Mixed language in flux? The various impacts of multilingual contact on Lánnang-uè's *wh*-question system

Wilkinson Daniel Wong Gonzales [iD]
The Chinese University of Hong Kong, Hong Kong SAR, People's Republic of China

Abstract

Aims and objectives: What exactly happens to a mixed language's system in a multilingual contact setting? This study aims to investigate the interactions between speakers' exposure to, frequency of, and proficiency in four languages (English, Tagalog, Hokkien, and Mandarin) and their influences on the *why*-fronting only *wh*-question system of Lánnang-uè, a mixed language used by the metropolitan Manila Lannangs. It also aims to test the validity of the assertion that symbiotic mixed languages are more likely to be in flux.
Methodology: The Lánnang-uè speakers participated in production and acceptability experiments.
Data analysis: Likelihood to front (production) and responses to a 7-point Likert-type scale (acceptability) were the dependent variables for several Bayesian linear mixed-effects models with age, frequency of language use, and language proficiency as primary fixed effects, sociolinguistic factors (e.g., attitudes, identity) as covariates, and participant (and when appropriate, item) as a random effect.
Findings: Both production and acceptability results showed that the effects of contact are numerous and far from homogeneous. They vary depending on the source language, *wh*-phrase type, and degree of consciousness. They corroborate the widely held belief that mixed languages are more stable in symbiotic contexts. However, this paper goes an extra step to show that this "instability" or variability is not always a consequence of contact-induced transfer. It demonstrates that when the source languages influence the stability or development of the mixed language, the effects can be diverse, encompassing aspects such as identity processes, language attitudes, structural transfer, and/or other sociolinguistic innovations.
Originality: This article is one of the first studies to examine the effects of contact between multiple languages on a mixed language variable using both production and acceptability experimental data in a five-language context. It is one of the very few variationist works in the Philippines that considers the effects of multilingualism on variation and change.

Keywords
Language variation and change, sociolinguistics, language contact, multilingualism, mixed languages, Bayesian regression, identity, language attitudes and ideologies, Lánnang-uè, Philippines

Corresponding author:
Wilkinson Daniel Wong Gonzales, Department of English, The Chinese University of Hong Kong, Fung King Hey Building, Shatin, New Territories, Hong Kong SAR, People's Republic of China.
Email: wdwonggonzales@cuhk.edu.hk

Introduction

One of the many outcomes of long-term multilingual contact is the emergence of mixed languages. These are languages that systematically incorporate elements from their source languages into their linguistic systems (Matras & Bakker, 2003; Meakins & Stewart, 2022; Thomason & Kaufman, 1988). In some cases of mixed language formation, the lexicon from one language and grammatical elements from another intertwine to form a new language (e.g., Media Lengua from Quichua and Spanish); in other cases, a systematic morphosyntactic frame is formed, with noun phrases coming from one language and verb phrases from another (e.g., Michif from French and Cree). There are also mixed languages that follow different developmental paths and display different structural properties (e.g., Sri Lanka Malay from Tamil, Sinhala, and Malay) (Meakins & Stewart, 2022).

Regardless of how the mixed languages are formed, scholars generally agree that languages of this type are independent, innovative linguistic systems (Bakker, 2003; Smith, 2000). Examples include Michif and Ma'a. However, there are other mixed languages similar to the Imbabura varieties of Media, which are spoken in Pijal, Ugsha, and Topo. These languages seem to be reliant on their source languages since they are still used alongside their parent language Quichua and, in certain instances, Spanish. The prior class of mixed languages—non-symbiotic mixed languages, or mixed languages that are in an linguistic ecology without the source languages—are characterized as more stable structurally speaking (Smith, 2000), whereas the latter group of mixed languages, the symbiotic ones, are claimed to have a linguistic system that is more "permeable" or susceptible to the ongoing influences of their source languages (Lipski, 2020, p. 410).

This paper delves into the study of the latter category of mixed languages, with a particular emphasis on the impact of multiple languages in contact on a Philippine language known as Lánnang-uè (Philippine Hybrid Hokkien). Lánnang-uè is predominantly an oral language that scholars have identified as a mixed language, coexisting with its source languages Tagalog, English, Hokkien, and Mandarin (Gonzales, 2018, 2021, 2022a, 2022b; Gonzales & Starr, 2020). A relatively recent study has provided quantitative evidence for the claim that mixed languages like Lánnang-uè could be less stable when used alongside their source languages: Lipski (2020) demonstrated that Media Lengua speakers in Imbabura are more likely to perceive Media Lengua as a distinct variety when the competing language is only Quichua, and less likely to do so when the speaker is both proficient in Spanish and Quichua. Following Lipski, I aim to further explore potential contact-induced "instability" in symbiotic mixed languages by examining the relationship between Lánnang-uè and its source languages. This will help assess the validity of the assertion that symbiotic mixed languages are more likely to be in flux. My aim is to investigate how the language background and exposure of the primary speakers, the Lannangs, specifically shape or influence the *wh*-question system of Lánnang-uè. Unlike Lipski's research, which primarily emphasizes perceived stability, my study takes a different approach by focusing on stability in both production and acceptability judgments. I adopt an experimental approach to test my hypotheses with greater control over the variables. While previous research, including Lipski's and other general mixed language literature, often focuses on "canonical" mixed languages with only two source languages, my investigation delves into the effects of multiple languages on a mixed language that incorporates not just two, but four languages.

The subsequent sections of this paper will follow this structure: First, I will provide an overview of the Lannangs and their languages, with a particular focus on Lánnang-uè. Next, I will delve into the *wh*-question system of Lánnang-uè and draw comparisons with the systems found in its source languages. Following that, I will present and discuss the results from the production and acceptability experiments. Subsequently, I will examine the effects of contact between multiple languages and their relevance to the development of Lánnang-uè. Finally, the paper will conclude with some closing remarks.

The Lannangs

The Lannangs (derived from the Hokkien phrase *lân láng* "our people"), many of whom identify as Chinese Filipinos (Gonzales, 2021), are broadly characterized in this paper as individuals with mixed Southern Chinese (Hokkien and Cantonese) and Filipino heritage. Specifically, they are a group that largely consists of late 19th-century to early 20th-century Southern Chinese immigrants to the Philippines and their descendants, the majority of whom are raised or born in the Philippines (Doeppers, 1986). The Lannangs residing in metropolitan Manila (henceforth, Lannangs)—the focus of this paper—live among those who identify only as Filipinos and those who identify only as Chinese (e.g., Mainland Chinese).

The relationship between the Lannang community and non-Lannang groups in Manila is complex. Generally, there is minimal animosity between Filipinos and Chinese. However, there are certain situations where tensions arise between these social groups. For instance, some non-Lannang Filipinos may classify Lannangs as Chinese, leading to stereotyping and generating friction between the two groups. In addition, Mainland Chinese, commonly referred to as *Taīdiōkâ* "Mainlander" by the Lannangs, may discriminate against the Lannangs and group them together with non-Lannang Filipinos in some instances. The Lannangs themselves, at times, have also been observed to marginalize or "other" the non-Lannang groups. Although Lannangs generally side with the Filipinos, based on Gonzales' (2022b) ethnographic interviews, several Lannangs—perceiving both the non-Lannang Filipinos and Mainlanders negatively—refused to fully identify as being part of either group, claiming that they form a distinct one.

What these dynamics highlight is the presence of a hybrid cultural identity—the Lannang identity. Continuously negotiating between their Filipino and Chinese identities, the Lannangs identify with a culture that draws from both Chinese and Filipino cultures, regardless of what their citizenship or ancestry is (Gonzales, 2021). That is, they do not necessarily need to have Filipino citizenship to identify as Lannang, nor do they need to have "pure" Chinese ancestry. What is common to all Lannangs is the shared experience of a hybrid culture. This includes being educated in a Lannang school with a multilingual curriculum (e.g., Filipino, English, Mandarin, Hokkien) and observing Lannang cultural traditions such as engaging in *Puāh Tiōngchū* "playing dice in Midautumn Festival" among others.

Multilingualism in the Lannang community

A key aspect of the hybrid Lannang culture is multilingualism. The Lannangs have a broad linguistic repertoire (Gonzales, 2018, 2022a, 2023b). They can, for example, speak in Philippine Hokkien, a dialect of Southern Min that has elements from Jinjiang Hokkien and Amoy Hokkien, both of which were brought to the Philippines in different waves of migration from Amoy and Jinjiang beginning in the 1850s (Doeppers, 1986, p. 382). The Lannangs are also knowledgeable in the regional language Tagalog—an indigenous language in the Austronesian group of Malayo-Polynesian languages—as well as the local English, Philippine English. It developed directly from contact between Tagalog and American English, introduced into mainstream Philippine society around the 1900s via an educational reform during the American occupation.

Some of the Lannangs claim to be proficient in Mandarin, a language that has been taught formally in most Lannang schools but is rarely used in peer-to-peer communication by mainstream Lannangs. Those with Cantonese heritage, forming around 10% of the estimated total Lannang population of 1.5 million, are also knowledgeable in "standard" Cantonese and/or Taishanese (Gonzales, 2022a), a Yue variety spoken in the Taishan region of Southern China. Apart from the enumerated languages, a language that most, if not all, Lannangs are familiar with is Lánnang-uè (derived from the Hokkien-origin phrase *lân láng uè* "our people speech/language").

Speakers of Lánnang-uè have varying levels of proficiency in other Lannang languages and are exposed to them in varying degrees, as mentioned above and as described in prior work (Gonzales, 2023b). For example, older speakers tend to have more Hokkien-speaking peers in their network and have been educated in Lannang schools that are either a split English-Chinese curriculum or a Chinese-dominant curriculum. These speakers have less exposure to Tagalog and English compared with Hokkien. In contrast, younger speakers have more Tagalog- and English-speaking peers and have been schooled in schools with an English- and Filipino/Tagalog-dominant curriculum (Chuaunsu, 1989b; Gonzales, 2023b). As a result, they have more exposure to Tagalog and English compared with Hokkien.

Lánnang-uè in Manila

Previous research characterized Lánnang-uè solely as unsystematic code-switching between Tagalog, English, and Hokkien, conflating it with the Hokkien-Tagalog-English "Hokaglish" code-switching phenomenon (Gonzales, 2016). New findings and preliminary investigations, however, indicate that Lánnang-uè is more than just ad hoc code-switching (Gonzales, 2018, 2022b, 2023a; Gonzales & Starr, 2020). The variety seems to have characteristics that are consistent with mixed languages: some of its elements are systematically drawn from English, Hokkien, and historically indigenous Philippine languages. By systematically, I mean in a non-random fashion—certain groups of linguistic elements tend to source from one language over the others.

The variety used in metropolitan Manila (henceforth, Lánnang-uè) draws from Tagalog, English, Hokkien, and occasionally, Mandarin elements in a systematic manner. Table 1 presents the distribution of Lánnang-uè elements by source language.

With respect to lexicon, among the 219 words examined,[1] approximately 49% are exclusively derived from Hokkien, while 5% originate from Tagalog, and 15% from English. For example, the word for "wide" is predominantly expressed using the English-sourced variant *wîde*, while "to sew" is conveyed as *tahî* from Tagalog. However, not all words in Lánnang-uè come from a single source language. About 31% of the basic lexicon tends to be expressed using variants

Table 1. Distribution of Lánnang-uè elements by source language (Gonzales, 2022b).

Linguistic components	Hokkien	Tagalog	English	Mandarin
Lexicon				
Origin (%)	72	29	37	0
Exclusive (%)	49	5	15	0
Grammar	• Components of the aspectual system • Negation system • Copula • Pronominal system • Some discourse particles • Tone system • Some conjunctions	• Components of the aspectual system • Plurality markers • Complementizers • Some conjunctions • Noun phrase prepositions • Interjections • Most discourse particles • Question marker • Post-modification structure • Part of the stress system	• Phrasal conjunctions • Some conjunctions and prepositions • Article system • Part of the stress system	• Some technical lexicon

from multiple source languages. For instance, the Lánnang-uè word for *worm* can be either the Tagalog-sourced *uôd* or the English-sourced *wòrm*. Overall, around 72% of the basic vocabulary in Lánnang-uè can be traced back to Hokkien, 29% to Tagalog, and 37% to English, based on the lexical distribution by source language. It becomes evident that Hokkien serves as the primary lexifier of this language variety, and it does not primarily acquire its basic vocabulary from Mandarin, potentially partially due to the history of Hokkien culture in the community, the late arrival of Mandarin in the linguistic ecology of the Lannang community, and the slightly negative attitudes toward the language ("Mainlander language") (Gonzales, 2022a). However, it is worth noting that some technical and culture-specific terms in Lánnang-uè are indeed sourced from Mandarin, like *siaukhaî*, which refers to "paper used for composition in Chinese."

In general, Lánnang-uè includes contributions from Hokkien (e.g., aspect markers, negation, copula), Tagalog (e.g., plurality marking, noun phrase derivational affixes), English (e.g., exclusively 15% of the lexicon, phrasal conjunctions), and Mandarin (Gonzales, 2023a). The data show that Lánnang-uè has a "composite" lexical-grammar split, meaning that elements of both the lexicon and grammar are drawn from multiple languages. Thus, based on initial explorations using structural criteria, Lánnang-uè has characteristics of a mixed language. It also meets the social criteria of mixed languages. For one, it is developed in a community where members want to express a mixed identity (see Gonzales, 2022b for an in-depth discussion).

On its emergence, Lannang language surveys, sociohistory, and linguistic experiments (Chuaunsu, 1989, p. 94; Gonzales, 2017, 2018) converge on the suggestion that Lánnang-uè emerged out of Hokkien-Tagalog "code-switching" by Hokkien-Tagalog bilinguals, who "interspersed Chinese with Pilipino [Tagalog]" (Chuaunsu, 1989, pp. 90–174). This code-switching was then deliberately developed into a "code-switching variety" or "language" (p. 95) by Hokkien-Tagalog-English-multilingual Lannang young professionals and teenagers as a means to establish community-wide belonging, to highlight their Filipino-ness (p. 94), and to avoid sounding "strange and alien" (pp. 175, 177) in the Lannang community. In its nascency, the Hokkien-Tagalog mixed language was only used in domains of the home and the community; however, children and teenagers began using it pervasively as their dominant language across all domains of speech communication (Chuaunsu, 1989), simultaneously infusing it with English elements as they generally become more knowledgeable in English. It is difficult to pinpoint when exactly the language emerged because of the lack of diachronic data, but I speculate that conventionalized Lánnang-uè emerged around the 1950s, based on the sociohistorical context on which the language emerged (see Tan (1993) for the role of the Filipinization movement) complemented by experiment findings (Gonzales, 2018).

Despite preliminary evidence of systematicity and languagehood, many Lánnang-uè speakers do not recognize Lánnang-uè as a distinct language, perceiving it as "adulterated" Hokkien (Ang See, 1990, p. 14; Chuaunsu, 1989; Uytanlet, 2014). Ang See (1990, p. 14), a Lannang scholar, has anecdotally characterized Lánnang-uè as an "adulterated" Chinese or "Chinese that is mixed with Filipino prefixes and suffixes . . . . syntax, and spoken in Filipino tones" (Ang See, 1990, p. 14). Another Lannang, Uytanlet (2014, p. 139), has also described this language as a mixture of English, Filipino, and Chinese—the result of "failure of transmission or mastery of the [Hokkien] language." There are, however, those that recognize it as a distinct language that sets them apart from the non-Lannangs (e.g., mainland Chinese and Filipinos) (1–3) (Chuaunsu, 1989).

(1) Q: *Dî tsiûwâ identifỳ taīdiōkâ?*
         How do you identify a Mainlander?
     A: *The way talk . . . Ūhmsī kap dân sáng lê kóng-uè ânī dapât si âni. Tsîgê tsâ sī Fīl-Chì. . . . Lánnang-uè pero the tòne. Intsîk na intsîk pà.*

> "The way they talk . . . They don't speak like us. It should be like what we are talking in
> now [Lánnang-uè]. This is a hallmark of being a Lannang. Lánnang-uè, but pay attention
> to the tone."
> (Female Lannang, 50s, Manila, interview) (Gonzales, 2022b)

(2) *lanlang-ue* 咱人話 *. . . symbolizes how we the Fil-Chi, Chi-Fil talks and communicates with each
other*
(Female Lannang, 20s, Quezon City, text message)

(3) *Hokkiên-uè sī kaûhuè lê ièng-e . . . khâ tshīm, Lánnang-uè si phóthōng dân lê kông-e.*
"Hokkien is used in church with deeper vocabulary Lánnang-uè is used in regular speech."
(Male Lannang, 20s, Manila, interview) (Gonzales, 2022b)

The skew toward viewing Lánnang-uè as broken Hokkien within the community and the common
mismatch between linguists' analyses of Lánnang-uè and community perceptions of the language
may have to do with the community-wide ideology of ethnic purity, and by extension, linguistic
purity. The notion of ethnic "purity" commonly surfaces in discussions of intermarriage and cul-
tural maintenance, where exposure to "foreign" or *huan-a* race and culture is perceived to be more
inferior than those who are not exposed to such. With respect to language, many community mem-
bers have also claimed that, in an ideal situation with no interactions with other languages, they
want to use "pure" Chinese. This ideology, I argue, could have contributed to the perception of the
community of mixed languages being "impure" versions of "standard" languages like Hokkien.

Wh-questions in Lánnang-uè and other Lannang languages

The focus of this paper is on the *wh*-question system of Lánnang-uè, arguably sourced from
Mandarin. One of the reasons for focusing on such is because the system betrays influences from
multiple languages. Furthermore, focusing on the *wh*-question system of Lánnang-uè and the vari-
ation found in it has the potential to provide valuable insights into the less explored area of syntax
and syntactic variation within the language. Syntactic variation has often taken a backseat to the
study of phonetic and lexical variation in Lánnang-uè, making this investigation particularly sig-
nificant and necessary. By delving into sociolinguistic and contact-induced patterns of variation at
the phonological and lexical levels, the findings of this study could help one determine whether
these patterns extend to other levels of language in Lánnang-uè. Furthermore, the focus on *wh*-
questions and variation offers a broader perspective on the nature and potential determinants of
syntactic variation, as well as variation in mixed languages in general. This could greatly contrib-
ute to our understanding of how sociolinguistic factors, including language contact factors, influ-
ence patterns of variation in "mixed languages" in distinct ways compared with other languages,
which, in turn, adds to the ongoing discussion on exceptionalism in contact languages.

    Based on data from a 7000-word Lánnang-uè question databank compiled in 2018 (Table 2), the
system adopted by Lánnang-uè speakers is as follows—when asking a *wh*-question, speakers typi-
cally have a simple or complex *wh*-phrase (Table 3) in the non-fronted position. In other words, for
most adjunct *wh*-questions, the *wh*-phrase should be positioned before the main verb (4–7).

(4) *În*   *tsiûwâ . . .*   *tsiûwâ*   *khāngkhô*   *a?*
3.PL   how   how   labor   PRT
"How did he or she labor?"
(PC0068)

Table 2. χ^2 distribution of Lánnang-uè *wh*-questions ($\chi^2 = 47.53$, $p < .001$).

	Why	*Non-why*
Fronted	50 (62.5%)	16 (14.4%)
Non-fronted	30 (37.5%)	95 (85.6%)

Table 3. Common *wh*-phrases and variants in Lánnang-uè.

General type	Lánnang-uè variants	Gloss
Argument	• *shāmmíh* • *siāmmíh* • *shāngá* • *siāngá*	"what" "who"
Adjunct	• *Kàna* • *tsiûwâ* • *tsâi-iùnn* • *tīsí* • *tōlóh* (simple) • preposition *tōlóh* (complex, e.g., *tiâm, tī,* or *ân tōlóh*)	"why" "how" "when" "where"

(5) *Tsî*   *sī*   ***tīsí***   *publîsh*   *-lê*   *a?*
     DEM   COP   when   publish   PF   PRT
     "When was this published?"
     (PC0068)

(6) *Î*   ***tōlóh***   *bêh*   *phiàng*   *a?*
     3.SG   where   want   explore   PRT
     "Where will they (SG) explode?"
     (PC0071)

(7) *Huaí*   *láng*   *bêh*   ***tiâm***   ***tōlóh***   *rescuè*   *a?*
     DEM.PL people   want   PREP   where   rescue   PRT
     "Where will the people rescue?"
     (PC0001)

The same goes for all *wh*-questions with an argument *wh*-phrase (e.g., *shāmmíh* "what," *shangá* "who") as the direct object (8 and 9).[2] The *wh*-phrase is located after the verb.

(8) *În*   *lê*   *sûhng*   *tsuê*   ***shammíh?***
     3.PL   PROG   count   as   what
     "What did they count that as?"
     (PC0068)

(9) *Lê*   *muhng*   *-le*   ***shangá?***
     PROG   ask   PF   who
     "Who are you asking?"
     (PC0068)

However, for *why*-questions, speakers generally front the *wh*-phrase before the subject, at the beginning of the question. For example, in (10), *kânâ* "why" is found directly before the subject *dîn* "2.PL," not before the verb *tsaû* "run."

(10) **Kânâ**   *dîn*   *bêh*   *tsaû*   *a?*
  why   2.PL   want   run   PRT
  "Why do you want to run?"
  (PC0071)

This is different from how *wh*-questions are formed in Lánnang-uè's source languages. In Hokkien, a *wh-in-situ* language (Sato, 2013, p. 300), the *wh*-phrase is by default never located at the beginning of the *wh*-question, regardless of the *wh*-phrase type (11–13). Of note is Hokkien *where*-phrases are always expressed as complex (prepositional) phrases, unlike Lánnang-uè.

(11) *Dìn*   **uisiammîh**   *tsuê*   *tsî*   *hang*   *taitsì?*
  2PL   why   do   DEM   CLF   affair
  "Why have you done this?"
  (The Amoy Audio Bible Project, Exodus 1:18b) (Biblical Seminary of the Philippines, 2011)

(12) *Gùn*   **siammîh sitsun**   *khuâkhîdî*   *iaū*   *âsi*   *tshuîtānn . . . ?*
  1PL   what   moment see   2.SG   hungry   or   thirsty
  "When did we see you hungry or thirsty?"
  (Matthew 25:44b)

(13) *Tsîde*   *áng*   *si*   **tui**   **tolóh**   *ū*   *tsîkhuân-ē*   *tìhuī*   *. . . ?*
  This   person   COP   PREP   where   have   this   wisdom
  "Where then did this man get all these things?"
  (Matthew 13:56)

In Tagalog, *wh*-questions with object *wh*-phrases are always formed with the *wh*-phrase at the beginning of the question (Schachter & Otanes, 1972, p. 51) (14–16).

(14) **Ano**   *ang*   *g-in-a-gawa*   *-mo?*
  What   ABS   RED-PERF-do   2.SG.ERG
  "What are you doing?"
  (Aldridge, 2002, p. 414)

(15) **Saan**   *-ka*   *b-um-ili*   *ng*   *libro?*
  Where   2.SG.ABS   PERF-buy   OBL   book
  "Where did you buy your books?"
  (Aldridge, 2002, p. 416)

(16) **Bakit**   *-mo*   *ako*   *pinatay?*
  Why   2.SG.ERG   ABS   PERF-kill
  "Why did you kill me?"
  (native speaker elicitation data 2019)

The way English forms *wh*-questions resembles Tagalog (17,18) in that it fronts the *wh*-phrase.

(17)   **What** did John kill?   (object argument *wh*-phrase)
(18)   **Why** did John kill Mary?   (adjunct *wh*-phrase)

In contrast with Tagalog and English, Mandarin speakers do not generally front the *wh*-phrase (Cheung, 2014). In (19), for instance, the *wh*-phrase *shenme dongxi* "what thing" is placed after post-verbally, at the end of the question.

(19) *Ni   mai   -le   **shenme   dongxi**?*
     2.SG   buy   -PERF   what   thing
     "What thing did you buy?"
     (Cheung, 2014, p. 398)

Despite the *in situ* nature of Mandarin's question system, *wh*-phrase fronting can occur when a speaker wants to mark questions with a high-scope *why* (see example below), resembling, but not identical to Lánnang-uè—a language that seems to front *why*-phrases regardless of scope. If the Mandarin speaker wants to indicate that the *why* phrase applies to the whole clause rather than the verb phrase, they may put it at the beginning.

In (20), for example, the person is asking for the reason why no person resigned, not the reasons nobody had for resigning. A Mandarin speaker would use the low-scope *yinwei shenme* (21) for the latter interpretation. In this sentence (21), the complex *wh*-phrase analog to high-scope *weishenme* is located after the subject *meiyou ren* "no one."

(20) ***Weishenme**   meiyou   ren   cizhi?*
     Why   no   person   resign
     "Why didn't anyone resign?"
     (Jin, 2014, p. 5)

(21) *Meiyou   ren   **yinwei**   **shenme**   cizhi?*
     No   person   because-of   what   resign
     "What reasons$_i$ did nobody have for resigning _i?"
     (Jin, 2014, p. 5)

Overall, then, Lánnang-uè, Hokkien, Tagalog, English, and Mandarin differ in the way their *wh*-questions are formed (Table 4). The *wh*-question system of Lánnang-uè exhibits the closest resemblance to that of Mandarin, implying that Mandarin could be the primary source language from which the *wh*-question system of Lánnang-uè has been derived, even if did not contribute any basic vocabulary to Lánnang-uè. The influence of Mandarin on the structure of Lánnang-uè is plausible due to its prominence in Lannang society. As mentioned earlier, Mandarin has been formally taught in schools, and its prevalence in the education sector has notably increased since at least the 2000s (Poa, 2004). Many members of the Lannang community have been gravitating toward Mainland China, leading to changes in language practices such as the use of simplified characters instead of traditional characters and adoption of a Mandarin curriculum over a Hokkien curriculum. Moreover, Mandarin holds instrumental value within the Lannang community, serving as a means of communication with Chinese speakers who do not share Lannang or Hokkien heritage. While Mandarin may not be the primary language for peer-to-peer communication among Lannangs, it remains significant in various aspects of their lives. As a result, Mandarin stands as a viable candidate source language alongside Tagalog, Hokkien, and English in shaping the *wh*-question system of Lánnang-uè.

However, the data reveal certain instances in Lánnang-uè where the rules of *wh*-questions are not consistently adhered to. In some cases, *why*-questions are not fronted, whereas non-*why questions* are fronted. This variation could potentially be influenced by the source languages like Hokkien, Mandarin, Tagalog, and English. To test the hypothesis with more control, I use experiments—production and acceptability experiments, in particular.

Table 4. Summary of *wh*-question systems in major Lannang languages.

Source language	Default *wh*-fronting (all)	Split *wh*-fronting (*why* vs others)	Scope fronting (*why*)
Lánnang-uè	✕	√	✕?
Hokkien	✕	✕	✕?
English	√	✕	✕
Tagalog	√	✕	✕
Mandarin	✕	✕	√

I specifically focus on the possible influences of four different languages on patterns of language mixed language variation, considering factors such as age, self-reported proficiency, frequency of language use, and other sociolinguistic factors like identity and language attitudes that have been found to condition variation in Lánnang-uè in previous work (Gonzales, 2023a). This approach allows for the identification of the relative and unique contributions of each language and other sociolinguistic factors on the system. Unlike prior work on language contact, which often views things simplistically at a broader level, this research seeks to explore the extent of influence and the specific factors related to each language (e.g., proficiency, frequency of use, structural factors) that mediate this influence. As a result, this study has the potential to enrich theories on language contact by providing a more nuanced probabilistic account of the influence of language contact and sociolinguistic factors on patterns of variation in contact languages, furthering our understanding of bilingualism and multilingualism.

Experiment 1: production task

The objective of this experiment is to investigate how multilingual contact influences the production of various types of Lánnang-uè *wh*-questions while taking into account the sociolinguistic factors known to influence language variation. To achieve this, I employ an elicitation task using a criminal investigation activity, specifically designed to elicit different types of *wh*-questions (e.g., *who, when, where*). I then analyze the data, considering the participants' age, self-reported proficiency in Lánnang-uè, and self-reported frequency of language use, all of which are obtained through a survey.

In the elicitation task, participants were tasked with gathering murder-related information from six witnesses and providing a description of the murder scene. To obtain this information, participants interrogated the witness character cards one by one, using *wh*-questions. Each witness character card was associated with a deck of response cards (stimuli) containing information corresponding to different types of *wh*-phrases, as illustrated in Figure 1.

After posing either a fronted or non-fronted *wh*-question to the character card and receiving responses through the use of response cards, participants were asked to describe the picture and take notes on a grid (Figure 1), which was designed to assist them in keeping track of the questions while also concealing the true purpose of the task.

The hypothesis put forward is that individuals with higher proficiency in the *wh*-fronting languages, such as English and Tagalog, would be more inclined to use the fronted construction when asking questions to the character card, regardless of the type of *wh*-question. The same tendency is expected for individuals who have greater exposure to these languages and those who frequently use them. Conversely, those with non-fronting Hokkien proficiency and exposure as well as those who frequently use Hokkien would be less likely to use the fronted construction. For speakers proficient in and exposed to Mandarin, it was anticipated that they would front more frequently, but

Figure 1. Production task (left: set-up of task; right: note-taking grid).

primarily in the context of *why*-questions due to the influence from Mandarin, a language known for its special *why*-fronting.

Participants

Seventy-seven adult Lánnang-uè speakers ranging from 21 to 99 years old participated—40 female, 37 male, equally spread across age decade groups (20s, 30s, etc.). They were all born and raised in the Philippines, had undergone Lannang education, and all have been at least exposed to Lánnang-uè, Tagalog, English, Hokkien, and Mandarin.

Procedure

Participants engaged in the task in-person. The sessions, conducted in summer 2019, were recorded using a monodirectional Zoom H6 recorder and a phone recorder, as backup. Participants were not informed of the research question; they were told that they were going to play a game and were not informed that I am investigating how they construct *wh*-questions.

Data analysis

After raw frequency analysis, multiple generalized linear mixed-effects regression models were fitted using R (Bürkner, 2017; R Core Team, 2015). One model was fitted for each of the subsets of the production data—(1) general *wh*-questions, (2) argument *wh*-questions, and (3) adjunct *wh*-questions—to examine the general (*wh*-questions in general) and local (arguments only, adjuncts

only) effects of multilingualism on the *wh*-question system of Lánnang-uè and the tendency to use *wh*-fronting.

The response or dependent variable was likelihood of using a fronted construction (coded as "0" for non-fronted and "1" for fronted), while the predictors included in the models can be found below, selected based on the literature on Lánnang-uè variation (Gonzales, 2023a; Gonzales & Starr, 2020). Multi-level categorical variables (e.g., type) were coded using Weighted Helmert coding conventions (Sonderegger, 2022). Random intercepts of participant were included for all models.

- *Wh*-type (categorical, *why, how, when, where, who, which*, inanimate *what*, animate *what*)
- Animacy (argument model only)
- Self-reported proficiency (*z*-scored: Hokkien, Mandarin, Tagalog, English)
- Self-reported frequency of use (*z*-scored: Hokkien, Mandarin, Tagalog, English)
- Ethnic identity (Chinese Filipino/Filipino-Chinese vs others)
- Pride in Lánnang-uè (*z*-scored)
- Acceptability of fronting (averaged values from Experiment 2)
- Age
- Sex (male vs female)
- Interaction terms
  - Age and sex
  - Type and proficiency (non-general models only)[3]
  - Type and frequency (non-general models only)
  - Type and identity (non-general models only)
  - Type and pride (non-general models only)

I utilized a Bayesian framework to examine the variation, aligning with contemporary research in variationist and general sociolinguistics (MacKenzie, 2020; Vasishth et al., 2018). Specifically, I employed a mixed-effects logistic model for analyzing a syntactic variable with two variants (Levshina, 2016). The Markov chain Monte Carlo (MCMC) algorithm was implemented using the brms package (Bürkner, 2017; R Core Team, 2015), and a total of 4 Markov chains were sampled. The MCMC algorithm comprised 30,000 iterations per chain for the general model and 10,000 iterations per chain for the argument and adjunct models. To account for initial sampling bias, each chain underwent a warm-up or burn-in period of 15,000 iterations (5,000 for argument and adjunct models). The thinning parameter was set at 2. For the intercept and slopes, I used weakly informative priors in the form of a normal distribution [0, 5] (Levshina, 2016, p. 252). As recommended by Vehtari et al. (2021, p. 683), I monitored the R̂ values and effective sample size (ESS) to verify convergence. Throughout the analysis, I ensured that the R̂ value remained within the range of 1.01, and the ESS value remained above 400 (see "Results and discussion" section for R̂ and ESS values).

I chose to use Bayesian methods instead of frequentist-oriented regression analysis. While Bayesian procedures are computationally intensive, they yield similar results to frequentist approaches (with *p*-values). However, the Bayesian approach offers distinct advantages. It provides direct and intuitive quantification of uncertainty through posterior distributions, which are easier to interpret than *p*-values and confidence intervals. These distributions represent the researcher's degree of belief in different parameter values. Furthermore, Bayesian methods allow insights into the absence of an effect, which is not possible in the frequentist framework. They also handle complex models with numerous predictors and interactions more effectively, avoiding issues of convergence that frequentist methods may encounter. In variationist regression analyses, where linguistic variation is

influenced by multiple factors and interactions, Bayesian methods offer reliable estimates even in highly complex models. This makes them particularly valuable for investigating linguistic variation and its underlying mechanisms (McElreath, 2020; Vasishth & Nicenboim, 2016).

After obtaining the results from the summary of Bayesian posterior draws (Bürkner, 2017), I employed the probability of direction (*pd*) measure to identify predictors that influence modified infinitive variation. A predictor is considered to have an effect on the dependent variable (i.e., modified infinitive variants) if its median value significantly deviates from zero or if the credibility intervals around the median do not encompass zero (Grafmiller et al., 2018; Levshina, 2016; MacKenzie, 2020; Makowski et al., 2019). To characterize the certainty or uncertainty of the effect's existence, this study employs the Bayesian statistical measure known as the "probability of direction" (*pd*)—which represents the proportion of posterior draws that share the same sign as the median. A higher *pd* value (close to 1) indicates a greater certainty that the positive or negative effect indicated by the median is present. However, a lower *pd* value (close to 0.5) suggests the possibility of the effect being non-existent (Makowski et al., 2019). For instance, if a predictor in this study's model has a median value of −1.3 and a *pd* of 0.95, it means that 95% of the posterior values are less than zero, and only 5% are greater than zero (Levshina, 2016). In simpler terms, there is a 95% likelihood that this predictor will have a negative effect (or reduce the likelihood) on choosing the fronting strategy over the non-fronting strategy, given that the reference category is set to "non-fronting."

Results and discussion

A frequency distribution of the production task is shown in Table 5. As expected, Lánnang-uè speakers generally do not place the *wh*-phrase at the beginning of the question, except in the case of *why*-questions, in which they are more likely to use fronted constructions. As can be observed, variation exists in the positioning of the *wh*-phrase: some speakers front in non-*why*-questions and some do not in *why*-questions.

In the general model fitting all *wh*-question data, there were main effects for certain types of *wh*-phrases, with the largest effect observed in the contrast between *why*-questions and non-*why* questions (β_{median}=2.89, SD=0.71, pd=1). Participants were more likely to front in *why*-questions compared with other *wh*-questions even after taking into consideration all other variables (e.g., gender, age, proficiency) (Figure 2(a)). They were also found to be more likely to front in questions with (inanimate) *what* compared with questions with *who* or *how*, but the effects are minimal.

Proficiency in all four languages did not seem to affect one's likelihood to front generally, but frequency of Hokkien and Tagalog use do (Table 6, Figures 2(b) and 3(c)). Fronting of questions is observed more among individuals who do not use Hokkien and those who use Tagalog frequently. Age, related to language exposure, has also been found to influence syntactic choice. Contrary to

Table 5. Distribution of *wh*-questions by position and phrase type (production).

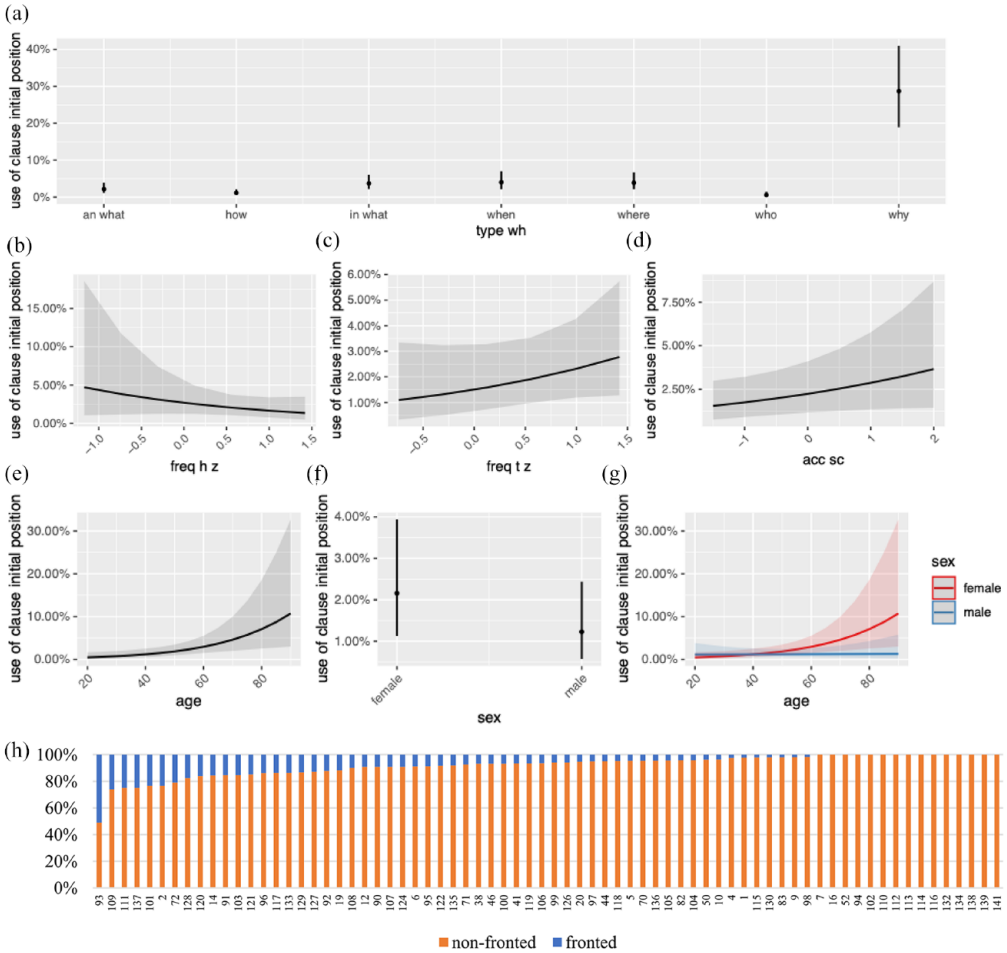| *Wh*-phrase position	Argument		Adjunct				Total
	What	*Who*	*Why*	*How*	*When*	*Where*	
Non-fronted	1,107 (95%)	324 (99%)	262 (62%)	717 (98%)	279 (94%)	332 (94%)	3,021 (91.6%)
Fronted	58 (5%)	3 (1%)	160 (38%)	15 (2%)	19 (6%)	21 (6%)	276 (8.37%)
Total	1,165	327	422	732	298	353	3,297

Figure 2. Plots showing the relationship between predictors and the likelihood to front (predictors: (a) type of phrase, (b) frequency of Hokkien use, (c) frequency of Tagalog use, (d) acceptability, (e) age, (f) sex, (g) age by sex, and (h) participant).

my initial expectations, older speakers, who are likely to have greater exposure to non-fronting Hokkien language, have been observed to favor fronted constructions over non-fronted ones (Figure 2(e)). This finding suggests that factors related to age, such as age-specific practices and degree of identification with the Lannang identity, might play a role in motivating older speakers to use fronted constructions more, rather than just their exposure to Hokkien. Identity and pride in Lánnang-uè do not play a notable role in conditioning syntactic variation, but sex and acceptability do. Female participants, particularly older female participants, were found to front more than other gender and age groups (Figure 2(f) and (g)). Acceptability of fronting, a measure derived from Experiment 2, appears to correlate with fronting likelihood ($\beta_{median}=0.25$, $SD=0.18$, $pd=0.92$) (Figure 2(d)). In other words, one who overtly accepts the fronting construction is also more likely to front. All in all, the results—controlled for individual effects (Figure 3(h))—suggest that *wh*-phrase type, frequency of Hokkien and Tagalog use, age, sex, and acceptability work hand in hand to encourage general fronting in Lánnang-uè.

Table 6. Bayesian model posterior draw estimates for predictors influencing likelihood to front (general model), reference levels in boldface, *pd* > 0.8 in gray.

Predictors	Median	SD	89% CI (HDI)	pd	R	ESS
Fixed effects						
Intercept	−5.36	1.13	[−7.2, −3.58]	1	1	28,080
Type (**animate *what*** vs inanimate *what*)	0.55	0.33	[0.03, 1.08]	0.96	1	23,799
Type (*who* vs ***what***)	−1.64	0.63	[−2.66, −0.68]	1	1	24,498
Type (*which* vs ***who, what***)	0.66	4.65	[−6.86, 7.97]	0.56	1	24,465
Type (*where* vs ***which, who, what***)	0.72	1.21	[−1.23, 2.61]	0.73	1	24,751
Type (*when* vs **all except why and how**)	0.61	0.98	[−0.91, 2.23]	0.73	1	24,677
Type (*how* vs **all except why**)	−0.77	0.83	[−2.11, 0.55]	0.82	1	24,868
Type (*why* vs **all**)	2.89	0.71	[1.75, 4.03]	1	1	24,423
Proficiency (Hokkien)	0.23	0.58	[−0.7, 1.15]	0.65	1	25,707
Proficiency (Mandarin)	0.01	0.44	[−0.7, 0.71]	0.51	1	26,406
Proficiency (Tagalog)	0.32	0.47	[−0.42, 1.08]	0.76	1	27,484
Proficiency (English)	−0.42	0.55	[−1.31, 0.43]	0.78	1	25,116
Frequency (Hokkien)	−0.5	0.52	[−1.32, 0.34]	0.84	1	26,325
Frequency (Mandarin)	−0.02	0.4	[−0.66, 0.62]	0.52	1	27,661
Frequency (Tagalog)	0.44	0.41	[−0.2, 1.11]	0.86	1	27,435
Frequency (English)	0.37	0.52	[−0.44, 1.22]	0.77	1	25,871
Identity (Chinese Filipino/Filipino Chinese vs **others**)	−0.43	0.92	[−1.88, 1.06]	0.68	1	29,046
Pride (Lánnang-uè)	−0.01	0.29	[−0.48, 0.45]	0.51	1	27,348
Acceptability	0.25	0.18	[−0.03, 0.54]	0.92	1	23,238
Age	0.03	0.02	[0, 0.05]	0.93	1	28,063
Sex (**male** vs female)	−1.76	1.26	[−3.76, 0.26]	0.92	1	24,823
Age: Sex	0.04	0.02	[0.01, 0.08]	0.97	1	24,900
Random effects (Intercepts only)						
Participant (*SD*, Intercept)	1.37	0.2	[1.06, 1.69]	1	1	17,406

Note. pd: probability of direction; *SD*: standard deviation; *CI*: confidence interval; *ESS*: effective sample size; *HDI*: highest density interval.

The argument model fitting only *who-* and *what-*question data—with an additional animacy covariate factor—indicated the influence of Hokkien, Mandarin, and English proficiency as well as frequency of Mandarin use on *wh*-fronting. All other factors considered equal, the more proficient the speaker is in the fronting language English (β_{median}=1.05, *SD*=1.13, *pd*=0.84) and special-fronting language Mandarin (β_{median}=0.81, *SD*=0.98, *pd*=0.8) and the less proficient the speaker is in non-fronting language Hokkien (β_{median}=−1.13, *SD*=1.27, *pd*=0.86), then the higher the likelihood of them fronting in argument *wh*-questions. The less often a participant uses Mandarin (β_{median}=−1.01, *SD*=0.93, *pd*=0.87), the more likely they will front as well. The divergence of results between frequency and proficiency in Mandarin is somewhat expected since Mandarin is a *wh-in-situ* default language with some licensed fronting conditions. Pride and acceptability also play some role in variation: those who are proud of Lánnang-uè and those who explicitly accept argument structures (Experiment 2) were found to increase fronting rates. Animacy, identity, age, and sex as well as Tagalog proficiency and frequency of use did not play a notable role in the syntactic variation in argument *wh*-questions.
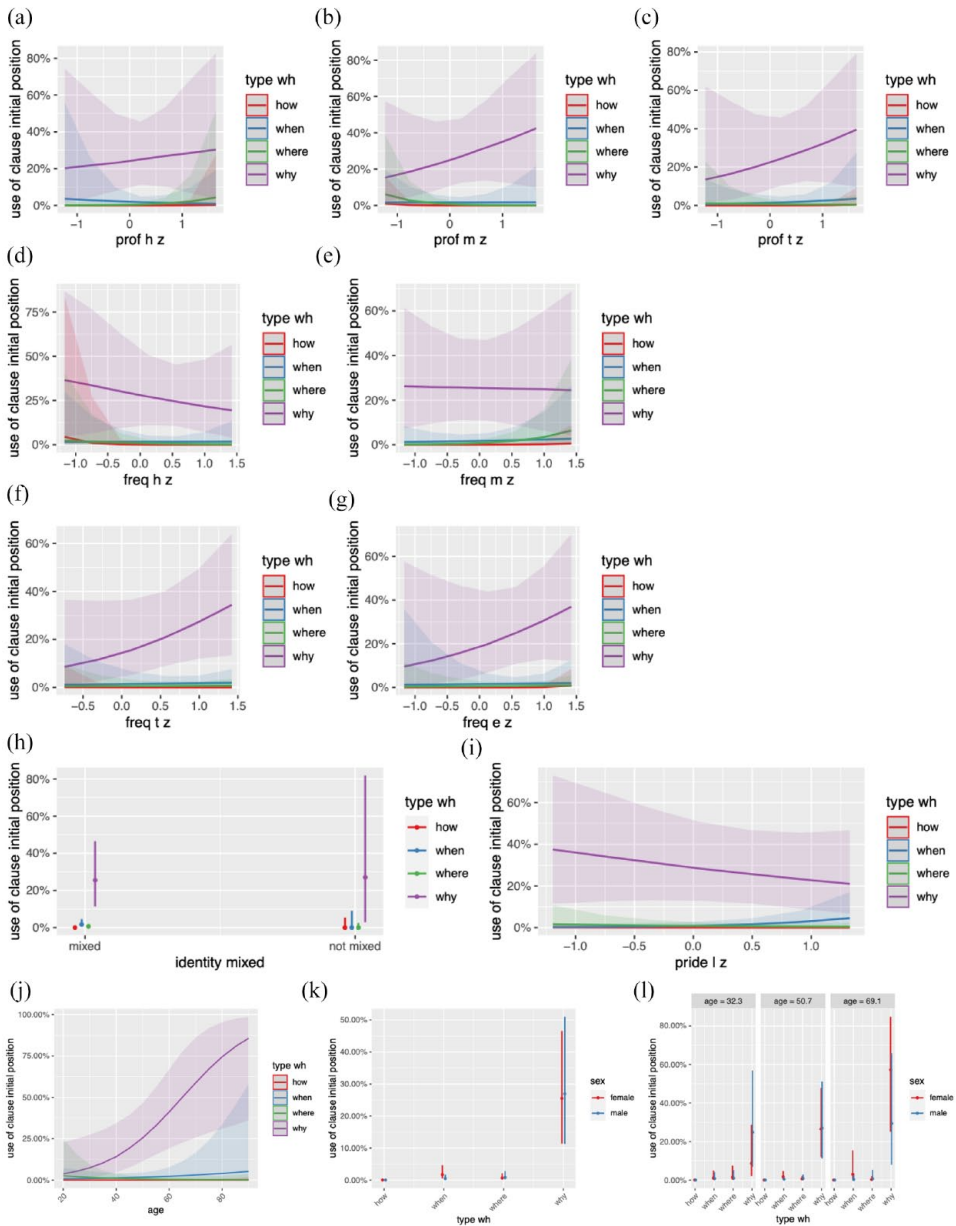
Figure 3. Plots showing notable interactions between *wh*-type and other predictors on likelihood to front (predictors: (a) Hokkien proficiency, (b) Mandarin proficiency, (c) Tagalog proficiency, (d) frequency of Hokkien use, (e) frequency of Mandarin use, (f) frequency of Tagalog use, (g) frequency of English use, (h) hybridity of identity, (i) pride in Lánnang-uè, (j) age, (k) type of phrase by sex, and (l) type by age and sex).

As expected, a significant effect of type (*why* vs *how/when/where*) was found in the adjunct model ($\beta_{\text{median}} = 1.86$, $SD = 1.89$, $pd = 0.84$). In adjunct questions, speakers generally tend to front in *why*-questions. The results of the adjunct model further revealed highly probable interaction effects between type and Mandarin proficiency ($\beta_{\text{median}} = 2.12$, $SD = 0.77$, $pd = 1$)

and between type and Hokkien proficiency (β_{median}=−0.85, SD=0.98, pd=0.81) (Figure 3(a) and (b)). Confirming my hypothesis, speakers proficient in special *why*-fronting Mandarin were more likely to front in *why*-questions. Those proficient in default non-fronting Hokkien were unexpectedly found to have slightly higher rates of *wh*-fronting in *why questions;* however, they were less likely to front in *when* questions, following expectations, given Hokkien does not also front in *when*-questions.

Frequency also interacts with *wh*-type to condition variation (Figure 3(d) to (g)): those who frequently use non-fronting Hokkien (β_{median}=1.15, SD=0.88, pd=0.91) tend to use more non-fronted *why* constructions. Those who frequently use fronting languages Tagalog (β_{median}=1, SD=0.92, pd=0.86) and English (β_{median}=−1.11, SD=0.89, pd=0.9) tend to use more fronted *why*-constructions. Interestingly, frequency of special-fronting Mandarin use makes participants more likely to front non-*why constructions* rather than *why constructions* (β_{median}=−1.41, SD=0.92, pd=0.86). The results altogether suggest that while proficiency in Mandarin contributed significantly to the dominant/default *why*-fronting system in Lánnang-uè, it is frequency of Mandarin use (frequent *why*-fronting use) that contributed to the variation in the system, that is, the generalization of the fronting strategy from just exclusively *why*-questions to all adjunct *wh*-question types.

Age and sex interact to influence the variation in adjunct *wh*-question production (Figure 3(l)). Older and female speakers tend to have markedly higher rates of fronting compared with other age and sex groups for *why* and *when questions.* Younger male speakers have the next highest rates of fronted *why*-question use.

Lannang participants who identify with hybrid identities such as Chinese Filipino or Filipino-Chinese and those who are proud of the mixed language Lánnang-uè tend to have lower but more consistent rates of fronted *why*-questions use than those who identify as exclusively Chinese or Filipino (Figure 3(h) and (i)), a stark departure from the argument findings, where identity factors play a minimal role and where the pride factor influences variation positively, such that those who are proud of Lánnang-uè tend to use more fronted argument *wh*-constructions. The results, controlling for individual effects, altogether show the complexity of the interactions between sociolinguistic factors and syntactic variation.

The findings collectively provide a first indication that the *wh*-question system of Lánnang-uè is influenced by the speakers' multilingualism or contact with other languages in the Lannang linguistic ecology. The speakers' proficiency in different languages and frequency of language use—along with other sociolinguistic factors—seem to impact the *why*-fronting system in Lánnang-uè.

When a language comes into contact with other languages, speakers may deliberately choose not to let these languages impact their language for a variety of reasons (e.g., language attitudes, etc.) (Thomason, 2007), but it is not unusual for the speakers' language use to be affected by other languages that they are knowledgeable in or are exposed to, subconsciously or consciously. Their use of languages that are already mixed, particularly mixed languages, is no exception—after all, mixed languages, like other "standard" languages, are not immune to effects of community multilingualism.

Experiment 2: scale-rating task

The second experiment utilized a scale-rating task designed to investigate the impact of multilingual contact on another aspect of grammaticality—acceptability. It examines how multilingualism can affect the speakers' explicit awareness and evaluation of the Lánnang-uè question *why*-fronting system. In a nutshell, the participants rated questions using a 7-point Likert-type scale—1 (*not acceptable in Lánnang-uè*) to 7 (*very acceptable*), with half of the questions manipulated for position. They rate questions that are not usually fronted in Lánnang-uè, and vice versa.

Table 7. Argument *wh*-question sample stimuli.

Type	Animacy	Position	Stimuli				sample
What	Inanimate	Non-fronted	*Hîgē*	*lāngkhêh*	*bêh*	*tūsôk*	***siāmmíh***[a]*?*
			the	visitor	will	stab	what
		Fronted	***Siāmmíh***[a]	*hîgē*	*lāngkhêh*	*bêh*	*tūsôk?*
			what	the	visitor	will	stab
	Animate	Non-fronted	*Hîgē*	*lāngkhêh*	*bêh*	*tūsôk*	***siāmmíh***[b]*?*
			the	visitor	will	stab	what
		Fronted	***Siāmmíh***[b]	*hîgē*	*lāngkhêh*	*bêh*	*tūsôk?*
			what	the	visitor	will	stab
Who	Animate	Non-fronted	*Hîgē*	*lāngkhêh*	*bêh*	*tūsôk*	***siāngá****?*
			the	visitor	will	stab	who
		Fronted	***Siāngá***	*hîgē*	*lāngkhêh*	*bêh*	*tūsôk?*
			who	the	visitor	will	stab

Note. Translation: "What/who will the visitor stab?"
[a]Object.
[b]Animal.

Table 8. Adjunct *wh*-question sample stimuli.

Position	Stimuli sample (frame)						
Non-fronted	*Hîgē*	*yayá*	*{ }*	*bêh*	*káh*	*hîgē*	*tshúhng?*
	the	maid	why	will	carry	the	bed
Fronted	*{ }*	*hîgē*	*yayá*	*bêh*	*káh*	*hîgē*	*tshúhng?*
	why	the	maid	will	carry	the	bed

Note. Translation: "{*Why/how/when/where*} will the maid carry the bed?" { } = *kàna* "why," *tsiûwâ/tsâi-iùnn* "how," *tīsí* "when," *tolóh* "where".

The predictions are similar to that of the production experiment. Increased exposure, frequency, and proficiency in the fronting languages English and Tagalog should make speakers rate fronted constructions higher. From the production experiment, the possibility that Mandarin *why*-fronting could be generalized to other *wh*-phrases was brought up, so, I would not be surprised if there are proficiency, frequency, and exposure effects on variation in acceptability that are related to Mandarin. Those who use non-fronted Hokkien frequently and have high proficiency in the language should encourage participants to give lower ratings for fronted constructions. Speakers proficient in Mandarin should be more likely to rate fronted *why*-questions higher than those who are not. I hypothesize opposite effects for the ratings for non-fronted constructions: higher Hokkien proficiency and higher frequency of Hokkien use should increase ratings, whereas lower Tagalog and English proficiency and frequency of Tagalog and English use are expected to decrease ratings for non-fronted constructions.

Participants

Out of the 77 adult Lánnang-uè speakers in the production experiment, 72 participated in the scale-rating task, as some speakers decided not to continue with the session.
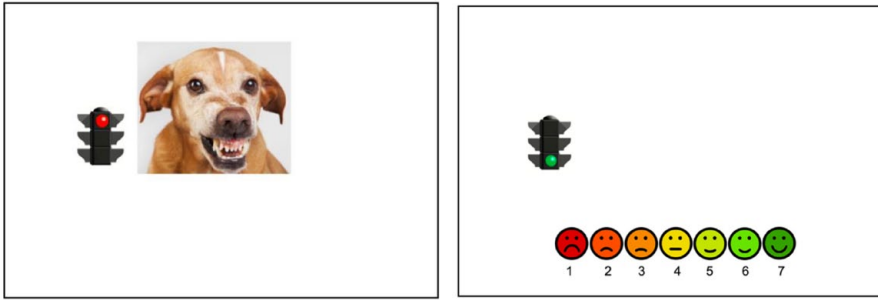
Figure 4. Screenshots of scale-rating task (first frame, left; last frame, right).

Materials

A young male native speaker of Lánnang-uè recorded 146 question stimuli distributed across 14 conditions (7 types of *wh*-phrases, including inanimate and animate *what* × position), with each condition having 10 or 11 items. Half of the stimuli have the dominant structural variants in Lánnang-uè (*why fronting*, non-*why* non-fronting), while the other half contain the uncommon variants (*why* non-fronting, non-*why fronting)* (Tables 7 and 8).

Procedure

Participants did the scale-rating task directly after the elicitation task. After doing a practice block of five trials to familiarize them with the task, they listened to the randomized stimuli on a portable laptop running PsychoPy 3.0 (Peirce, 2007), using over-ear headphones. Each participant did not rate all 146 stimuli; while they were exposed to all conditions, each only encountered two to five items per condition, as opposed to the total number of items (i.e., 11 for argument questions, 10 for adjunct questions).

Initially, participants listened to the audio stimuli twice and subsequently viewed a picture associated with the stimuli. Afterward, they were prompted to provide their rating. To do so, they had to wait for a green light to appear on the screen before clicking the appropriate button on the laptop keyboard. A special keyboard cover was used—the number keys (1–7) were superimposed with smiley faces that correspond to the acceptability scale (Figure 4). This was done to make the task more intuitive for older participants, who often made mistakes (e.g., treating 1 as acceptable) using the numerical Likert-type scale during the pilot study. The participants' responses in the task were then recorded using the software.

Data analysis

After analyzing the means of the responses, multiple linear mixed-effects regression models were fitted using R. One model was fitted for each of the six subsets of the production data: (1) fronted general *wh*-questions, (2) fronted argument *wh*-questions, (3) fronted adjunct *wh*-questions, (4) non-fronted general *wh*-questions, (5) non-fronted argument *wh*-questions, and (6) non-fronted adjunct *wh*-questions.

The acceptability models generally share the same dependent variables, predictors, coding system, and modeling algorithm parameters as the production models. The only differences involve the inclusion of the production factor instead of the acceptability factor—a frequency index derived

Table 9. Summary of scale-rating responses (z-scored, by type).

Position	*Wh*-phrase type	*n*	*z-scores*	*SD*
Non-fronted	*how*	360	0.56	0.72
	what	1,131	0.66	0.69
	when	360	0.69	0.69
	where	360	−0.53	0.87
	who	450	0.52	0.82
	why	360	0.68	0.69
Fronted	*how*	360	−0.35	0.85
	what	1,135	−0.84	0.71
	when	360	−0.14	0.90
	where	360	−0.69	0.79
	who	452	−0.66	0.80
	why	360	0.53	0.75

from Experiment 1—and the inclusion of random intercepts for item, to control for the possible effects of individual experiment items on the ratings.

Results and discussion

An analysis of the data reveals that the participants held a general aversion toward fronted constructions (mean *z*-score = −0.493), whereas they accepted non-fronted constructions (mean *z*-score = 0.492). Among the fronted constructions, only *why*-questions received high ratings, and among the non-fronted ones, all except where-questions were well-received (refer to Table 9). While the former result was anticipated, the latter outcome was unexpected.

The low rating for *where*-questions could be accounted for by most participants' preference for complex *where*-phrases (e.g., *tī tōlóh* "at where") over simple *where*-phrases (e.g., *tōlóh* "where"), as evidenced in their production. The experiment did not have complex *wh*-phrases stimuli to allow for sufficient control. Had it included such phrases, the results should show a positive rating for non-fronted *where*-questions.

Contrary to expectations, non-fronted *why*-questions were also rated high, possibly because Lánnang-uè is still viewed by many as a variety of Hokkien, as mentioned earlier and reported in previous work (Gonzales, 2022b). That is, participants could be judging the stimuli using their knowledge of Hokkien despite being instructed to judge Lánnang-uè. While it may be impossible to experimentally control for Hokkien proficiency, the statistical modeling used can allow us to virtually control for the said proficiency, so that the results can be assumed to have come from speakers with equal Hokkien proficiency.

Ratings for fronted constructions

As expected, the overall model examining fronted data displayed the main effects of type consistent with the earlier description (Figure 5(a)). In addition, it revealed main effects related to proficiency in Hokkien and English (Figure 5(b) and (c)), as well as the frequency of Hokkien, Tagalog, and English usage (Figure 5(d) and (e)). Participants proficient in non-fronting Hokkien tended to provide higher ratings for fronted constructions ($\beta_{median} = 0.63$, $SD = 0.4$, $pd = 0.94$), while those proficient in fronting English tended to give lower ratings for such constructions ($\beta_{median} = 0.81$, $SD = 0.38$, $pd = 0.98$). Moreover, the frequency of non-fronting Hokkien ($\beta_{median} = −0.65$, $SD = 0.35$,
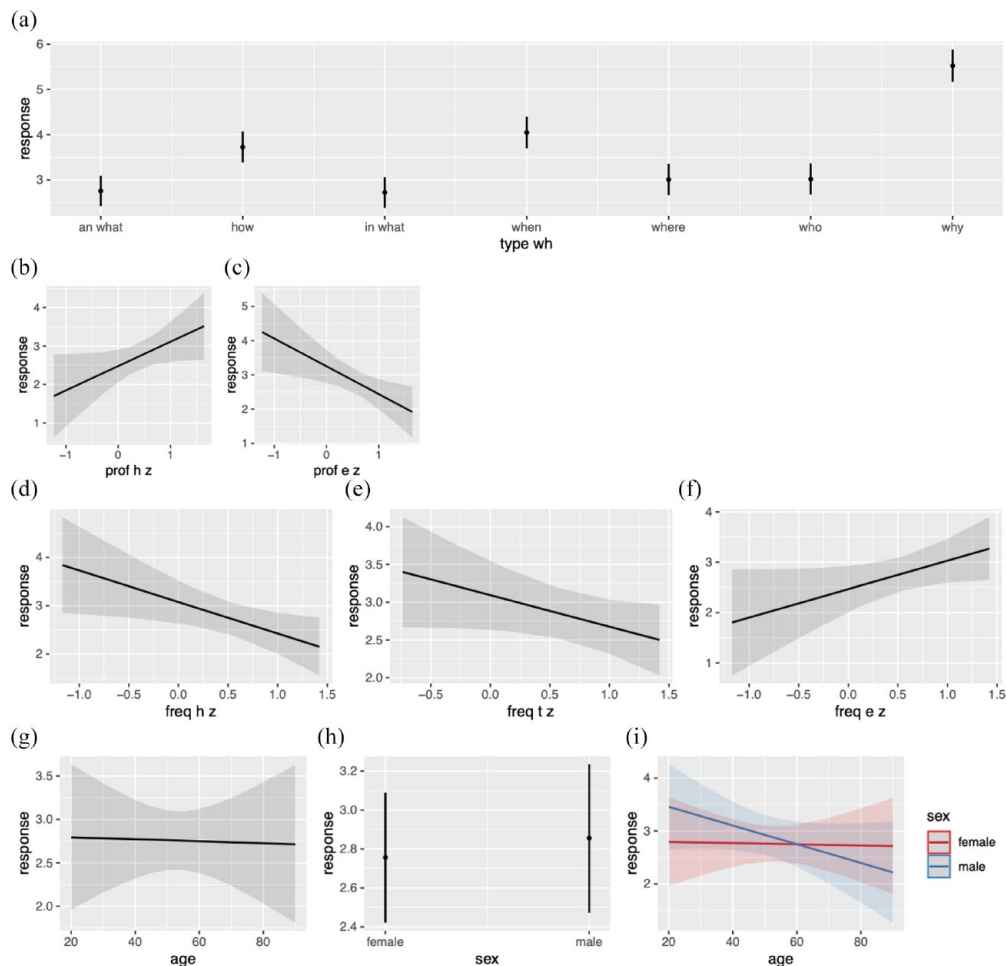
Figure 5. Plots showing relationship between notable predictors and acceptability of fronted constructions (predictors: (a) type of phrase, (b) Hokkien proficiency, (c) English proficiency, (d) frequency of Hokkien use, (e) frequency of Tagalog use, (f) frequency of English use, (g) age, (h) sex, and (i) age by sex).

$pd=0.97$) and fronting language Tagalog usage ($\beta_{median}=-0.42$, $SD=0.29$, $pd=0.93$) was associated with lower ratings for fronted constructions. Conversely, higher English usage was linked to an increased likelihood of rating fronted constructions more favorably ($\beta_{median}=0.57$, $SD=0.37$, $pd=0.94$).

Sex ($\beta_{median}=-0.99$, $SD=0.85$, $pd=0.88$) and its interaction with age ($\beta_{median}=0.02$, $SD=0.02$, $pd=0.86$) were also found to shape ratings: male participants (Figure 3(h)), particularly those who are young (Figure 3(i)), tend to provide higher ratings for fronted constructions. Older male participants provided the lowest ratings out of the sex and age groups. Identity as Chinese Filipino/ Filipino-Chinese, pride in Lánnang-uè, and production frequency did not seem to influence ratings significantly.

The argument model revealed the same effects outlined in the general model. For the adjunct model, type (*why* vs other adjuncts) expectedly had an effect. Speakers still rated fronted *why*-questions higher
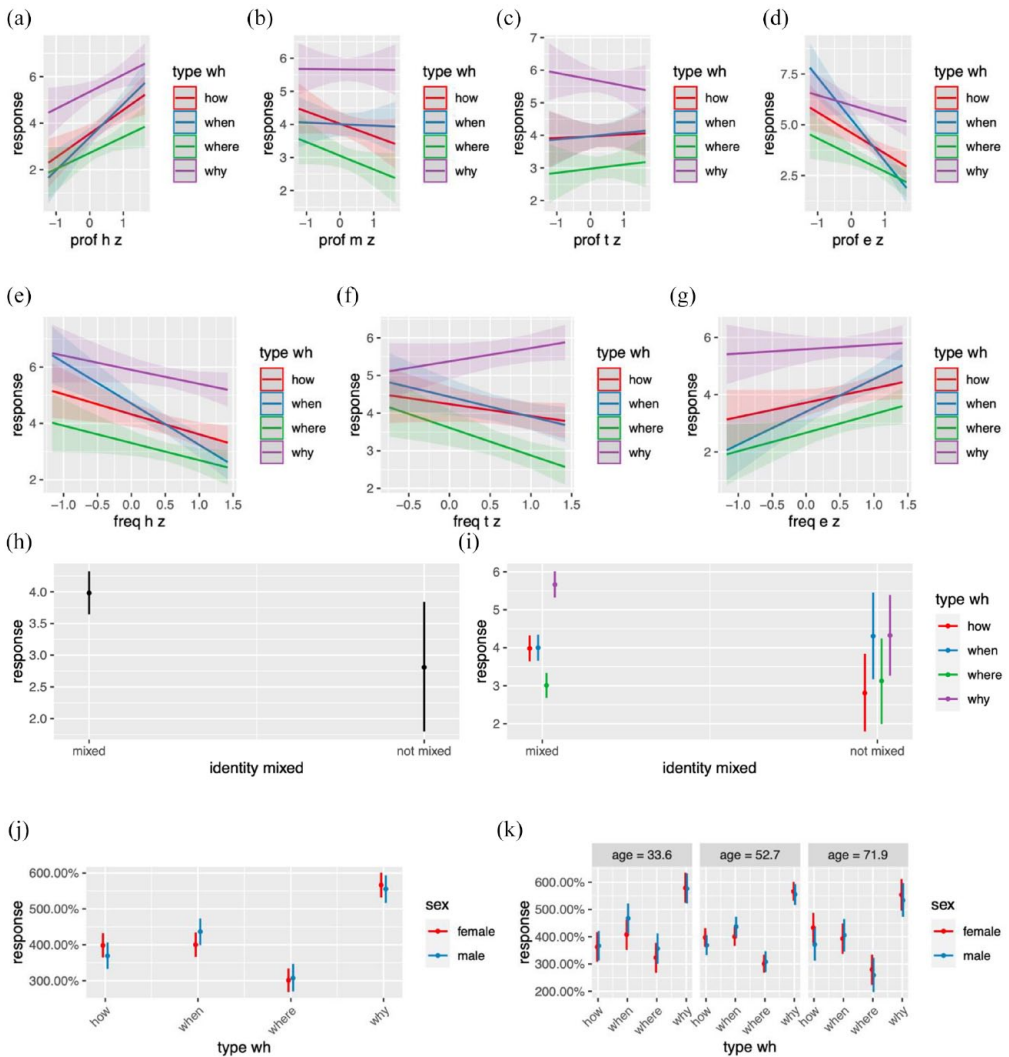
Figure 6. Plots showing notable interactions between *wh*-type and other predictors on acceptability of adjunct fronted constructions (predictors: (a) Hokkien proficiency, (b) Mandarin proficiency, (c) Tagalog proficiency, (d) English proficiency, (e) frequency of Hokkien use, (f) frequency of Tagalog use, (g) frequency of English use, (h) hybridity of identity, (i) hybridity of identity by type, (j) type by sex, and (k) type by age and sex).

than non-*why* fronted questions. The model reveals that the type of wh-question interacts with various factors such as proficiency, frequency, age, sex, and Chinese Filipino/Filipino-Chinese identity, influencing the ratings of fronted constructions. Proficiency in non-fronting Hokkien appears to have led to higher ratings for fronted adjunct *wh*-questions (Figure 6(a)) ($\beta_{median} = -0.31$, $SD = 0.28$, $pd = 0.86$). However, proficiency in fronting Mandarin has no notable effect on *why*-fronting and *when*-fronting but appears to encourage lower ratings for *how*- and *where*-fronting (Figure 6(b)) ($\beta_{median} = 0.27$, $SD = 0.21$, $pd = 0.9$). When it comes to fronting languages like Tagalog and English, proficiency seems to have decreased ratings for fronted *why*-questions (Figure 6(c) and (d)). Interestingly, Tagalog

proficiency appears to have increased ratings for *when-*, *where-*, and *how-*fronting ($\beta_{median}=-0.29$, $SD=0.23$, $pd=0.9$), whereas English proficiency shows the opposite effect ($\beta_{median}=0.8$, $SD=0.27$, $pd=1$) (Figure 6(c) and (d)).

Although the patterns related to proficiency generally contradict my expectations, the patterns associated with frequency largely align with them. Increased frequency of non-fronted Hokkien use is correlated with lower ratings for fronted constructions (Figure 6(e)) ($\beta_{median}=0.8$, $SD=0.27$, $pd=1$). Likewise, higher frequency of fronted English use is correlated with higher ratings for fronted constructions (Figure 6(g)) ($\beta_{median}=-0.61$, $SD=0.25$, $pd=0.99$). In addition, a higher frequency of fronted Tagalog use is correlated with higher ratings for fronted *why-*question constructions (Figure 6(f)) ($\beta_{median}=0.88$, $SD=0.21$, $pd=1$). Surprisingly, a higher frequency of fronted Tagalog use is also associated with lower ratings for fronted *how-, when-*, and *where-*questions, contrary to what was expected.

Male participants were more inclined to give higher ratings for fronted *when* constructions, while female participants tended to give higher ratings for fronted *how* constructions ($\beta_{median}=0.69$, $SD=0.59$, $pd=0.88$). The data indicate that younger female participants lead in the acceptance of fronted *why constructions*, while younger male participants show greater acceptance of fronted *when* and *where* constructions but exhibit aversion toward fronted *how* constructions ($\beta_{median}=-0.01$, $SD=0.01$, $pd=0.81$). It is worth noting that these younger speakers have more exposure to fronting languages such as Tagalog and English. The results imply that exposure to fronting languages may lead to increased acceptance of fronted structures for *why, when*, and *where* constructions, but for *how-*fronting, it seems to reduce the acceptance of such structures, potentially showcasing the divergent effects of language contact.

Of particular interest is the interaction between identity and syntactic variation. Individuals who identify as Chinese Filipino or Filipino-Chinese demonstrate a higher acceptance of fronted *wh-*questions ($\beta_{median}=0.52$, $SD=0.55$, $pd=0.83$). Moreover, they appear to possess clearer and more consistent intuitions about what is considered acceptable or not ($\beta_{median}=1.08$, $SD=0.48$, $pd=0.99$). For instance, in Figure 6(i), among Chinese Filipinos and Filipino-Chinese, fronted *why* consistently receives acceptance, while fronted *when/where/how* does not. In contrast, those who identify exclusively as Filipino or Chinese exhibit not only lower ratings but also less consistency, as indicated by the variance in the data. These findings indicate that identity plays a significant role in the maintenance of the Lánnang-uè *wh-*question system in the Lannang community. This complements existing data that highlight positive attitudes as one of the contributing factors to the mixed language structural preservation (Gonzales, 2023a).

Overall, the acceptability data regarding fronted constructions showed some expected effects. For instance, frequent use of non-fronted Hokkien was associated with lower ratings for fronted constructions, while frequent use of fronted English was expected to result in higher ratings for such constructions. Moreover, younger speakers with more exposure to fronting languages like Tagalog and English tended to give higher ratings for fronted constructions. However, certain results contradicted my initial expectations. Instead of leading to lower ratings for fronted constructions, proficiency in Hokkien was found to increase ratings. Similarly, proficiency in English and the frequency of Tagalog use (except for non-*why* questions) did not lead to higher ratings for fronted constructions; on the contrary, they resulted in decreased ratings. In addition, among younger men exposed to fronting languages, there was a tendency to reject the fronted *how* construction. These divergent results indicate that knowledge, use, and exposure to the source languages and their *wh-*question systems can, but do not always necessarily, align with the explicit assessment of fronted *wh-*question structures. This finding is not surprising if we consider the possibility that Lánnang-uè has some level of independence from its source languages. (Gonzales, 2022b).
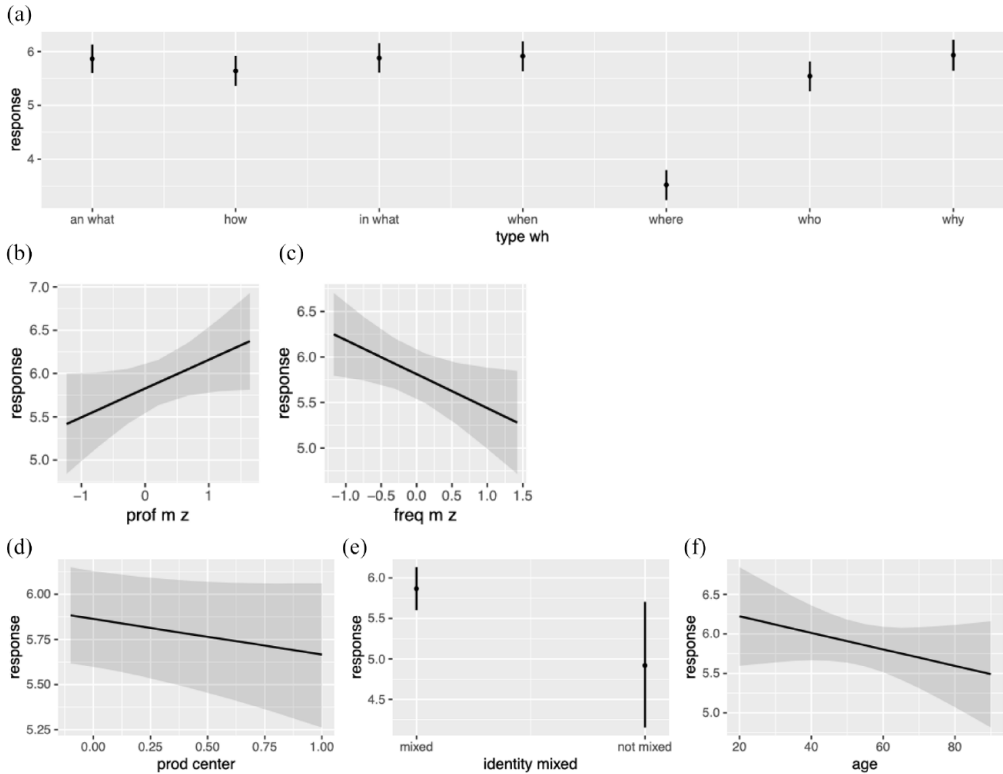
Figure 7. Plots showing relationship between notable predictors and acceptability of non-fronted constructions (predictors: (a) type of phrase, (b) Mandarin proficiency, (c) frequency of Mandarin use, (d) production, (e) hybridity of identity, and (f) age).

Ratings for non-fronted constructions

As for the general model fitting the non-fronted data, the results indicated that the type of phrase had effects on the ratings (Figure 7(a)). Participants rated all non-fronted constructions relatively high except for those involving *where*. Out of all the proficiency and frequency factors, only proficiency in Mandarin ($\beta_{median}=0.33$, $SD=0.22$, $pd=0.93$) and frequency of Mandarin use ($\beta_{median}=-0.38$, $SD=0.21$, $pd=0.96$) had a high likelihood of influencing the ratings for non-fronted constructions. Those proficient in Mandarin were more likely to rate non-fronted constructions high (Figure 7(b)) and those who use Mandarin frequently tend to rate them low (Figure 7(c)), highlighting the mediating effect of dimension of language interaction (i.e., knowledge, use) on the relationship between Mandarin and non-fronted construction acceptability ratings.

Age and ethnic identity seem to play a role in conditioning the variation as well. Younger speakers, who tend to be more exposed to fronting languages Tagalog and English, tend to rate non-fronted constructions higher compared with older speakers ($\beta_{median}=-0.01$, $SD=0.01$, $pd=0.92$). Those who identify as Chinese Filipino or Filipino-Chinese were more likely to rate non-fronted constructions higher than those who do not ($\beta_{median}=0.94$, $SD=0.47$, $pd=0.98$).

The effects observed in the general model also apply to the argument model, where factors such as Mandarin proficiency, frequency of Mandarin use, identity, and age similarly influence the acceptability of non-fronted constructions in the same direction.

Table 10. Summary of effects: acceptability ratings for non-fronted adjunct *wh*-questions (+, positive relationship, − negative relationship, ~ effect not notable, cells/shading in gray indicate that the direction is in line with my hypotheses).

Dimension	Language	*Why*	*How*	*When*	*Where*
Proficiency	Hokkien	+	−	−	−
	Mandarin	+	+	+	−
	Tagalog	~	+	~	−
	English	−	+	~	+
Frequency	Hokkien	−	+	~	+
	Mandarin	−	−	−	−
	Tagalog	+	+	~	−
	English	~	~	~	+

In the adjunct model, all proficiency and frequency factors interact with type to influence the ratings, as presented in Table 10. However, the overall effects only partially support my initial hypotheses (Table 10). Some factors align with expectations and make individuals more likely to accept non-fronted *why constructions*, such as high Hokkien proficiency ($\beta_{median}=0.37$, $SD=0.26$, $pd=0.92$), low English proficiency ($\beta_{median}=-0.23$, $SD=0.25$, $pd=0.81$), and low frequency of Mandarin use ($\beta_{median}=0.17$, $SD=0.18$, $pd=0.83$). In addition, it is unsurprising that certain factors make individuals more likely to accept non-fronted questions involving *how, when*, and *where*. High Mandarin proficiency has been found to increase ratings for non-fronted *how* and *when* constructions ($\beta_{median}=0.32$, $SD=0.19$, $pd=0.95$), while a high frequency of Hokkien use encourages higher ratings for non-fronted *how* and *where* constructions ($\beta_{median}=-0.21$, $SD=0.23$, $pd=0.82$). Furthermore, a low frequency of use in the fronting language Tagalog appears to promote higher ratings for non-fronting constructions involving *where* ($\beta_{median}=0.31$, $SD=0.19$, $pd=0.95$). However, several other effects do not align with my hypotheses (cells not shaded in Table 10). For example, besides high Hokkien proficiency, high Mandarin proficiency was also found to encourage acceptance of non-fronted *why constructions*, which is surprising as Mandarin does not typically exhibit this construction.

Age has an influence on the ratings of non-fronted constructions involving adjuncts ($\beta_{median}=-0.02$, $SD=0.01$, $pd=0.96$). Younger participants tend to give higher ratings compared with older participants. In addition, age interacts with sex ($\beta_{median}=0.01$, $SD=0.01$, $pd=0.85$). Among younger participants, men seem to be leading in providing high ratings for non-fronted constructions in the context of adjunct *wh*-words. If we consider age as a proxy for language exposure, the results do not align with my initial hypotheses. Participants exposed to fronting languages like English and Tagalog behaved differently, rating non-fronted constructions higher than those with less exposure. However, as discussed earlier, it is important to note that age is not solely correlated with language exposure, as other age-specific factors might also come into play.

Again, we also find the role of identity in *wh*-question acceptance ($\beta_{median}=1.05$, $SD=0.52$, $pd=0.98$). Similar to the argument and general models, those who identify as Chinese Filipino or Filipino-Chinese were more likely to rate non-fronted constructions higher than those who do not. Their intuitions about what is considered acceptable or not appear to be clearer and more consistent than those who do not identify as such.

Synthesis

Collectively, the results of the scale-rating models reveal the influence of multilingual contact and sociolinguistic factors on *wh*-question acceptability, representing another aspect of Lánnang-uè

wh-question grammar. Contact with languages possessing distinct question systems from Lánnang-uè appears to impact how speakers evaluate Lánnang-uè *wh*-question constructions in diverse and sometimes unpredictable manners. In some cases, contact with other languages aligns with the expected direction, providing potential evidence for structural transfer and the partial dependence of Lánnang-uè on its source languages. However, there are cases where it does not.

As an illustration, factors such as young age (indicative of more exposure to English and Tagalog and less to Hokkien), high self-reported English proficiency, and a higher self-reported frequency of Tagalog use appear to influence Lánnang-uè's *why*-fronting system, but in a manner opposite to what was observed in the production experiment. Rather than reinforcing system-wide transfer/fronting, contact with English and Tagalog in the form of exposure, frequency, and proficiency seems to have led speakers to disfavor (at least some) fronted constructions, contradicting initial expectations. One possible explanation involves Lánnang-uè perception and conscious linguistic dissimilation between Hokkien and English/Tagalog. Multilingual speakers might have attempted to accentuate the differences among the languages in their perceived linguistic repertoire (Hokkien, Tagalog, English). During the experiment, many participants, when explicitly asked to justify their low ratings for fronting, expressed that the fronted stimuli appeared too similar to Tagalog/English fronting, which they believed did not belong in Lánnang-uè (essentially Hokkien).[4] They strongly held the view that Lánnang-uè is a non-fronting language in general. Their perception appears to account for the unexpected trends observed in the scale-rating data. Another plausible explanation, as previously discussed, is that the patterns contradicting my hypotheses could be indicative of the (partial) linguistic independence of Lánnang-uè from its source languages.

Using acceptability judgments on minoritized languages with no historical standardization has been known to yield "unreliable" results (Benmamoun et al., 2013; Lipski, 2020). The findings from the scale-rating experiment, along with metalinguistic comments, support this notion: some speakers appeared to perceive the stimuli as Hokkien, even though the task explicitly called for judgments of Lánnang-uè. Therefore, the collective findings related to acceptability might illustrate the influence of multilingualism on conscious judgments of Lánnang-uè, not as a distinct mixed language, but as a language that is still perceived by many as a variant of (broken) Hokkien.

Lánnang-uè *wh*-questions "in flux"

Based on the findings, a question arises: Is Lánnang-uè, like other mixed languages, less stable when used alongside its source languages? To address this question, it is crucial to assess the stability of Lánnang-uè, considering the historical multilingualism of the Lannangs (Gonzales, 2017). We should clarify whether there was a point in Lannang sociohistory when Lánnang-uè served as the dominant language of the community. The answer, using an inferred chronology based on age differences (Sankoff, 2006), is affirmative. According to a language survey conducted in 2017 (Gonzales, 2017, 2023b), Lánnang-uè was the predominant language among the Lannangs from the 1950s (represented by the 70s group) to the 1980s (40s group), with its usage peaking in the 1980s (Figure 8).

Chuaunsu's (1989, p. 95) community-wide survey three decades ago corroborate this: she found prevalent Hokkien-Tagalog "code-switching variety" or Lánnang-uè use in the home and community domains by 21–45-year-old Lannangs, now in their 50s to 70s. She also reported robust, expanded Lánnang-uè use by 13–18-year-old Lannangs (now in their 40s) in all communication domains. Both surveys converge on the suggestion that Lánnang-uè and its (*wh*-question) grammar had conventionalized at least three decades ago, perhaps even earlier if we adopt the apparent time model.
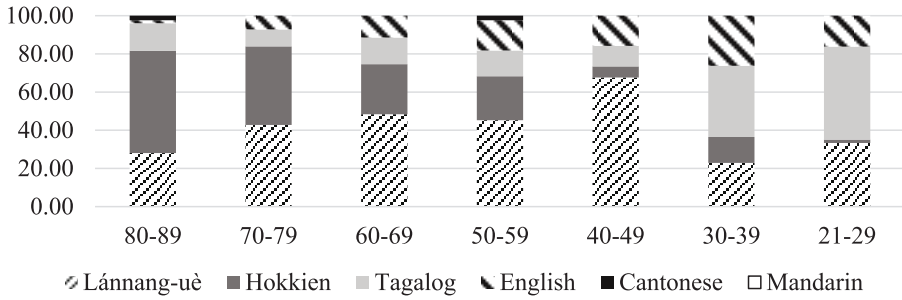
Figure 8. Self-reported dominant languages used with various interlocutors among 65 Lannangs.
Source. Adapted from Gonzales (2017).

Having established Lánnang-uè's early stability, I now return to the question of whether increased multilingualism or contact has affected it. To answer this, I examined for potential effects of language exposure, frequency, and proficiency on Lánnang-uè's *wh*-question system, by looking at both speaker's unconscious use and conscious assessment or acceptability of the *wh*-question constructions. The results in both experiments suggest that multilingual contact indeed did have an impact on Lánnang-uè's question system. While Lánnang-uè generally has a conventionalized *wh*-question system that systematically fronts *why*-questions and does not front non-*why questions*, this convention seems to be partially affected by its speakers' knowledge of the source languages, frequency of source language use, and arguably exposure to the source languages. In both language use and acceptability, speakers who had increased contact with the source languages generally exhibit tendencies to deviate from the said system.

Echoing Lipski's (2020) findings on the perceptual instability of Media Lengua in a situation of Spanish-Quichua-Media Lengua trilingualism, this case study shows the potential of extended community multilingualism (Hokkien, Tagalog, English, Mandarin, and Lánnang-uè) to "destabilize" the grammar of a mixed language. The fluctuation of the *wh*-question system (green bars) across time is particularly salient if we treat age variation as a proxy for non-existent diachronic data.

If the community had Lánnang-uè as the only language in their repertoire and other potential sociolinguistic causes for language maintenance and change such as deliberate change were ruled out (Thomason, 2007), speakers might more uniformly regard *why*-fronting and non-*why* non-fronting as more conventional (i.e., red bars would be less salient in Figure 9)—both in use and acceptability, as they would have no other reference languages to compare Lánnang-uè with. But in this case, the Lánnang-uè speakers were continually exposed to languages that noticeably had a "destabilizing" impact on Lánnang-uè conventions for the most part. From a perspective of variation as instability, mixed languages like Lánnang-uè can then be interpreted to be less stable in a setting where it is used alongside its sources, partially supporting the widely held assertion that symbiotic mixed languages are less stable (Lipski, 2020; Smith, 2000).

However, as can be gleaned from the findings, exclusively attributing the "instability" of Lánnang-uè's *wh*-question system to its symbiotic nature (i.e., constant contact with its source languages in terms of exposure, frequency of use, proficiency) can be problematic. The symbiotic-non-symbiotic dichotomy appears to be too simplistic, as evident in the varied effects of contact on the Lánnang-uè *wh*-question system: there are cases when contact with its source languages clearly reinforces the system, other cases when contact introduces innovations to the system, and cases when contact had no notable effect on the system. Furthermore, we have also seen how sociolinguistic factors like hybrid identity and pride in Lánnang-uè interact with these factors to influence
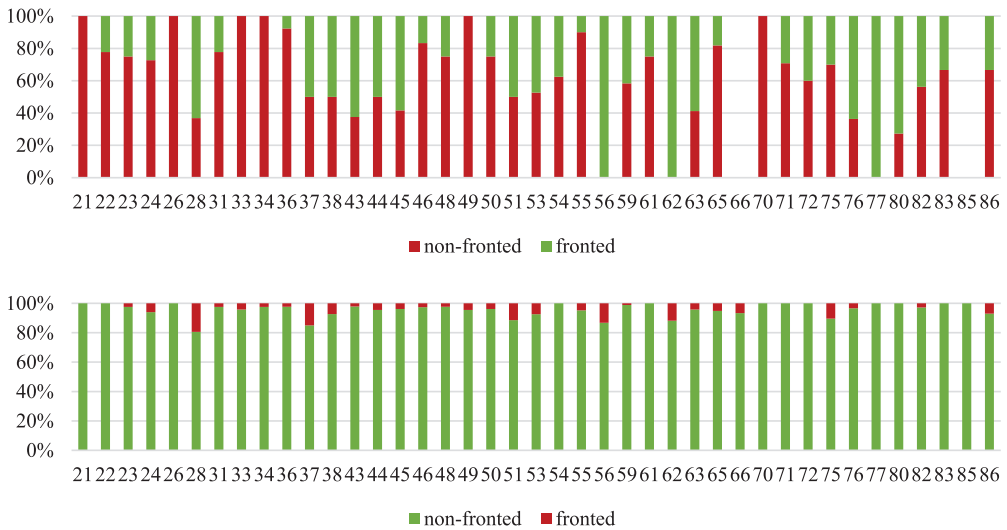
Figure 9. Proportion of fronted and non-fronted constructions by age (top = *why*, bottom = non-*why*, red = less common variant, green = more common variant).

the *wh*-question conventions in Lánnang-uè. The results of this study show that variation exists even with the class of symbiotic mixed languages, and that this variation is not exclusively conditioned by knowledge, exposure, and proficiency in the source languages, but also other factors. The findings point to a need for nuanced characterizations of contact varieties.

Divergent effects of multilingual contact on Lánnang-uè *wh*-questions

Lánnang-uè *wh*-question system was likely affected by multilingualism, along with other factors (e.g., identity). But how exactly was the system affected? What is the nature of the effect? As both production and acceptability results showed, the effects of language proficiency, frequency, and exposure (among others) are numerous and far from homogeneous; they vary depending on the source language, *wh*-phrase type, and degree of consciousness (see Table 11).

For example, special *why*-fronting Mandarin can reinforce the *why*-fronting convention of Lánnang-uè, while system-wide fronting language English seem to alter the *why*-fronting convention to include fronting in other *wh*-phrases. More examples can be found in Table 11. Noticeably, from a putative diachronic view, the different source languages had different effects on Lánnang-uè's system, simultaneously reinforcing and destabilizing the mixed language's system. The different source languages could also be interpreted as triggers for innovative contact-induced change in Lánnang-uè.

The source languages can also have heterogeneous effects depending on the *wh*-phrase type. In production, for example, high English proficiency only increased the speakers' likelihood to front in argument *wh*-phrase questions but not in adjunct ones. Decreased frequency in Mandarin use and increased proficiency in Mandarin seemed to have encouraged fronting in argument *what-* and *who*-questions, but not in other questions. The asymmetrical effect shows that a source language does not necessarily have to affect a linguistic system wholesale, as in the case of Chinese languages on the aspectual system of Colloquial Singapore English (Bao, 2005), but can selectively yet systematically condition or influence components of a linguistic system differently.

Table 11. General summary of factors that encourage general fronting, conventional fronted placement of *why* and conventional *in situ* placement of *how, when, where, what*, and *who*: production and acceptability (− = decreased, + = increased).

Factors that encourage . . .	Production (unconscious)	Acceptability (conscious)
General fronting	• Hokkien frequency (−) • Tagalog frequency (+) • Age (old) • Gender (female) • Acceptability (+)	• Hokkien proficiency (+) • Hokkien frequency (−) • Tagalog frequency (−) • English proficiency (−) • English frequency (+) • Age: Sex (young male) • Identity (Chinese Filipino/Filipino-Chinese)
Conventional fronted placement of *why*	• Hokkien proficiency (+) • Hokkien frequency (+) • Tagalog frequency (+) • English frequency (+) • Mandarin proficiency (+) • Identity (not Chinese Filipino/Filipino-Chinese) • Pride in Lánnang–uè (−) • Age: Sex (older women)	• Hokkien proficiency (+) • Hokkien frequency (−) • Tagalog proficiency (−) • Tagalog frequency (+) • English proficiency (−) • English frequency (+) • Mandarin frequency (+) • Age: Sex (younger women) • Identity (Chinese Filipino/Filipino-Chinese)
Conventional *in situ* placement of *how, when, where, what*, and *who*	*what, who* • Hokkien proficiency (+) • English proficiency (−) • Mandarin frequency (+) • Mandarin proficiency (−) • Pride in Lánnang–uè (−) • Acceptability (−) *how, when, where* • Mandarin frequency (−) *when* • Hokkien proficiency (+) • Age Sex (not older women)	*what, who* • Hokkien proficiency (+) • Hokkien frequency (−) • Tagalog frequency (−) • English proficiency (−) • English frequency (+) • Mandarin proficiency (+) • Mandarin frequency (−) • Age: (young) • Identity (Chinese Filipino/Filipino-Chinese) • Age: Sex (young male) *how, when, where* • Hokkien proficiency (+) • Hokkien frequency (+) • Tagalog proficiency (−) • Tagalog frequency (+) • English proficiency (+) • English frequency (−) • Mandarin proficiency (+) • Mandarin frequency (−) • Age: (young) • Identity (Chinese Filipino/Filipino-Chinese) *how, where* • Mandarin proficiency (+) • Hokkien frequency (+) *how* • Sex (Male) *when* • Sex (Female) • Mandarin proficiency (+) *when and where* • Age: Sex (not young men)

Degree of consciousness is another possible factor of effect variation apart from source language and phrase type, reflecting stylistic variation reminiscent of the attention-paid-to-speech model (Labov, 2002; Meyerhoff, 2018). Knowledge of source languages seemed to have affected participants differently in situations when they were not aware that their Lánnang-uè was being monitored (the elicitation task) compared with those where they were (scale-rating task). In the production task, speakers were directly asked to participate in the game after I introduced myself in Lánnang-uè. As such, they were not actively aware. Contrastively, in the acceptability experiment, speakers were explicitly asked to rate the Lánnang-uè of the speaker—however they perceive it—on a scale of "good-soundedness" or acceptability. Both tasks arguably differ in speakers' awareness, and the effects of multilingualism on their responses in both tasks are notably divergent. For example, English proficiency does not play a notable role in encouraging general fronting when speakers are not aware but have significant effects when they are. Those with increased English proficiency rate fronted constructions lower than those who do not have such proficiency. Likewise, frequency of Tagalog use increases fronting rates when speakers are not aware, but decreases such rates when speakers are aware of language. Finally, heightened exposure to Tagalog and English (as indicated by age) appears to make participants more inclined to give lower ratings to fronted constructions when they are aware of the task. However, this effect is not present in the production experiment when they are unaware of language use.

When paired with metalinguistic comments that equate Lánnang-uè to Hokkien, it appears that participants with heightened awareness tended to perceive Lánnang-uè as (broken) Hokkien and actively utilized their knowledge of English and Tagalog to make judgments on it. In some cases, they even corrected the stimuli. This stands in contrast to the production scenario, where speakers did not actively conflate Lánnang-uè and Hokkien, nor did they consciously draw on their knowledge of the source languages to influence their production.

The findings suggest that Lánnang-uè's *wh*-question system will be likely to be perceived as belonging to Hokkien if speakers are constantly in contact with the source languages and are conscious of language use. Speakers will actively "Hokkien-ify" non-Hokkien Lánnang-uè elements using their knowledge of Tagalog and English, inevitably changing the language. At the same time, however, Lánnang-uè speakers are often not conscious of language use in daily conversations, allowing the source languages to selectively interact with Lánnang-uè's question grammar.

Conclusion

This study aimed to explore the impact of multilingual contact on the system of a mixed language. By examining Lánnang-uè and its *wh*-question system as a case study, the findings revealed several and, in certain instances, divergent effects of the source languages and sociolinguistic factors on the language's notably variable *wh*-question system. First, the source languages of Lánnang-uè had variable effects on the *wh*-question grammar of the language; second, the source languages of Lánnang-uè have been observed not to influence the question system across-the-board, but instead systematically affect components of the question system differently, interacting with sociolinguistic factors like attitudes and identity. By exploring the variability partially induced by community multilingualism, this study contributes to the scarce amount of mixed language work that treats variation as a foundational feature, not something to be downplayed. The findings that align with my hypotheses of language transfer are consistent with the literature on symbiotic mixed languages, assuming Lánnang-uè belongs to this category of contact languages. However, the findings that contradict these hypotheses, combined with evidence of sociolinguistically conditioned variation—an inherent characteristic of language—provide additional support for Lánnang-uè's independence and status as a distinct language, even though this is contested by some of its speakers. The utilization of the

experimental method in this study allows scholars to make observations and draw inferences about the development of languages like Lánnang-uè, which have limited historical documentation.

In conclusion, it is evident that mixed languages are not unique or exceptional languages. Similar to other standard languages, they are characterized by variation influenced by sociolinguistic factors. In addition, they undergo processes of change and innovation, just like any other language. However, these processes seem to accelerate in environments where the mixed languages are in constant contact with their source languages, as substantiated by the some of the impacts of Hokkien, Tagalog, English, and Mandarin on the *wh*-question system of Lánnang-uè. In this sense, the study corroborates Lipski's (2020) work and the widely held belief that mixed languages are indeed more stable in situations where they have limited to no contact with the source languages. However, this paper goes an extra step to show this "instability" or variability is not always a consequence of transfer induced by proficiency in, frequency of, and exposure to the source languages. It demonstrates that when the source languages influence the stability or development of the mixed language, the effects can be diverse, encompassing aspects such as identity processes, language attitudes, structural transfer, and/or other sociolinguistic innovations. The nature of these effects is dependent on numerous factors, including the sociohistorical context in which the mixed language(s), the source languages, and their speakers interact.

Declaration of conflicting interests

Funding

ORCID iD

Wilkinson Daniel Wong Gonzales 🆔 https://orcid.org/0000-0001-6073-256X

Notes

1. These consist of lexicon from a 219-word Swadesh list used in Gonzales and Starr (2020).
2. I focus on questions with object *wh*-phrases to highlight the distinctiveness of Lánnang-uè.
3. With the exception of the age and sex interaction term, I attempted to add the interaction terms to the general model, but doing so exponentially increased the computing and processing cost of the data, making it impossible to derive results.
4. Recall that speakers perceive Lánnang-uè as a Hokkien variety.

References

Aldridge, E. (2002). Nominalization and wh-movement in Seediq and Tagalog. *Language and Linguistics*, *3*(2), 393–426.

Ang See, T. (1990). *The Chinese in the Philippines: Problems and perspectives* (Vol.1). Kaisa Para Sa Kaunlaran.

Bakker, P. (2003). Mixed languages as autonomous systems. In Y. Matras (Ed.), *The mixed language debate: Theoretical and empirical advances* (pp. 107–150). Mouton de Gruyter.

Bao, Z. (2005). The aspectual system of Singapore English and the systemic substratist explanation. *Journal of Linguistics*, *41*(2), 237–267. https://doi.org/10.1017/S0022226705003269

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, *39*(3–4), 129–181.

Biblical Seminary of the Philippines. (2011). *The Amoy Audio Bible Project*. Biblical Seminary of the Philippines. https://bsop.edu.ph/the-amoy-bible-project/

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Cheung, C.-H. (2014). Wh-fronting and the left periphery in Mandarin. *Journal of East Asian Linguistics*, *23*, 393–431.

Chuaunsu, R. (1989). *A speech communication profile of three generations of Filipino-Chinese in Metro Manila: Their use of English, Pilipino and Chinese languages in different domains, role-relationships, speech situations and functions* [Master's thesis]. University of the Philippines.

Doeppers, D. (1986). Destination, selection and turnover among Chinese migrants to Philippine cities in the nineteenth century. *Journal of Historical Geography*, *12*(4), 235–260.

Gonzales, W. D. W. (2016). Trilingual code-switching using quantitative lenses: An exploratory study on Hokaglish. *Philippine Journal of Linguistics*, *47*, 106–128.

Gonzales, W. D. W. (2017). Language contact in the Philippines: The history and ecology from a Chinese Filipino perspective: The history and ecology from a Chinese Filipino perspective. *Language Ecology*, *1*(2), 185–212. https://doi.org/10.1075/le.1.2.04gon

Gonzales, W. D. W. (2018). *Philippine Hybrid Hokkien as a postcolonial mixed language: Evidence from nominal derivational affixation mixing* [Master's Thesis]. National University of Singapore.

Gonzales, W. D. W. (2021). Filipino, Chinese, neither, or both? The Lannang identity and its relationship with language. *Language & Communication*, *77*, 5–16.

Gonzales, W. D. W. (2022a). Interactions of Sinitic languages in the Philippines: Sinicization, Filipinization, and Sino-Philippine language creation. In Z. Ye (Ed.), *The Palgrave Handbook of Chinese Language Studies* (pp. 369–408). Springer Nature Singapore. https://doi.org/10.1007/978-981-16-0924-4_31

Gonzales, W. D. W. (2022b). *"Truly a Language of Our Own" A Corpus-Based, Experimental, and Variationist Account of Lánnang-uè in Manila* [Ph.D. dissertation]. University of Michigan.

Gonzales, W. D. W. (2023a). Spread, stability, and sociolinguistic variation in multilingual practices: The case of Lánnang-uè. *International Journal of Multilingualism*. https://doi.org/10.1080/14790718.2023.2199998

Gonzales, W. D. W. (2023b). Variability in clusters and continuums: The sociolinguistic situation of the Manila Lannangs in the 2010s. *Asia-Pacific Language Variation*, *9*(1), 83–124. https://doi.org/10.1075/aplv.22009.gon

Gonzales, W. D. W., & Starr, R. L. (2020). Vowel system or vowel systems? Variation in the monophthongs of Philippine Hybrid Hokkien in Manila. *Journal of Pidgin and Creole Languages*, *35*(2), 253–292.

Grafmiller, J., Szmrecsanyi, B., & Hinrichs, L. (2018). Restricting the restrictive relativizer. *Corpus Linguistics and Linguistic Theory*, *14*(2), 309–355. https://doi.org/10.1515/cllt-2016-0015

Jin, D. (2014). Contradiction, why-questions and complex islands in Chinese. *University of Pennsylvania Working Papers in Linguistics*, *21*(1), 1–8.

Labov, W. (2002). The anatomy of style-shifting. In P. Eckert & J. R. Rickford (Eds.), *Style and sociolinguistic variation* (1st ed., pp. 85–108). Cambridge University Press. https://doi.org/10.1017/CBO9780511613258.006

Levshina, N. (2016). When variables align: A Bayesian multinomial mixed-effects model of English permissive constructions. *Cognitive Linguistics*, *27*(2), 235–268. https://doi.org/10.1515/cog-2015-0054

Lipski, J. (2020). Reconstructing the life-cycle of a mixed language: An exploration of Ecuadoran Media Lengua. *International Journal of Bilingualism*, *24*(2), 410–436.

MacKenzie, L. (2020). Comparing constraints on contraction using Bayesian regression modeling. *Frontiers in Artificial Intelligence*, *3*, Article 58. https://doi.org/10.3389/frai.2020.00058

Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, *10*, Article 2767. https://doi.org/10.3389/fpsyg.2019.02767

Matras, Y., & Bakker, P. (Eds.). (2003). *The mixed language debate: Theoretical and empirical advances*. Mouton de Gruyter.

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). CRC Press, Taylor and Francis.

Meakins, F., & Stewart, J. (2022). Mixed Languages. In S. Mufwene & A. M. Escobar (Eds.), *The Cambridge handbook of language contact* (1st ed., pp. 310–343). Cambridge University Press. https://doi.org/10.1017/9781316796146.016

Meyerhoff, M. (2018). *Introducing sociolinguistics* (3rd ed.). Taylor and Francis.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Poa, D. (2004). From quadrilingual to bilingual: On the multilingual teaching in the Chinese schools in the Philippines. In Q. Dai (Ed.), *Bilingual studies* (Vol.2, pp. 96–107). Minzu Chubanshe.

R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org

Sankoff, G. (2006). Age: Apparent time and real time. In K. Brown (Ed.), *The encyclopedia of Language & Linguistics* (2nd ed., pp. 110–116). Elsevier.

Sato, Y. (2013). Wh-questions in colloquial Singapore English: Adaptive traits from vernacular Malay and typological congruence. *Journal of Pidgin and Creole Languages*, *28*(2), 299–322.

Schachter, P., & Otanes, F. T. (1972). *Tagalog reference grammar*. University of California Press.

Smith, N. (2000). Symbiotic mixed languages: A question of terminology. *Bilingualism: Language and Cognition*, *3*(2), 79–99.

Sonderegger, M. (2022). *Regression modeling for linguistic data* (1st ed.). Open Science Framework. https://osf.io/pnumg/

Tan, S. V. (1993). *The education of Chinese in the Philippines and Koreans in Japan*. University of Hong Kong.

Thomason, S. (2007). Language contact and deliberate change. *Journal of Language Contact*, *1*(1), 41–62.

Thomason, S., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Univeristy of California Press.

Uytanlet, J. L. (2014). *The Hybrid Tsinoys: Challenges of hybridity and homogeneity as sociocultural constructs among the Chinese in the Philippines* [PhD dissertation]. Ashbury Theological Seminary.

Vasishth, S., & Nicenboim, B. (2016). Statistical methods for linguistic research: Foundational ideas—Part I: Statistical methods for linguistics—Part I. *Language and Linguistics Compass*, *10*(8), 349–369. https://doi.org/10.1111/lnc3.12201

Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics*, *71*, 147–161. https://doi.org/10.1016/j.wocn.2018.07.008

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved R^ for assessing convergence of MCMC (with Discussion). *Bayesian Analysis*, *16*(2), 667–718. https://doi.org/10.1214/20-BA1221

Author biography

Wilkinson Daniel Wong Gonzales is an Assistant Professor of Applied English Linguistics at The Chinese University of Hong Kong. His research focuses on language variation and change in multilingual contexts, with a particular interest in the Philippines and wider East Asia. He uses corpus-based, computational, and ethnographic methods to investigate Sino-Philippine and other East Asian language varieties such as Colloquial Singapore English, Philippine English, Hong Kong English, and Lánnang-uè. He is the founder of The Lannang Archives.