grow/ raise'	*criar*	Sp	ia	ia, ea	87.5%	79.2%	16
**tejena*	*tejena, tejina, tijina*	'knit'	*tejer*	Sp	e	e, i	80% (RI)	97% (avg)	10
					e	e, i	60% (RF)		
**entendena*	*entendena, entendina*	'understand'	*entender*	Sp	e	e	100% (I)	70% (avg)	15
					e	e	100% (RI)		
					e	e, i	40% (RF)		
**querena*	*querena, querina, quirina, quirena, kirina, kerina*	'want'	*querer*	Sp	e	e, i	72% (RI)	32% (avg)	70
					e	e, i	50% (RF)		

Another clear trend in the data shows that Kichwa origin words borrowed into Spanish (including place names and words of Kara origin) where ⟨i⟩ and ⟨u⟩ shifted to ⟨e⟩ and ⟨o⟩, respectively, tend to retain the ⟨e⟩ and ⟨o⟩ in Media Lengua.

Table 12 provides both non-standard vowel-stable and vowel-variable words that fit this pattern. For instance, *kucha* /kutʃa/ 'lake' is found in numerous place names in and around Imbabura and throughout the Ecuadorian Andes, such as *Cuicocha* 'Guinea Pig Lake', *Yahuarcocha* 'blood lake', the lakes of Mojanda, which are known locally as *Caricocha* 'Lake of the Man', *Huarmicocha* 'Lake of the Woman', *Yanacocha* 'Black Lake', and the largest lake in Imbabura, Lago San Pablo, known locally as *Imbacocha* 'Lake of Imba', among many other examples. In our dataset, *kucha* never appears with a ⟨u⟩ (or a ⟨k⟩), and in all seven instances, it is spelled as *cocha*.

Other words with non-standard stable vowel usage and/or variable vowel usage include, the community of *Waykupungu*, Spanish *Huaycopungo*, which retains the Spanish spelling in all instances in the dataset; Kichwa *sampu* 'fig-leaf gourd', Spanish *sambo,* appearing with ⟨o⟩ roughly 90% of the time (e.g., *sambo/tsambo*); Kichwa *chukllu* 'corn', Spanish *choclo,* appearing with ⟨o⟩ approximately 30% of the time (e.g., *choklo, chokllo, chuglu, chugllu* or *choclo*); and Kichwa *purutu* 'bean', Spanish *poroto,* which appears with ⟨o⟩ in nearly every instance (e.g., *poroto*). As mentioned in Section 4.2.1, Kichwa *kuyi* 'guinea pig', Spanish *cuy,* shows a strong preference (78%) from both transcribers to use non-Unified Kichwa ⟨e⟩ (*cuye* /kuje/) even though Spanish drops the final syllable in favor of an open vowel sequence. Media Lengua may do this to either emphasize the existence of the second syllable, disassimilating /j/ from /i/ in the Unified Kichwa pronunciation (/kuji/) or distinguishing it from the Spanish monosyllabic pronunciation (/kui/).

Additional words with variable ⟨u/o⟩ usage include Kichwa *maytuna* 'wrap', *kukawi* 'snack', and *quipi* 'bundle'; all three words used in rural Ecuadorian Spanish. In its nominal form, *mayto* refers to a "set of diapers used by mothers in rural areas to dress newborn babies" (INPC Ecuador 2012: 178). *Cocabi* colloquially refers to shared snack food (it also contains the root *coca* from the coca plant and is found in Spanish as *cucayo* (Real Academia Española, 2025)), and *quipe* refers to a knapsack (HarperCollins 2024). A more commonly used Kichwa word found in Ecuadorian Spanish is Kichwa origin *wachu* 'furrow', spelled as *huacho* in Spanish. Similarly, place names of pre-Kichwa origin from the Kara language that appear with mid vowels in Spanish are often retained in Media Lengua, even though Unified Kichwa prefers high vowels. Examples include *Otavalo* (city name), *Peguche* (community name), and *Cayambe* (city name). This trend is also prevalent in non-standardized Kichwa.

Interestingly, only three Spanish origin words and the one Spanish bound morpheme in the dataset fall into this list, *bestia* 'animal', *colera* 'angry', *hacienda* 'ranch', and the agentive marker *-dor.* Each of these items are older borrowings into Kichwa, with *bestia* referring to 'horse', *colera* used in its archaic sense of

'angry', *hacienda* originating during the colonial hacienda system, and *-dor* used alongside the Kichwa agentive *-k* in non-Unified Kichwa (e.g., Unified Kichwa *hampi**k***, Community Kichwa *jambi**dur*** for 'healer'). These items fall into the vowel-variable category, with *bestia* appearing as both *bestia* and *bistia, colera* as *culera* and *culira* (the latter only attested in the Media Lengua Collection), *hacienda* with numerous spellings (see Table 12), and *-dor* also appearing as *-dur*.

A notable older Spanish borrowing into Kichwa is *intintina* 'understand', from Spanish *entender*. As shown in Figure 9, most instances were written with ⟨e⟩ in both initial- and root-internal positions, while ⟨i⟩ appeared in root-final position in approximately 40% of the tokens, placing it in the previous category (see Table 11).

Table 12. Non-standard vowel stable (=⟨20%) and vowel variable (between ~20% and ~80%) usage of vowels in reborrowings with at least three instances in the dataset

Non-standard Vowel-stable — Reborrowings									
Word	**Variations**	**Gloss**	**Cognate**	**Or**	**St V**	**ML V**	**Actual %**	**Pred %**	**Count**
sambu	*tsambo, sambo, sambu*	' fig-leaf gourd'	*sambu*	K	u	o, u	9%	10%	23
kucha	*cocha*	'lake'	*lago, cocha*	K	u	o	0%	0%	7
Kayampi	*Cayambe, Cayambi*	'Cayambe'	*Cayambe, Kayampi*	Ka	i	e, i	15%	18%	13
Kutakachi	*Cotacachi*	'Cotacachi'	*Cotacachi*	K	u	o	0%	5%	8
**Waykupunku*	*Huaycopungo*	'Huaycopungo'	*Huaycopungo, Waykupunku*	K	u	o	0% (RI$_1$)	38% (avg)	4
						u	100% (RI$_2$)		
						o	0% (RF)		
**purutu*	*poroto, porotu*	'bean'	*poroto*	K	u	o	0% (RI$_1$)	4% (avg)	33
					u	o	0% (RI$_2$)		
					u	o,u	5% (RF)		
**Utawalu*	*Otavalo, Otavalu*	'Otavalo'	*Otavalo*	Ka	u	o	0% (I)	11% (avg)	12
						o, u	16% (RF)		
Vowel-Variable — Reborrowings									
bestia	*bestia, bistia*	'animal'	*bestia*	Sp	e	e, i	75%	86%	4
colera	*colera, culera*	'angry'	*colera*	Sp	o	o, u	67%	72%	3

Table 12. *(continued)*

hacienda	*asienda, acienda, asinda, atsenda, hacienda, hasienda, hatsenda*	'ranch'	*hacienda*	Sp	ie	ie, e, i	74%	77%	19
-dor	*-dur, -dor*	'agentive marker'	*-dor*	Sp	o	u, o	25%	30%	4
wachu	*huacho, wachu, huachu*	'furrow'	*wachu*	K	u	o, u	21%	49%	14
maytuna	*maytuna, maitona*	'wrap'	*maytuna*	K	u	u, o	50%	73%	4
kukawi	*cukabi, kucabi, cocabi*	'snack'	*kukawi*	K	u	u, o	33%	70%	6
kuyi	*cuye, cuyi*	'guinea pig'	*kuyi*	K	i	e, i	22%	23%	9
**quipi*	*quipi, quipe*	'knapsack'	*kipi*	K	i	i	100% (RI)	78% (avg)	8
					i	i, e	25% (RF)		

Another interesting trend is observed in numerous Kichwa origin words and morphemes not borrowed into Ecuadorian Spanish, which nonetheless appear with ⟨e⟩ or ⟨o⟩. The motivation for this pattern is not immediately clear. One plausible explanation lies in item-specific phonetic tendencies shaped by Spanish bilingualism and literacy. These tendencies may be reinforced by the transcribers' limited familiarity with Unified Kichwa orthographic norms, which explicitly restrict the use of mid vowel graphemes. Given the broad acoustic range of Kichwa high vowels and their overlap with Spanish mid vowels, ⟨e⟩ and ⟨o⟩ might be favored in cases where Kichwa origin high vowels are consistently realized lower in acoustic space in particular lexical items or bound morphemes.

Another, more controversial, possibility is that Kichwa may, in practice, exhibit a richer vowel system than the traditional three-vowel analysis suggests. Mid vowels have been shown to be marginally contrastive in Imbabura Kichwa in Spanish borrowings (Stewart 2018), and such borrowings constitute a substantial portion of Kichwa's non-standardized lexicon. From a broader phonological perspective, segments need not show robust minimal-pair contrasts in all environments to be treated as phonemic. Rather, they may function as 'marginal phonemes' (segments that are only weakly or infrequently contrastive but nonetheless patterned in the grammar; see Hall 2013). From this perspective, the consistent item-specific use of ⟨e⟩ and ⟨o⟩ in Kichwa origin words and bound morphemes may reflect the marginal phonemic status of mid vowels, rather than instability. However, further phonetic and phonological research on Imbabura

Kichwa would be necessary to evaluate whether mid vowels function as marginal phonemes.

Table 13. Non-standard vowel stable (=⟨20%) and vowel variable (between ~20% and ~80%) usage of vowels in Kichwa origin words (not reborrowings) with at least three instances in the dataset

Non-standard Vowel-stable — Kichwa origin									
Word	**Variations**	**Gloss**	**Cognate**	**Or**	**St V**	**ML V**	**Actual %**	**Pred %**	**Count**
kusa	*cosa*	'husband'	*kusa*	K	u	o	0%	1%	4
**utuhurina*	*otojorina*	'make holes'	*utuhurina*	K	u	o	0% (I)	5% (avg)	6
					u	o	0% (RI_1)		
					u	o	0% (RI_2)		
Vowel-variable — Kichwa origin									
-nti	*-ndi, -nde*	'comitative marker'	*-nti*	K	i	i, e	57%	59%	14
-ri	*-ri, -re*	'reflexive marker'	*-ri*	K	i	i, e	80%	80%	341
-ni	*-ni, -ne*	'1S marker'	*-ni*	K	i	i, e	84%	84%	179
-mu	*-mu, -mo*	'translocative marker'	*-mu*	K	u	u, o	56%	60%	27
=chu	*=chu, =cho*	'negative marker'	*=chu*	K	u	u, o	71%	77%	105
=chu	*=chu, =cho*	'polar question marker'	*=chu*	K	u	u, o	61%	69%	127
-ku	*-gu, -go*	'diminutive marker'	*-ku*	K	u	u, o	70%	74%	318
tukuy	*tukuy, tocui, tocuy*	'all/ every'	*tukuy*	K	u	u, o	29%	64%	17
ullka	*ulca, olca*	'ball of wool'	*ulka*	K	u	u, o	75%	84%	4
**pukyu*	*poguio, pucyu*	'slope'	*pukyu*	K	u	o, u	25% (RI)	29% (avg)	8
					u	io, u	25% (RF)		

Next, Table 14 presents 16 Spanish origin words that exhibited root-internal changes not falling into the previously discussed categories. In most cases, vowels trend toward Spanish origin ⟨e/o⟩ shifting to ⟨i/u⟩, as expected under Kichwa influence. However, only one word appears with a stable high vowel, *ll**o**vena* 'rain', which occurs as *ll**u**vena* or *ll**u**vina* in approximately 85% of its 13 instances. Conversely, three words with stable vowel usage demonstrate the opposite pattern, where Spanish origin ⟨i⟩ shifts to ⟨e⟩. These include *c**i**nturas* 'waists', consistently appearing as *s**e**nturas*, *s**e**ntoras*, or *s**e**turas*; ***i**mpedir*, appearing as ***e**mpedena* or ***e**mpedina*, with the initial vowel consistently written as ⟨e⟩ and the root-final vowel varying between ⟨e⟩ and ⟨i⟩ approximately 50% of the time, while the

root-medial vowel (⟨e⟩) remains unchanged; and *discapacitado* 'disabled', which appears as *descapacitado* or *descapasitado*, where ⟨e⟩ consistently replaces the first ⟨i⟩, but the second ⟨i⟩ remains stable.

Several instances of monophthongization are also noted, such as *hierba* 'herb' and *quién* 'who', both appearing with either ⟨i⟩ or ⟨e⟩; likely indicating Kichwa influence. One instance of diphthongization is observed in the pronoun *voste* 'you', an older borrowing (related to the archaic *vosted*), though this is likely a reflex of *vuestro,* also an archaic form of the second person possessive in Latin America (compare with Spanish *nuestro* 'our' which is also used as a subject pronoun in Media Lengua). Finally, *ahora* 'now' could potentially be considered a reborrowing, as the root *hora* 'hour' appears in the Kichwa univerbation *imauras* 'when' (*ima* 'what' + *uras* 'time').

Table 14. Non-standard vowel stable (=⟨20%) and vowel variable (between ~20% and ~80%) usage of vowels in Spanish origin words with at least three instances in the dataset that showed root-internal change

Non-standard Vowel-stable									
Word	**Variations**	**Gloss**	**Cognate**	**Or**	**St**	**ML V**	**Actual %**	**Pred %**	**Count**
llovena	*llovina, lluvina*	'rain'	*llover*	Sp	o	u, o	15%	17%	13
cinturas	*senturas, sentoras, seturas*	'waist'	*cintura*	Sp	i	e	0% (RI$_1$)	i 12%	6
					u	u, o	33% (RI$_2$)	u 17%	
**impedina*	*empedena, empedina*	'impede'	*impedir*	Sp	i	e	0% (I)	i	4
					e	e	0% (RI)	43% (avg)	
					i	e, i	50% (RF)	e 100%	
**discapacitado*	*descapacitado, descapasitado*	'disabled'	*discapacitado*	Sp	i	e	0% (RI$_1$)	57% (avg)	4
					i	i	100% (RI$_2$)		
Vowel Variable									
pedina	*pedina, pidena, pidina*	'ask'	*pedir*	Sp	e	e, i	70%	72%	10
ahorita	*aurita, ahorita, auryta*	'just now'	*ahorita*	Sp	o	u, o	33%	35%	12
ahora	*aura, ahora, aora*	'now'	*ahora*	Sp	o	u, o	49%	49%	171
podrina	*pudrina, podrina*	'rot/ spoil'	*podrir*	Sp	o	u, o	33%	39%	3

Table 14. *(continued)*

morina	*murina, morina, morena*	'die'	*morir*	Sp	o	u, o	45%	46%	22
babosa	*babosa, babusa*	'slime'	*babosa*	Sp	o	o, u	75%	79%	4
correna	*correna, currina*	'run'	*correr*	Sp	o	o, u	75%	79%	4
rogana	*rogana, rrugana*	'beg'	*rogar*	Sp	o	o, u	60%	63%	5
voste	*boste, buste, voste, votes, vueste, vostedes*	'2s'	*usted*	Sp	o	o, u, ue	74%	78%	19
balanceado	*balanceado, balanciadu*	'type of feed'	*balanceado*	Sp	ea	ea, ia	33%	64%	12
hierba	*guerba, hierba, hierva, guierba, guirba, guirya, hirba*	'herb'	*hierba*	Sp	ie	ie, e, i	48%	51%	33
quien	*quien, quin, quen*	'who'	*quién*	Sp	ie	ie, i, e	57%	63%	13

Finally, a small number of lexical items also suggest vowel-to-vowel alignment within words, superficially resembling harmony-like behavior. For example, forms such as *ocopana* 'occupy' (< *ocupar*), *longo* 'youth' (< *lunku*), *aunu* 'not yet' (< *auno*), *otojorina* 'make holes' (< *utuhurina*), *choclo* 'corn' (< *chukllu*), *poroto* 'bean' (< *purutu*) etc. show multiple vowels shifting in the same direction, and others that are resistant to losing alignment. For example, as demonstrated root-final ⟨e⟩ tends to shift to ⟨i⟩, but not in *tenena* and *tejena*. There are other examples of variable vowels that go in and out of alignment e.g., *uno* ~ *unu* and *querena* ~ *querina/quirina*. Given the limited number of items, the presence of counterexamples (e.g., *venina* 'come'), and the likelihood of Spanish orthographic influence, these patterns are best treated as item-specific tendencies rather than evidence for a productive vowel harmony system. Further work with targeted datasets along with phonetic evidence would be required to evaluate this possibility more directly.

5.3 Implications for orthography

The findings also speak to the status of Unified Kichwa orthography as a single written standard for a highly diverse set of Kichwa varieties. In the Media Lengua context, examined here, both writers consistently preserve mid vowels in borrow-

ings and place names and show patterned mid-high alternations at morphological boundaries. These are not random errors, but appear to be stable preferences tied to etymology and morphology, which align with well-established cases of linguistic stratification in other languages.

For example, in Japanese, Sino-Japanese vocabulary observes phonotactic patterns such as permitting coda /ɴ/ that native Yamato words do not, and the standard orthography reflects this distinction through different scripts; native vocabulary, grammatical elements, and verb/adjective endings are typically written in hiragana, whereas foreign loanwords, foreign names, onomatopoeia, and emphatic or stylistic uses are written in katakana. In such a system, source language features are maintained for borrowings not only to index their origin but also to aid recognition and reduce processing cost for literate users.

A similar flexibility in Unified Kichwa could allow mid vowels to signal borrowed vocabulary without undermining the broader three-vowel system. For instance, Spanish-literate speakers of Kichwa/ Media Lengua have entrenched expectations of Spanish spelling for Spanish origin vocabulary. Forms like *kumina* 'eat' can introduce unnecessary processing difficulty where *comena/ comina* 'eat' is immediately transparent as being of Spanish origin (*comer* 'eat'), revealing etymology and reducing cognitive load.

Importantly, not all orthographies require strict prescriptive uniformity. Even English, widely considered to have highly standardized norms, permits acceptable variation in formal writing (e.g., *gray/grey*, *center/centre*, *favor/favour, realize/ realise, burned/burnt*) and widely tolerates informal spellings (*gonna/ going to*, *nite/night, drive-thru/ drive-through*). Such flexibility allows written language to accommodate regional identity, literacy development, and register-specific conventions. Until we better understand the degree of acceptable variation across the broader Media Lengua-speaking population, it would be premature to impose a strict prescriptive orthographic system that may introduce unnecessary complications or limit the natural evolution of written practices.

From this perspective, the current Unified Kichwa standard may be overly restrictive, contributing to challenges in uptake among speakers who are more accustomed to Spanish-based representations. Allowing predictable, etymologically-motivated variation, such as the use of mid vowels in Spanish origin lexemes, may improve accessibility and recognition without undermining overall system coherence, linguistic prestige, and standardization.

In summary, orthographic variation in Media Lengua is informative rather than incidental, revealing how speakers conceptualize their language and which distinctions they treat as meaningful. The patterns observed here reflect an emergent writing system shaped by perception, writer preference, structural and contextual cues, and etymology rather than strict adherence to external norms. While

these trends may eventually inform community-driven discussions about orthography, should such efforts arise, imposing prescriptive intervention at this stage risks constraining natural patterns of usage and literacy preference.

5.4 Limitations

This study has several limitations that should be acknowledged. First, the analysis relies on data from only two transcribers who transcribed different portions of the corpus. While their orthographic practices provide valuable insights, this limits the broader applicability of the findings and may not generalize beyond these individuals.

Second, Spanish dominates literacy in Pijal, and the annotators received their formal schooling in Spanish. Although their spellings appear perceptually grounded in Media Lengua phonology rather than strictly following Standard Spanish conventions (see support for this in §2.3), it is certain that Spanish literacy shapes aspects of their written production. Even with these challenges, the transcribers' consistent use of phonetic spelling in their Spanish translations made this study possible.

It should also be pointed out that although the dataset covers a sizable portion of the Media Lengua Collection, it does not encompass all available recordings, and infrequent lexical items yield limited evidence for assessing stability. Expanding the dataset and increasing the number of transcribers involved would help refine the frequency-based analyses used here. Despite these limitations, the findings represent the first quantitative account of written vowel behavior in Media Lengua and illustrate how speakers negotiate orthographic choices in a nascent spelling system.

Finally, as with many studies based on naturally produced writing in minority languages with limited texts and speakers, the data are susceptible to some non-independence among observations. Orthographic behavior often clusters within lexical items and may also be influenced by discourse context. Because the focus of this study is on item-level spelling patterns rather than speaker- or text-driven effects, *Word* was modelled as a random intercept to account for the most theoretically relevant source of dependence. However, I acknowledge that *Speaker* or *Text* random effects may contribute to residual dependence in the data. These constraints reflect broader practical challenges in documenting emergent orthographic systems with limited corpora. Future work with more detailed metadata and larger speaker samples would allow a fuller assessment of such influences on orthographic variation.

6. Conclusion

This study examined written vowel variability and stability in Media Lengua. The findings demonstrate a strong retention of source vowels from both Standard Spanish and Unified Kichwa. For the words that showed written vowel variation or stable non-standard vowel usage, two main effects were identified (1) a tendency for root-final vowels to shift and (2) vowel shifts in "reborrowings" and place names. At morphological boundaries (root-final position), mid vowels ⟨e⟩ and ⟨o⟩ often shift to high vowels ⟨i⟩ and ⟨u⟩ (e.g., *comi-na* 'eat' from Spanish *comer*), with this pattern being most common in verbs but also present in other parts of speech. "Reborrowings" include Kichwa words borrowed into Spanish and back into Media Lengua, as well as Spanish words borrowed into Kichwa and back into Media Lengua. If ⟨i⟩ and ⟨u⟩ shifted to ⟨e⟩ and ⟨o⟩ in Spanish borrowings (e.g., *Cotacachi* (place name) from Kichwa *Kutakachi*), they were likely retained when borrowed back into Media Lengua, and if Spanish ⟨e⟩ and ⟨o⟩ were borrowed into Kichwa as ⟨i⟩ and ⟨u⟩, they were likely retained when borrowed back into Media Lengua (e.g., *pagui* 'thank you' from Spanish *[Dios se lo] pague* 'God bless you'). Despite the small number of transcribers, this work provides important insights into Media Lengua's written vowel system and informs much of the phonetic work previously conducted on Media Lengua's vowel system.

Abbreviations

TOP	topic marker	ACC	accusative
DIM	diminutive	PST	past
TOT	totality	CONJ	conjunction
1	first person		

References

Baayen, H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge University Press.

Bahr, Howard M. 2004. *The Navajo as Seen by the Franciscans, 1898–1921: A Sourcebook*. Lanham, MD: Scarecrow Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Cerda Ch., Mariano, and Andrés Malaver, D. 2005. *Shimiyuk Kamu: Pastaza markapa ishkay shimi yachayta pushak*. edited by J. Proaño, S. Puyo: DIPEIB-P.

Corominas, J., & Pascual, J.A. (1983). *Diccionario crítico etimológico castellano e hispánico: RI-X: Vol. V*. Gredos.

Corominas, J., & Pascual, J.A. (1984). *Diccionario crítico etimológico castellano e hispánico: G-MA: Vol. III*. Gredos.

Flamand, Rita. 2002. *Michif Conversational Lessons for Beginners*. Winnipeg: Metis Resource Centre.

Floyd, Mary Beth. 1995. "Spanish Dialect Features in Chicano Literature: Short Stories by Alfonso Rodríguez." *Confluencia* 10(2):145–59. https://www.jstor.org/stable/27922293

van Gijn, Rik. 2009. "The Phonology of Mixed Languages." *Journal of Pidgin and Creole Languages* 24:91–117.

Gómez Rendón, Jorge. 2005. "La Media Lengua de Imbabura." Pp. 39–58 in *Encuentros y Conflictos: Bilingüismo y Contacto de Lenguas en el Mundo Andino*, edited by H. Olbertz and P. Muysken. Madrid: Iberoamericana.

Gómez Rendón, Jorge. 2008. *Mestizaje Linguístico En Los Andes: Génesis y Estructura de Una Lengua Mixta*. Quito: Abya-Yala.

Grzech, Karolina, Anna Schwarz, and Georgia Ennis. 2019. "Divided We Stand, Unified We Fall? The Impact of Standardization on Oral Language Varieties: A Case Study of Amazonian Kichwa." *Revista de Llengua i Dret* 71:123–45.

Gualapuro Gualapuro, Santiago David. Forthcoming. "Writing in Minority Languages: Imbabura Kichwa Case Study."

Guion, Susan. 2003. "The Vowel Systems of Quichua-Spanish Bilinguals: Age of Acquisition Effects on the Mutual Influence of the First and Second Languages." *Phonetica* 60:98–128.

Hadley, Sarah. (2026). An Ultrasound Investigation of Articulatory Overlap in Ecuadorian Media Lengua, Kichwa, and Spanish High- and Mid-vowels. MA Thesis. University of Saskatchewan.

Hall, K.C. (2013). A typology of intermediate phonological relationships. *The Linguistic Review*, *30*(2), 215–275.

HarperCollins. 2024. *Quipe*. In *Collins Spanish-English Dictionary*. 10th Edition. HarperCollins Publishers. https://www.collinsdictionary.com/dictionary/spanish-english/quipe

Hutchinson, Gerald M. 1988. "Evans, James." *Dictionary of Canadian Biography 7*.

INPC Ecuador. 2012. *Glosario Del Patrimonio Cultural Inmaterial Del Azuay*. 1st ed. Cuenca: Instituto Nacional de Patrimonio Cultural.

Johnson, K.E. (2008). *Second Language Acquisition of the Spanish Multiple Vibrant Consonant*. University of Arizona.

Lapesa, R. (1981). *Historia de la Lengua Española* (9th ed.). Gredos

Laverdure, Patline, Ida Rose Allard, and John C. Crawford. 1983. *The Michif Dictionary : Turtle Mountain Chippewa Cree*. Winnipeg: Pemmican Publications.

Limerick, Nicholas. 2018. "Kichwa or Quichua? Competing Alphabets, Political Histories, and Complicated Reading in Indigenous Languages." *Comparative Education Review* 62(1):103–24.

Limerick, Nicholas. 2020. "Speaking for a State: Standardized Kichwa Greetings and Conundrums of Commensuration in Intercultural Ecuador." *Signs and Society* 8(2):185–219.

Lipski, John. 2015. "Colliding Vowel Systems in Andean Spanish." *Linguistic Approaches to Bilingualism* 5:91–121.

Lipski, John. 2017. "Ecuadoran Media Lengua: More than a 'Half'-Language?" *International Journal of American Linguistics* 83(2):233–62.

Lipski, John. 2020. "Can a Bilingual Lexicon Be Sustained by Phonotactics Alone? Evidence from Ecuadoran Quichua and Media Lengua." *The Mental Lexicon* 15(2):330–65.

de la Mahautiére, Duvivier. 1757. *Lisette Quitté La Plaine*.

Ministerio de Educación. 2009. *Runakay Kamukuna: Yachakukkunapa Shimiyuk Kamu*. Ecuador: Ministerio de Educación.

Mitchell, Mary. 1988. *Introductory Ojibwe (Severn Dialect), Part One*. edited by R. Valentine and L. Valentine. Thunderbay: Native Language Office, Lakehead University.

Muysken, Pieter. 1981. "Halfway between Quechua and Spanish: The Case for Relexification." in *Historicity and variation in Creole studies*. Vols. 57–78, edited by A. R. Highfield. Ann Arbor: Karoma Publishers.

Muysken, Pieter. 1997. "Media Lengua." Pp. 365–426 in *Contact languages: A Wider Perspective*, edited by S. Thomason. Amsterdam: J. Benjamins Pub. Co.

Muysken, Pieter. 2013. "Two Linguistic Systems in Contact: Grammar, Phonology, and Lexicon." Pp. 193–215 in *The handbook of bilingualism and multiculturalism*, edited by T. Bhatia and W. Ritchie. Malden, MA: Blackwell Publishing.

Narbona Jiménez, A., Cano Aguilar, R., & Morillo Velarde-Pérez, R. (2022). *El español hablado en Andalucía*. Universidad de Sevilla.

Navarro, T. (1956). Apuntes sobre el español dominicano. *Revista Iberoamericana*, *21*(41–42), 417–430.

Onosson, S., & Stewart, J. (2021a). A multi-method approach to correlate identification in acoustic data: The case of Media Lengua. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *12*(1), 1–30.

Onosson, S., & Stewart, J. (2021b). The Effects of Language Contact on Non-Native Vowel Sequences in Lexical Borrowings: The Case of Media Lengua. *Language and Speech PaPE 2019 Special Issue*, 1–30.

Onosson, S., & Stewart, J. (2023). Cotopaxi Media Lengua vowels: A preliminary analysis. *Proceedings of the 20th International Congress of Phonetic Sciences*, 3360–3365.

Papen, Robert. 2005. "On Developing a Writing System for Michif." *Linguistica Atlantica* (26):75–97.

Perrault, Charles. 1862. *Umiumi Uliuli (Bluebeard Trans.)*. Honolulu: Ka Nupepa Kuokoa.

Prado Ayala, G., Gonza Inlago, L., & Stewart, J. (2021). *Recetacunaca Yopa Comunidadmanta* (1st ed.). Lulu. https://www.lulu.com/en/ca/shop/jesse-stewart-and-lucia-gonza-inlago-and-gabriela-prado-ayala/recetacunaca-yopa-comunidadmanta/hardcover/product-zn7pgz.html

Pukui, M. K., & Elbert, S. H. (1986). *Hawaiian Dictionary*. University of Hawaii Press.

R Development Core Team. (2025). *_R: A Language and Environment for Statistical Computing_* (Version 4.4.1) [Windows 11]. R Foundation for Statistical Computing. https://www.R-project.org/

Real Academia Española. (2025). *Cucayo*. In *Diccionario de la lengua española* (23a ed.). Real Academia Española. https://dle.rae.es/cucayo

Rhodes, Richard A. 1985. *Eastern Ojibwa-Chippewa-Ottawa Dictionary*. Vol. 3. Berlin: De Gruyter Mouton.

Saint-Exupéry, Antoine de. 1946. *Quyllur Llaqtayuq Wawamanta*. Cusco, Peru: CBC — Centro de Estudios Regionales Andinos Bartolomé de las Casas.

Santo Tomás, Domingo de. 1560. *Grammatica o Arte de La Lengua General de Los Indios de Los Reynos Del Perú (Facsimile Edition)*. Mexico City: Impresso en Valladolid: por Francisco Fernandez de Cordoua, impressor de la M.R.

Schieffelin, Bambi B., and Rachelle Charlier Doucet. 1994. "The 'Real' Haitian Creole: Ideology, Metalinguistics, and Orthographic Choice." *American Ethnologist* 21(1):176–200.

Slaughter, Helen B., and Karen Watson-Gegeo. 1988. *Hawaiian Language Immersion: Evaluation of the Program's First Year, SY 1987–88. Evaluative/Feasibility (142)*. FL 018 274. Hilo: Hawaii Univ., Honolulu. Social Science Research Inst.

Stewart, J. (2011). *A Brief Descriptive Grammar of Pijal Media Lengua and an Acoustic Vowel Space Analysis of Pijal Media Lengua and Imbabura Quichua*. University of Manitoba.

Stewart, J. (2013). *Stories and traditions from Pijal: Told in Media Lengua* (1st ed.). CreateSpace.

Stewart, J. (2014). A comparative analysis of Media Lengua and Quichua vowel production. *Phonetica*, *71*, 159–182.

Stewart, J. (2015a). Intonation patterns in Pijal Media Lengua. *Journal of Language Contact*, *8*(2), 223–262.

Stewart, J. (2015b). Production and perception of stop consonants in Spanish, Quichua, and Media Lengua [PhD Dissertation, University of Manitoba].

Stewart, J. (2018a). Voice onset time production in Spanish, Quichua, and Media Lengua. *Journal of the International Phonetic Association*, *48*(2), 173–197.

Stewart, J. (2018b). Vowel perception by native Media Lengua, Quichua, and Spanish speakers. *Journal of Phonetics*, *71*, 177–193.

Stewart, J., Prado Ayala, G., & Gonza Inlago, L. (2020). Media Lengua Dictionary. *Dictionaria*, *12*, 1–3216.

Stewart, J. (2020). A preliminary, descriptive survey of rhotic and approximant fricativization in Northern Ecuadorian Andean Spanish varieties, Quichua, and Media Lengua. In R. Rao (Ed.), *Spanish Phonetics and Phonology in Contact: Studies from Africa, the Americas, and Spain*. John Benjamins.

Stewart, J. (accepted). Media Lengua in the Ecuadorian Andes. In L. Cerno, H.-J. Döhla, M. Gutiérrez Maté, R. Hesselbach, & J. Steffen (Eds.), *Contact varieties of Spanish and Spanish-lexified contact varieties*. Mouton De Gruyter.

Stewart, J., Gonza Inlago, L., & Prado Ayala, G. (2023). Cotopaxi Media Lengua is still very much alive. *Language Documentation and Conservation*, *17*, 49–63. http://hdl.handle.net/10125/74690

Stewart, J., & Onosson, S. (2023). Vowel raising in Cotopaxi Quichuan languages. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 3311–3315). Guarant International.

Stewart, J., Hadley, S., Gagne, M., & Osborne, C. (under-review). *The Status of the Voiceless Labial Fricative in Imbabura Media Lengua*.

Stewart, J., & Prado Ayala, G. (2025). Media Lengua Collection. *The Archive of Indigenous Languages of Latin America (AILLA)*. https://ailla.utexas.org/collections/1923/

Upton, G. J. G. (2017). *Categorical data analysis by example*. Wiley.

Young, Robert W., and William Morgan. 1972. *The Navajo Language*. Salt Lake City: Deseret Book Company.

Address for correspondence

Jesse Stewart
Department of Linguistics
University of Saskatchewan
916 Arts Building
Saskatoon, SK S7N 5A5
Canada

stewart.jesse@usask.ca

Publication history

Date received: 30 June 2025
Date accepted: 5 January 2026