

**“Truly a Language of Our Own”
A Corpus-Based, Experimental, and Variationist Account of Lánnang-uè in Manila**

by

Wilkinson Daniel Wong Gonzales

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
In The University of Michigan
2022

Doctoral Committee:

Professor Marlyse Baptista, Co-chair
Professor Sarah G. Thomason, Co-chair
Professor Julie Boland
Professor Andries Coetzee

“Talagà Dân-e Láng-e Uè”

Tsīgē Corpus-Bâsed, Experimentàl, kâp Variationîst na Accoûnt of Lánng-uè ti Manilá

Wilkinson Daniel Wong Gonzales

wdwg@umich.edu

ORCID iD: [0000-0001-6073-256X](https://orcid.org/0000-0001-6073-256X)

© Wilkinson Daniel Wong Gonzales 2022

DEDICATION

... to the Lannang community ...
... *para dân mga Lánnáng* ...

ACKNOWLEDGMENTS

I didn't believe the others when they said it, but it's true! Writing the acknowledgments section has indeed been the most difficult thing I've had to do in my thus-far short academic journey. I spent hours looking for models, hoping to get a sense of what is considered 'suitable' for this strange genre. I tried looking for the most appropriate (and fun!) way to thank everyone in my life who have made this dissertation possible. Towards the end of my research, I realized that my attempts were futile – not because I couldn't find great models, but because there were simply too many, and doing the research has left me in a worse position than when I started. I realized that I should just get right into writing it.

First, I must thank the roughly 140 Lannang participants in Manila who let me visit their homes, hardware stores, churches, temples, warehouses, restaurants, etc. Thank you so much for sharing your time, experiences, stories, and in many cases, food! – this dissertation would literally not have been successful without you. I am especially indebted to my Lannang friends Dr. Juliet and Samson Uytanlet (Biblical Seminary of the Philippines), Cindy Hao (Menukha Center), Ivan Man Dy (Old Manila Walks), Fr. Aristotle Dy (Xavier School), Rev. Tony Cheng (United Evangelical Church of the Philippines), Rev. Michael Lim (United Evangelical Church of Greenhills), who have all taken special interest in my work, so much so that they introduced me to more prospective participants (more data, yay!). *Tosia dîn talagà!* If you read this dissertation, I hope you'll find it enjoyable.

Next on my endless list of people to thank is my dissertation committee: Marlyse and Sally, you have both been with me since day one. You were both the reason why I chose The University of Michigan for my Ph.D. studies. Thank you for being super patient with me, and for believing in me. There were honestly times I thought about giving up, but seeing how much time, effort, and care you gave into guiding me stopped me and became my source of strength. I told myself that if my advisers cared enough to write single-spaced comments ranging from 5 to 20 pages for each version of each chapter (what? I know!), then the work I am doing must be

valuable to them as well. I hope I have made you proud. Andries, I approached you right after I took phonology with you in my second year because I valued how you approached linguistic problems and how constructive your comments were. I hope you did not regret your decision to take me in as one of your advisees (kidding!). Julie, thank you so much for agreeing to be part of my committee – your feedback on the statistical and experimental aspects of my dissertation has been crucial to its success. Thank you all for taking a chance on me, and for being my advocates, academic role models, and mentors. You have all completely shaped and guided my understanding and perspective towards linguistics.

Thanks to the wonderful collaborators I've been lucky to have throughout my Ph.D. journey:

To Mie Hiramoto, Rebecca Starr, Jakob Leimgruber, Zhengdao Ye, Ariane Macalinga Borlongan, Andrew Cheng, Loretta Cheng, and Pocholo Umbal, Ariana Bancu, Felicia Bisnath, Danielle Burgess, Sophie Eakins, Joy Peltier, Moira Saltzman, Yourdanis Sedarous, and Alicia Stevers – thank you for either giving me an opportunity to work on a writing project and/or publish with you. I have greatly benefited from your insights and your perspectives. I hope to continue working with you beyond the doctorate program.

To Pam Beddor, Steve Abney, Emily Atkinson, Will Styler, Savithry Namboodiripad, and Stephen Tobin – thank you for taking an interest in my work, shelling out time to discuss and provide feedback on my research, teaching me computational skills, and/or leading me to resources crucial to the success of my dissertation.

To the Department of Linguistics at the University of the Philippines, Diliman (especially Jem Javier and Elsie Or) – thank you for inviting me to become a research fellow during my data collection summer. Without your help, I would not have been able to train transcribers and recruit a significant number of my participants for my studies.

I had many dear colleagues in my program that I will fondly remember. There are too many to list (insert whole roster of department graduate students here), so I'll focus on individuals in my 'inner circle' (hello, Braj Kachru!):

To Chiawen Lo, Lucy Chiang, and Tzuyun Tung, thank you for the gift of companionship. You certainly made my stay at Michigan feel more like my second home, Taiwan. I will miss our small gatherings. Special thanks to Lucy for surprising me with

Taiwanese delicacies regardless of holidays like Chinese New Year. They certainly made me feel a little less homesick.

To members of YAWL (also known as /mywl/): Yushi Sugimoto, Mathew Kramer, and Laretta Cheng. It is safe to say that I wouldn't have survived my Ph.D. journey without the social support you provided! Yushi, thank you for coining the lambda gesture and for introducing the ghost operator. I am very thankful for your advice regarding the syntactic aspects of the dissertation as well as our pre-pandemic coffee trips. I will also never forget how you invited me to your place when I first landed in the U.S. with nowhere to stay. Mathew – you have an impeccable taste in food and drink. Thanks for all the recommendations (places with food I got to try because of you). A special thank you to you for helping me out with historical linguistics and programming when I was a total noob. Laretta, if Mathew helped me out with historical and programming, you helped me out with phonetics and phonology stuff – I really appreciated how much effort and time you spent explaining concepts that were totally Greek to me (e.g., acoustics, theories about perception, optimality theory). I also appreciated our pre-pandemic meet-ups – coffee shop writing sessions, Scrabble sessions in the department graduate student lounge, etc.

As a group, thank you for our almost monthly YAWL sessions (sometimes holiday-themed, sometimes planned, sometimes spontaneous). There are too many to enumerate in one sitting: Indian dinners at Namaste, Studio-Ghibli-themed puzzle/movie sessions, linguistics-themed game night, Asian fusion BBQ lunch, Dimsum Night, D.I.Y. hotpot, Dance and Champagne party, YAWLsgiving, pumpkin picking at Erwins', etc. They have all been YAWLsome and you guys have been “so fetch” (Mean Girls, 2004). I'm glad to have crossed paths with YAWL (okay, I promise this is the last!).

To my family, thank you for being supportive and for inspiring me with love for academia from a young age. Special thanks to Papá Wilson and Mamá Marissa, who have supported me not only financially when I needed it, but also spiritually with prayers and guidance.

I would also like to thank the Rackham School of Graduate Studies, the Kenneth G. Lieberthal and Richard H. Rogel Center for Chinese Studies, the International Institute at the University of Michigan, as well as the Linguistics department at the University of Michigan for travel grants, research grants, and other financial aid that made it possible for me to complete this

Ph.D. All errors that remain in this dissertation, despite the brilliant guidance I received, are my own.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGMENTS	iii
LIST OF TABLES	xiii
LIST OF FIGURES	xvii
LIST OF APPENDICES	xx
LIST OF ABBREVIATIONS	xxi
ABSTRACT (English)	xxiii
ABSTRÂCT (Lánnang-uè)	xxv
Chapter 1 : Introduction	1
Chapter 2 : General Background	6
2.1 The Lannangs	6
2.2 Lánnang-uè.....	8
2.2.1 Terminology	8
2.2.2 Investigated features in the scholarly literature on Lánnang-uè	9
2.2.3 Domains of use.....	10
2.2.4 Speakers and language background	10
2.2.5 Speakers' perceptions.....	12
2.3 Language and languageness.....	13
2.4 Spread and stability	16
Chapter 3 : Description of Lánnang-uè	18
3.1 Preliminaries	18
3.2 Lexicon.....	18

3.3	Phonetics and phonology	19
3.3.1	Phoneme inventory.....	19
3.3.2	Phonotactics	22
3.3.3	Suprasegmentals.....	24
3.3.3.1	Tone.....	24
3.3.3.2	Stress.....	32
3.3.3.3	Phrase-final lengthening.....	34
3.4	Noun phrases.....	34
3.4.1	The determiner system	35
3.4.2	Classifiers	37
3.4.3	Quantifiers.....	39
3.4.4	Pluralization	41
3.4.5	Numerals	46
3.4.6	Interrogatives.....	51
3.4.7	Adjectivals.....	52
3.4.7.1	Modification	52
3.4.7.2	Comparative constructions	57
3.4.7.3	Superlative constructions.....	60
3.4.7.4	Approximation.....	61
3.4.7.5	Negation and adjectives.....	62
3.4.7.6	Adverbial modification.....	63
3.4.8	The pronominal system	64
3.4.8.1	Personal pronouns.....	64
3.4.8.2	Reflexive pronouns.....	65
3.4.8.3	Idiomatic constructions.....	66
3.4.9	Inflectional affixes and clitics	67
3.4.10	Derivational prefixes	68
3.5	Verb phrases.....	73
3.5.1	Modality	73
3.5.2	Aspect.....	83
3.5.3	The ability suffix -dit.....	96
3.5.4	Causative markers	97
3.5.5	Benefactive markers.....	100
3.5.6	Directional markers	101
3.5.7	Resultative marker.....	102
3.5.8	Negation	103
3.5.9	Adverbial modification	106
3.5.10	The copula	112
3.6	Clause-level constituent order.....	113

3.7	Questions.....	120
3.7.1	Yes-no questions	120
3.7.2	Wh-questions.....	122
3.8	Prepositions	126
3.9	Conjunctions	137
3.9.1	Coordinating conjunctions	137
3.9.2	Subordinating conjunctions.....	140
3.9.3	General patterns & summary.....	152
3.10	Interjections.....	157
3.11	Discourse particles	159
3.12	Summary and concluding remarks.....	172
Chapter 4 : Stress and Tone Features.....		176
4.1	Introduction	176
4.2	The prosody of Lánnang-uè and its source languages: A summary	178
4.3	Hypotheses	179
4.4	Methodology	182
4.4.1	Task	182
4.4.2	Stimuli	182
4.4.3	Datasets and preparation	183
4.4.4	Analytical method	186
4.4.4.1	Descriptive analyses	186
4.4.4.2	Regression analyses.....	186
4.4.4.3	Criteria for hypothesis testing.....	188
4.4.5	Measuring spread and stability.....	189
4.4.6	Participants	191
4.5	Results	192
4.5.1	Lexical stress	192
4.5.1.1	Descriptive analysis.....	192
4.5.1.2	Regression analysis.....	194
4.5.2	Lexical tone	196
4.5.2.1	Descriptive analysis.....	196
4.5.2.2	Regression analysis.....	198
4.5.3	Tone distributional patterns.....	200
4.5.3.1	Descriptive analysis.....	200
4.5.3.2	Regression analysis.....	204
4.6	Discussion	209
4.6.1	Hypotheses 1 and 2: Spread and stability.....	209

4.6.1.1	Possible reinforcers of the stress pattern	210
4.6.1.2	Possible reinforcers of the lexical tone feature.....	211
4.6.1.3	Possible reinforcers of the CVT/CV and CVR-English/Tagalog tone patterns	212
4.6.2	Hypothesis 3: Structured variation (Systematicity)	213
4.6.2.1	Age.....	214
4.6.2.2	Sex	216
4.6.3	Hypothesis 4: Linguistic independence.....	217
4.7	Conclusion	219

Chapter 5 : Conjunction and Preposition Lexical Patterns.....223

5.1	Introduction	223
5.2	The patterns: A summary	225
5.3	Hypotheses	228
5.4	Methodology	231
5.4.1	Approach and dataset preparation	231
5.4.2	Analytical method	234
5.4.2.1	Descriptive analyses	234
5.4.2.2	Regression analyses.....	235
5.4.2.3	Criteria for hypothesis testing.....	236
5.4.3	Measuring spread and stability.....	237
5.4.4	Speakers	238
5.5	Results	239
5.5.1	The conjunction lexical distributional pattern.....	239
5.5.1.1	Descriptive analyses	239
5.5.1.2	Regression analysis.....	242
5.5.2	The preposition lexical distributional pattern.....	246
5.5.2.1	Descriptive analyses	246
5.5.2.2	Regression analysis.....	248
5.6	Discussion	252
5.6.1	Hypotheses 1 and 2: Spread and stability.....	252
5.6.1.1	Possible reinforcers of the distributional patterns (general).....	253
5.6.1.2	Possible explanations for the (specific) nature of the distributional patterns.....	257
5.6.2	Hypothesis 3: Structured variation (Systematicity)	260
5.6.2.1	General patterns	260
5.6.2.2	Source-language-specific patterns.....	264
5.6.3	Hypothesis 4: Linguistic independence.....	266
5.6.4	Synthesis.....	268
5.7	Conclusion	269

Chapter 6 : The *Wh*-phrase Position Distributional Pattern272

6.1	Introduction	272
6.2	Wh-phrase position in the source languages of Lánnang-uè	273
6.3	The wh-phrase position distributional pattern in Lánnang-uè	277
6.4	Hypotheses	279
6.5	Methodology	281
6.5.1	Approach	281
6.5.2	Production experiment	282
6.5.2.1	Design and stimuli	282
6.5.2.2	Dataset and preparation	286
6.5.3	Acceptability experiment	287
6.5.3.1	Design and stimuli	287
6.5.3.2	Dataset and preparation	290
6.5.4	Analytical method	291
6.5.4.1	Descriptive analyses	291
6.5.4.2	Regression analyses	291
6.5.4.3	Criteria for hypothesis testing.....	294
6.5.5	Measuring spread and stability.....	295
6.5.6	Participants	296
6.6	Results	296
6.6.1	Production	296
6.6.1.1	Descriptive analyses	296
6.6.1.2	Regression analyses.....	299
6.6.2	Acceptability	302
6.6.2.1	Descriptive analyses	302
6.6.2.2	Regression analysis.....	307
6.7	Discussion	312
6.7.1	Hypotheses 1 & 2: Spread and stability	312
6.7.2	Hypothesis 3: Structured variation (Systematicity)	313
6.7.2.1	General patterns	313
6.7.2.2	Type-specific patterns.....	314
6.7.2.3	Younger speakers and pattern-non-conforming sentence-medial <i>why</i> constructions	315
6.7.2.4	Age effect asymmetry.....	317
6.7.3	Hypothesis 4: Linguistic independence.....	318
6.7.3.1	Testing the hypothesis using results from a linguistic transfer perspective	319
6.7.3.2	Testing the hypothesis using results from an agentive perspective.....	319
6.8	Conclusion	322

Chapter 7 : Epilogue	325
7.1 Brief review.....	325
7.2 Research questions revisited	327
7.2.1 Lánnang-uè in the cline of languageness	327
7.2.2 Lánnang-uè in the constellation of contact languages	328
7.2.2.1 Lánnang-uè as a mixed language?.....	329
7.2.2.2 Lánnang-uè as a Hokkien variety?	331
7.2.2.3 Lánnang-uè as a novel contact language	337
7.3 Limitations and future directions	338
7.4 Final remarks.....	340
APPENDICES	342
REFERENCES	371

LIST OF TABLES

Table 1. Consonants of Lánnang-uè	20
Table 2. Diphthongs in Lánnang-uè.....	22
Table 3. Syllable permutations found in Lánnang-uè with examples.....	23
Table 4. Tones in Lánnang-uè	25
Table 5. Tones in Mandarin and Mandarin-sourced words in Lánnang-uè (words are in Pinyin orthography).....	32
Table 6. The classifiers of Lánnang-uè.....	38
Table 7. Personal pronouns.....	64
Table 8. Derivational prefixes in the noun phrase domain	73
Table 9. Modals in Lánnang-uè	82
Table 10. Aspect markers in Lánnang-uè	95
Table 11. Negative markers in Lánnang-uè.....	106
Table 12. List of the most common verb phrase adverbs, with origin and distribution	111
Table 13. Frequency distribution of variants for selected adverbs	112
Table 14. Clause-level constituent orders	120
Table 15. The prepositions of Lánnang-uè	134
Table 16. The conjunctions of Lánnang-uè	154
Table 17. Common interjections in Lánnang-uè (with linguistic source)	158
Table 18. Discourse particles.....	170
Table 19. Distribution of stimuli (prosody)	183
Table 20. Linear regression results for syllable duration (observations = 4,786, conditional $R^2 = 0.694$, random intercepts for participant and item).....	195
Table 21. Frequency of Lánnang-uè words that did not have lexical tone, by source language	198

Table 22. Generalized linear mixed-effects regression (with logistic link function) results – likelihood to use lexical tone in a Lánnang-uè word (observations = 3,294, R^2 Tjur = 0.16, random intercepts for participant included).....	199
Table 23. Frequency syllables that did not adhere to the CVT/CV tone pattern, by syllable structure (grave \grave{a} = falling tone, circumflex \hat{a} = high tone).....	201
Table 24. Frequency of syllables that did not adhere to the CVR-Tagalog/CVR-English tone pattern, by source language (grave \grave{a} = falling tone, circumflex \hat{a} = high tone).....	203
Table 25. Linear regression results for final syllable pitch slope (observations = 2,391, conditional R^2 = 0.41, random intercepts for participant and item).	205
Table 26. Function-based classes of conjunctions in Lánnang-uè, criteria, most frequently used tokens, and source language	225
Table 27. Function-based classes of prepositions in Lánnang-uè, criteria, most frequently used tokens, and source language	227
Table 28. distributional pattern of speakers by self-reported sex and age group	238
Table 29. Distribution of conjunction tokens that did not conform to the conjunction distributional pattern, by class and source language.....	241
Table 30. Regression results – adherence to the conjunction distributional pattern (observations = 15,769, R^2 = 0.209, random intercepts for speaker).	242
Table 31. Distribution of preposition tokens that did not conform to the preposition distributional pattern, by class and source language	248
Table 32. Regression results – adherence to the preposition distributional pattern (observations = 8,134, R^2 = 0.08, random intercepts for speaker). Reference levels are highlighted in boldface; in the p -values column, statistically significant values are in boldface. .	248
Table 33. Associations between tokens that did not conform to distributional pattern and age/sex groups, by word class and source language.....	265
Table 34. Associations between tokens that did not conform to distributional pattern and proficiency, by word class and source language.....	267
Table 35. Blocks, condition, and stimuli for the <i>wh</i> -question production experiment.....	286
Table 36. Conditions, and items for the <i>wh</i> -question acceptability experiment.....	289
Table 37. Sample item set for argument <i>wh</i> -questions. Translation: ‘What/who will the visitor stab?’	289

Table 38. Sample item set for adjunct <i>wh</i> -questions. Translation: ‘{Why/how/when/where} will the maid carry the bed?’	290
Table 39. Participant matrix (<i>wh</i> -questions).....	296
Table 40. Distribution of <i>wh</i> -phrases (in the questions) that did not conform to the position distributional pattern, by type	298
Table 41. Linear regression results for likelihood to put the <i>wh</i> -phrase sentence-initially (observations = 3,163, conditional $R^2 = 0.427$, random intercepts for participant).....	300
Table 42. Linear regression results for likelihood to put the <i>why</i> -phrase sentence-initially (observations = 390, conditional $R^2 = 0.469$, random intercepts for participant); in the Predictors column, reference levels are highlighted in bold; in the <i>p</i> -values column, statistically significant values are in bold.	301
Table 43. Linear regression results for likelihood to put the how/when/where/who/what-phrase sentence-initially (observations = 2,770, conditional $R^2 = 0.302$, random intercepts for participant)	302
Table 44. Distribution of ratings that did not conform to the position distributional pattern, by type (cell with highest percentage highlighted).....	304
Table 45. Spread scores by position and <i>wh</i> -phrase type (highlighted cells indicate low degrees of spread)	305
Table 46. Intra- and interspeaker variation scores by position and <i>wh</i> -phrase type (highlighted cells indicate high degrees of inconsistency).....	307
Table 47. Linear regression results – likelihood to adhere to <i>wh</i> -question distributional pattern based on acceptability judgments (observations = 6,048, conditional $R^2 = 0.05$, random intercepts for participant).....	307
Table 48. Linear regression results – ratings of sentence-initial constructions featuring <i>why</i> -phrases (observations = 360, conditional $R^2 = 0.274$, random intercepts for participant and item)	308
Table 49. Linear regression results – ratings of sentence-medial constructions featuring <i>why</i> -phrases (observations = 360, conditional $R^2 = 0.191$, random intercepts for participant and item)	309

Table 50. Linear regression results – ratings of sentence-initial constructions featuring <i>how/when/where/object who/object what</i> -phrases (observations = 2,667, conditional $R^2 = 0.106$, random intercepts for participant and item)	310
Table 51. Linear regression results – ratings of constructions featuring sentence-medial <i>how-/when-/where</i> -phrases and sentence-final object <i>who-/what</i> -phrases (observations = 2,661, conditional $R^2 = 0.07$, random intercepts for participant and item)	311
Table 52. Spread scores of conjunctions by class - <i>Speaker (with)</i> : speakers who derived at least one conjunction from that class from the expected language; <i>Speakers (all)</i> : speakers who had at least one conjunction from that class	365
Table 53. Mean intraspeaker consistency scores and interspeaker pattern inconsistency scores (conjunctions) by class.....	366
Table 54. Regression results – likelihood to use Hokkien-derived conjunctions (observations = 15,901, $R^2 = 0.632$, random intercepts for speaker)	367
Table 55. Regression results – likelihood to use Tagalog-derived conjunctions (observations = 15,901, $R^2 = 0.447$, no random effects)	367
Table 56. Regression results – likelihood to use English-derived conjunctions (observations = 15,901, $R^2 = 0.038$, no random effects)	368
Table 57. Spread scores of prepositions by class - <i>Speaker (with)</i> : speakers who derived at least one preposition from that class from the expected language; <i>Speakers (all)</i> : speakers who had at least one preposition from that class	368
Table 58. Mean intraspeaker consistency scores and interspeaker pattern inconsistency scores (prepositions) by class.....	369
Table 59. Regression results – likelihood to use Hokkien-derived prepositions (observations = 8,134, $R^2 = 0.506$, random intercepts for participant)	369
Table 60. Regression results – likelihood to use Tagalog-derived prepositions (observations = 8,134, $R^2 = 0.368$, random intercepts for participant)	369
Table 61. Logistic regression results – likelihood to use English-derived prepositions (observations = 8,134, $R^2 = 0.473$, random intercepts for participant).....	370
Table 62. Logistic regression results – likelihood to use Mandarin-derived prepositions (observations = 8,134, $R^2 = 0.733$, random intercepts for participant).....	370

LIST OF FIGURES

Figure 1. Summary of dominant codes across age groups (n = 65, in percentage) (Gonzales 2017c:203).....	11
Figure 2. Vowel phonemes of Lánnang-uè, adapted from Gonzales & Starr (2020).....	21
Figure 3. Acoustic comparison of duration between a minimal pair in Lánnang-uè: (a) <i>recôrd</i> [ˈɹe ³³ koɪd ⁵⁵] vs. (b) <i>recôrd</i> [ɹe ³³ ˈkoɪd ⁵⁵].....	33
Figure 4. Acoustic comparison of duration between <i>percênt</i> [pɹ̩ˈsent] and <i>simpèr</i> [ˈsim pɹ̩]..	34
Figure 5. Histogram (frequency), density, and box plot: individual intraspeaker consistency scores (conformance to the stress pattern); broken line indicates mean	193
Figure 6. Marginal means/effects of age, sex, and proficiency on duration and expected stress	196
Figure 7. Histogram (frequency), density, and box plot: individual intraspeaker consistency scores (lexical tone in Lánnang-uè words); broken line indicates mean.....	197
Figure 8. Marginal effects of proficiency in Tagalog and English on likelihood to use lexical tone in a Lánnang-uè word.....	199
Figure 9. Histogram (frequency), density, and box plot: individual intraspeaker consistency scores (adherence to the CVT/CV tone pattern); broken line indicates mean	200
Figure 10. Histogram (frequency), density, and box plot: individual intraspeaker consistency scores (adherence to the CVR-English/CVR-Tagalog tone pattern); broken line indicates mean	202
Figure 11. Syllable structure and pitch slope by source language of the word (final syllables of Tagalog- and English-origin words).....	206
Figure 12. Marginal means/effects of (a) age, (b) sex, and (c) proficiency on the CVT/CV tone pattern in Tagalog- and English-origin words; ignore CVR, which was automatically included in plotting package; slope means close to zero suggest use of high tone, slope means -5 and below suggest use of falling tone, slope means near -3 (e.g., CVR syllables) indicate mixed use of falling and high tone.	207

Figure 13. Marginal means/effects of (a) age, (b) sex, and (c) proficiency – the CVR-Tagalog/CVR-English tone pattern; slope means close to zero suggests use of high tone, slope means -5 and below suggests use of falling tone, slope means near -3 (e.g., non-CVR syllables) suggests mixed use of falling and high tone. (n.s. = not significant).....	208
Figure 14. Final pitch slope, structure, age, and source language of the word (Tagalog- and English-origin words).....	209
Figure 15. Histogram (frequency), density, and box plot of proportions (adherence to conjunction distributional pattern); broken line indicates mean	240
Figure 16. Marginal means/effects of sociolinguistic factors on likelihood to adhere to the distributional pattern.....	243
Figure 17. Marginal means/effects of sociolinguistic factors on likelihood to adhere to the distributional pattern (Hokkien-derived conjunctions)	244
Figure 18. Marginal means/effects of proficiency on likelihood to (not) adhere to the distributional pattern (Tagalog-derived conjunctions)	245
Figure 19. Histogram (frequency), density, and box plot of proportions (adherence to preposition distribution pattern); broken line indicates mean	246
Figure 20. Marginal means/effects of sociolinguistic factors on likelihood to (not) adhere to the distributional pattern (Hokkien-derived prepositions)	250
Figure 21. Marginal means/effects of sociolinguistic factors on likelihood to use Tagalog-derived prepositions	251
Figure 22. Marginal means/effects of sociolinguistic factors on likelihood to (not) adhere to the distributional pattern (English-derived prepositions).....	252
Figure 23. Speakers' z-scored ratings of Lánnang-uè being <i>konyò</i> (n= 117) Diamonds represent the mean; the heavy line indicates the median; the dots represent outliers.....	262
Figure 24. Set-up of criminal investigation task.....	283
Figure 25. Sample note-taking grid (filled)	284
Figure 26. Screenshots of acceptability experiment	288
Figure 27. Histogram (frequency), density, and box plot of proportions (adherence to <i>wh</i> -phrase distributional pattern, production); broken line indicates mean.....	297

Figure 28. Histogram (frequency), density, and box plot of proportions (adherence to <i>wh</i> -phrase position distributional pattern, stratified by <i>wh</i> -type; left, <i>why</i> questions only, right, <i>how-/when-/where-/who-/what</i> -questions only); broken line indicates mean	299
Figure 29. Marginal means/effects of <i>wh</i> -phrase type, age, sex, and proficiency on likelihood to put <i>wh</i> -phrase at sentence-initial position (production)	301
Figure 30. Marginal means/effects of age, sex, and proficiency on likelihood to put <i>why</i> -phrases at sentence-initial position.....	302
Figure 31. Histogram (frequency), density, and box plot of proportions (adherence to <i>wh</i> -phrase position distribution, acceptability); broken line indicates mean.....	303
Figure 32. Histogram (frequency), density, and box plots of proportions (adherence to <i>wh</i> -phrase distributional pattern, acceptability), stratified by position and <i>wh</i> -phrase type; broken line indicates mean	306
Figure 33. Marginal effects of age, sex, and proficiency on likelihood to adhere to <i>wh</i> -question pattern based on acceptability judgments.....	308
Figure 34. Marginal effects of age, sex, and proficiency in <i>insitu</i> languages on ratings of sentence-initial constructions featuring <i>why</i> -phrases	309
Figure 35. Marginal effects of age, sex, and proficiency in <i>insitu</i> languages on ratings of sentence-medial or final constructions featuring <i>why</i> -phrases.....	310
Figure 36. Marginal effects of age, sex, and proficiency in <i>wh</i> -fronting languages on ratings of sentence-initial constructions featuring <i>how/when/where/who/what</i> -phrases	311
Figure 37. Marginal effects of age, sex, and proficiency in <i>wh</i> -fronting languages on ratings of sentence-medial and final constructions featuring <i>how/when/where/object who/object what</i> -phrases.....	312
Figure 38. Syllable duration (indicative of actual stress) in syllables expected to be stressed/unstressed for each participant	362
Figure 39. Pitch slope (imputed) in CV and CVT syllables in Tagalog- and English-origin words for each participant; slope means close to zero indicate use of high tone, slope means -5 and below indicate use of falling tone.....	363
Figure 40. Pitch slope (imputed) in CVR-English and CVR-Tagalog syllables for each participant; slope means close to zero indicate use of high tone, slope means -5 and below indicate use of falling tone.....	364

LIST OF APPENDICES

Appendix A. Completed Expanded Swadesh List (24-year-old Male Lannang)	342
Appendix B. Means, Standard Deviations and Standard Errors of Sixteen Lannangs' Likelihood of Using Lánnang-uè Word by Word Origin (1 = very likely, 0 = unlikely), Based on a 219-word Expanded Swadesh List.....	349
Appendix C. Sample Wordlist for Stress and Tone Production Experiment.....	356
Appendix D. Interview Questions.....	358
Appendix E. Supplementary Linear Regression Results - Stress Perception	360
Appendix F. Supplementary Statistical Tables for Chapter 4 : Stress and Tone Features	362
Appendix G. Supplementary statistical tables for Chapter 5 : Conjunction and Preposition Lexical Patterns.....	365

LIST OF ABBREVIATIONS

Abbreviation	Gloss
1	first person
2	second person
3	third person
ABI	ability marker/ modal of ability
ADV	adverb marker
APPROX	approximator
ART	article
BEN	benefactive
CAUS	causative/cause-marking
CLS	classifier
CMPL	complementizer
CMPV	comparative
CND	conditional
COLL	colleague marker
CONJ	conjunction
COP	copula
DAT	dative
DEF	definite
DEM	demonstrative
DES	desiderative
DET	determiner
DIR	direction marker
DUR	durative
EMPH	emphasis marker
EXC	exclusive
FUNC	function marker
GEN	genitive
INC	inclusive
INCH	inchoative
INDEF	indefinite
ITER	iterative
LNK	linker
MAN	manner marking
MOD	modifier
NEC	modal of necessity
NEG	negative marker
ORD	ordinal marker

ORIG	origin marker
PER	modal of permission
PF	perfect
PFV	perfective
PL	plural
POS	modal of possibility
PREP	preposition
PROG	progressive
PROH	modal of prohibition
PRT	discourse particle
Q	question marker
REL	relativizer
RES	reservational
RLM	relation specifier
RSLT	resultative
SG	singular
SUP	superlative

ABSTRACT (English)

Lánnang-uè [³⁵la²²naŋ⁵¹?we] (Philippine Hybrid Hokkien), as used in Manila, is a predominantly oral Sino-Philippine variety that has elements derived from Hokkien, Mandarin, English, and Tagalog. Its users, the Lannangs, are divided in their perceptions towards the variety – some view it as what is commonly known as a ‘language’, while the majority view it as broken Hokkien, ad-hoc code-switching, or an unstructured admixture. My previous research on Manila Lánnang-uè (henceforth, Lánnang-uè), which focused on three features, has shown some evidence for the former – that Lánnang-uè has high degrees of ‘linguageness’. It also revealed an intriguing pattern: a mismatch between popular folk belief and linguistic practice.

This dissertation seeks to further explore the patterns found through a comprehensive investigation of the variety. It aims to answer the question: Where does Lánnang-uè fall in the cline of linguageness? I do this by analyzing linguistic data across multiple levels of language with respect to established key properties relevant to linguageness, such as systematicity, spread, stability, linguistic independence, clustering, and user attitudes. I employ a wide range of methods and tools (e.g., descriptive, experimental, computational, corpus-based, ethnographic, sociolinguistic) in the hopes of answering this question. Furthermore, using the evidence collected, I hope to situate Lánnang-uè in the constellation of contact varieties/phenomena.

The results suggest that Lánnang-uè is highly language-like. A series of investigations across multiple features in the variety indicates high levels of systematicity in the variety. For example, adopting the view that structured variation is a fundamental part of language, I found that Lánnang-uè has variation that is systematically conditioned by social and linguistic factors, just like varieties with high levels of linguageness (e.g., Light Warlpiri, Gurindji Kriol, Chabacano). For at least some features, there is a strong indication that variation is systematically used to express particular social meaning(s). Another major finding in my investigation is that the features in Lánnang-uè have a high degree of spread and stability within the community. From a perspective of linguistic transfer, my findings also suggest that the patterns/features are relatively independent from the source languages of Lánnang-uè.

Furthermore, although many speakers do not perceive it as a full-fledged language, there are those who do, referring to it as ‘secret code’ and ‘mixed language’. There is also evidence of feature/pattern clustering. The findings corroborate my previous work on Lánnang-uè, which also suggest that the variety has high degrees of languageness.

The close resemblance of many Lánnang-uè lexical and grammatical elements to Hokkien and the fact that many community members refer to Lánnang-uè as Hokkien (e.g., broken Hokkien, adulterated Hokkien, nativized Hokkien) might, at first glance, lead one to definitively conclude that it is a variety of Hokkien. However, a closer examination of the sociohistorical and linguistic patterns involving Lánnang-uè and its users indicate that that may not be the case. The systematic derivation of lexical and grammatical elements from specific source languages, the presence of linguistic elements that cannot be traced back to any of the source languages, and the sociohistorical context (e.g., the emergence of the hybrid Lannang identity) show that Lánnang-uè has features of “mixed languages” in the Thomasonian and Bakkerian sense. Pending more research, the findings of this dissertation suggest that the most likely scenario is that Lánnang-uè is a mixed language or – if one shifts away from the idea of rigid typological categories and aligns with a view of a linguistic continuum – a variety situated somewhere in a continuum from ‘Hokkien’ to ‘mixed language’, leaning closely towards ‘mixed language.’ If that is the case, then my findings point to the existence of the first documented mixed language in the Philippines. They also point to a mixed language with linguistic patterns that diverge from the patterns observed in other mixed languages. However, if Lánnang-uè is analyzed as a Hokkien variety, then my findings point to another variety of Hokkien in the Philippines alongside Manila/Philippine Hokkien. Overall, Lánnang-uè has features that set it apart from other linguistic varieties and language types in its linguistic ecology. It is rightfully labeled Lánnang-uè – a language that its users can truly call their own.

ABSTRÂCT (Lánnang-uè)

Lánnang-uè [³⁵la²²naŋ⁵¹] (Philippine Hybrid Hokkièn), tiekpiet Manilá ièng e, si tsige halos oràl e variety na u mga elemènts na derivêd galing Hokkièn-uè, Kogî, Iengbún, kâp Huaná-uè. Í e mga usêrs, huaí Lánnáng, tuí tsíge variety e khuâhuât bo sáng. Uwé kamkâk î sî giêngû, pero majority kamkâk î sî barôk e Hokkièn-uè, ad-hoc e code-switching, âsî tsige admixtûre na tshîntshaí halô è. Guá tiengmaí e reseârch na focûs lê sa-e featurês ho dân evidênce na Lánnang-uè talagà parang giêngû (na î u yá kuí e ‘pagka-languagê’). Guá e reseârch reveâl tîoh tsige interestîng e pattèrn: láng tuí Lánnang-uè e khuâhuât kâp lang tsiúwâ ieng Lánnang-uè bo sáng.

Tsíge dissertatiòn bêh kay-explòre tsuaí mga pattèrn througħ tsīgē khâ comprehensivê e investigatiòn of hîge variety. Guá balâk kay-answèr hîge questiòn: Ti cline of pagka- languagê, Lánnang-uè tolóh situatêd a? Para guá uhuâtthang answèr tsíge, guá kay-analyzê linguístic e data across yá tsue levels of giêngû munâ. Guá iêng tsuaí mga establishêd e parametêrs: pagka-systematic, pagka-spreâd-oût, pagka-stablê, pagka-independênt, clustering, kâp láng tuí hîge variety e attitudês. Guá kay-emplòy tsintsue mga methòd kâp tool (e.g., descriptivê, experimentâl, computationâl, corpus-basêd, ethnographíc, sociolinguístic). Guá hôpe tsuaí ũhuâtthāng pāngtsān answèr tsíge questiòn. Iêng huaí mga evidencê, guá balâk kay-situatê Lánnang-uè tî hîge constellatiòn of mga contact variety âsî phenomènòn.

Guá-e resúlts suggêst na Lánnang-uè tsin parang giêngû. Tsige seriês of investigatiòn across yá tsue featurês ti hîge variety reveâl-tîoh yá kuí e systematicity tî hîge variety. Halimbawâ, habang guá lê adôpt hîge viêw na structurêd e variatiòn si tsige yá tiong-iaù e bahagî ng mga giêngû, guá lamân-tîoh na Lánnang-uè ũ variatiòn nā hō tsikuá sociâl kâp linguístic e feature condition-tîoh, parang mga variety na u yá kuí e pagka-languagê, nántshiū Light Warlpiri, Gurindji Kriòl, kâp Chabacanò. Para at leâst tsikuá mga featurê, guá ũ evidencê na láng lê iêng tsíge variatiòn para uhuâtlang exprêss particulâr e (mga) social meanìng. Kô tsīgē findìng si Lánnang-uè e featurês u ya kuí e pagka-spreâd kâp stability within hîge community. Tuí tsige linguistic transfèr e khuâhuât, guá e mga findìngs hînt na Lánnang-uè e mga pattèrn kâp features tuí î-e source languagês mejo independênt. Tapôs, kahit na yá tsuê Lánnáng bo perceivê

Lánang-uè tsuê tsige independênt e giengû, u Lánáng na ũ. Tsuaí Lánáng refèr Lánang-uè tsuê ‘secrèt e côde’ kâp ‘míxed e languagê’. Guâ u evidencê of featurè âsî pattern clustering dîn. Guâ e mga findîng corroboratê guâ tiengmaí e wôrk na u suggêst-tioh na hîge variety u yá kuí e pagka-languagê.

Dahil yá tsue Lánang-uè e mga featurè kâp Hokkiên-uè e mga feature satáng kâp dahil yá tsue Lánáng kóng Lánang-uè sî Hokkiên-uè (e.g., baròk e Hokkiên-uè, adulteratêd e Hokkiên-uè, nativizêd e Hokkiên-uè), lang baka e kóng Lánang-uè talagà sî tsi khuân e Hokkiên-uè. Pero pag na dân kay-examinè huaí mga sociohistorical kâp linguistic e pattern na involvê Lánang-uè kâp î-e mga usêrs, tsîge hypothesis yá unlikely. Hîge systematic e pagka-derîve ng pang-lexicòn kâp pang-grammàr e elemênts ân specific e mga source languagê, hîge presencê of elemênts na buē tracê-dit, kâp hîge sociohistorical contêxt (e.g., halo-halô e Lánang identity) ho dân khuà na Lánang-uè u mga featurês ng mga “mixed languagê,” tsiaû Thomasòn kâp Bakkèr e khuâhuât. Dân khiâm-ieng khâ tsue reseârch pà pero tsîge dissertatiòn e mga findîng suggêst na hîge pinaka-khódiéng e scenariò sî Lánang-uè sî tsîgê mixed languagê âsî – na dî tsiaû hîge khuâhuât na giengû bo huât thāng khē lê yá rigid e mga category – tsige giengû na ti ‘Hokkiên’ kâp ‘mixed languagê’ e gitnâ, khâ kûn ‘mixed languagê’. Na tsîge si tióh-e, guâ e mga findîngs ho dân khuà na Lánang-uè sî Huīdīpīn-e thaū tsîgê documentêd e mixed languagê. În presênt tsige mixed languagê na ũ mga pattèrn na kâp pade mga mixed languagê bosáng dîn. Pero pag Lánang-uè analyzê tsuê Hokkiên-uè e dialect (tsige khâ bo khódiéng-e scenariò), guâ e findîngs highlight kô tsige variety ng Hokkiên-uè ti Huīdīpīn, kasamà Manila/Philippine Hokkiên (Bānlām-uè). Overall, Lánang-uè sî tsige languagê na u mga featurês na distinguîsh ì tuí pade dialêcts kâp côdes tî î-e linguistic ecology. Í tamang-tamâ labèl tsuê Lánang-uè – tsige languagê na î-e mga usêrs uhuâtthāng talagà kiò tsuê î-n-e.

Chapter 1 : Introduction¹

Lánnang-uè [ˈla³⁵naŋ²²ʔwe⁵¹] (derived from Hokkien *lân lâng uè* ‘our people speech’) or “Philippine Hybrid Hokkien”² (Gonzales 2018:1) is a predominantly oral Sino-Philippine linguistic variety with elements from (Philippine) Hokkien, Mandarin, (Philippine) English, and Austronesian languages in the Philippines (e.g., Tagalog). It is primarily used by an estimated 1.5 million Lannangs, individuals with a mixed Southern Chinese and Filipino cultural heritage (Uytanlet 2014). Although some speakers perceive it as a language that systematically incorporates linguistic elements from particular source languages (Gonzales’ field notes, summer 2019), many perceive it as a dialect of Hokkien “adulterated” by words or phrases of non-Hokkien origin (Ang See 1990:14).³ Sometimes analogized to *halo-halô* ‘mix-mix’ (a local dessert that consists of a heterogeneous concoction of sweet ingredients), Lánnang-uè is perceived by many of its speakers as ad-hoc code-switching or random mixing (Gonzales’ field notes, summer 2019). Several speakers anecdotally claim that Lánnang-uè is unstructured and unstable, citing high levels of intra- and interspeaker variation. They also believe that many of its linguistic features are idiolectal or familectal – used only by an individual or among members of a family or other small intimate group (i.e., not widespread in the community) (Gonzales’ field notes, summer 2019). Overall, they perceive Lánnang-uè as not being language-like.⁴

¹ The title of this dissertation is partially inspired by the title of Bakker’s (1997) much-cited book on the mixed language Michif – *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Métis*.

² Although the term ‘Lánnang-uè’ is used to refer to both Philippine Hokkien and Philippine Hybrid Hokkien, many speakers report a distinction. Some speakers claim, for instance, that ‘Lánnang-uè’ is used for day-to-day life while ‘Hōkkiên-uè’ (Philippine Hokkien) is used for church. For this dissertation, I will use the term ‘Lánnang-uè’ to refer to Philippine Hybrid Hokkien.

³ The rest report not having an insight or acknowledge that they do not have the expertise to comment on this.

⁴ I do not view ‘language’ in binary terms (e.g., language vs. ephemeral code-switching/admixture involving multiple languages) Instead, I view it as something gradient. See Chapter 2.3 for a discussion of languageness and properties of languages.

My initial investigations of the metropolitan Manila variety of Lánnang-uè (henceforth Lánnang-uè, for convenience) showed evidence of an intriguing but certainly not novel pattern: a mismatch between popular folk belief and actual linguistic practice. For one, I found evidence of high degrees of systematicity within the linguistic variety. For example, I observed that Lánnang-uè incorporates elements from Tagalog,⁵ English, and Hokkien systematically (Gonzales 2017a; Gonzales 2018; Gonzales and Starr 2020), consistent with some descriptions of mixed languages (Matras and Bakker 2003:1; Winford 2013; Gonzales and Starr 2020). I found some evidence that the mixing in the Lánnang-uè is not random, contrary to what many Lánnang-uè speakers report. In the same preliminary investigations, I also found high rates of stability and spread of Lánnang-uè features within the community. For one, the features observed were not only used by a specific speaker or group of speakers, but by most speakers in the sample. Furthermore, the features were used or judged to be acceptable with high levels of consistency: most speakers used/judged them with little variation, contradicting what many Lánnang-uè speakers have reported. The patterns of variation were also not ‘random’ and were found to be conditioned by social factors (e.g., age, sex) (Gonzales 2018; Gonzales and Starr 2020).

But while I have found evidence supporting a highly systematic, widely-used, and stable Lánnang-uè – a Lánnang-uè with high degrees of languageness – there is still a need for a more comprehensive and in-depth study. An examination of a few linguistic features is not enough to generalize about the nature of Lánnang-uè. It would also be beneficial to formally examine the languageness of Lánnang-uè using identified properties of ‘language’ (e.g., systematicity, stability) (Chapter 2.3).

As such, the overall goal is to investigate Lánnang-uè comprehensively to determine whether the variety has features of languageness, such as high degrees of systematicity and independence from peripheral languages (Chapter 2.3). The findings will help me get closer to answering the question of how language-like Lánnang-uè is. It might also offer us clues to where the variety is situated within the typology of contact languages. Because I intend to test the hypothesis that Lánnang-uè is highly language-like, I refer to Lánnang-uè using terms that do not

⁵ In this dissertation, I am not interested in the distinction between Filipino and Tagalog. Filipino, in the strictest sense, refers to a variety of Tagalog that has been enriched by local languages including English, Spanish, and Hokkien. Tagalog, on the other hand, refers to the variety that has limited borrowings or foreign influence.

prematurely commit me to a particular linguistic theory of Lánnang-uè (e.g., Lánnang-uè as ad-hoc code-switching, Lánnang-uè as a language). Such terms include ‘linguistic variety’, ‘admixture’, and ‘code’.

The research questions that underlie this dissertation are as follows:

1. Where does Lánnang-uè fall in the cline of languageness?
 - a. Is Lánnang-uè highly systematic? Are the patterns of variation in it structured? That is, can they be explained by (socio)linguistic factors?
 - b. Are the features of Lánnang-uè used at all by most speakers in the Lánnang-uè-speaking community or only a small subset?
 - c. Is Lánnang-uè highly stable? Will speakers be consistent in the use of its features?
 - d. Is Lánnang-uè linguistically independent? Are the features/patterns of the variety independent from the features/patterns of its source languages?
2. If it has a high degree of languageness, where does it fall in the typology of contact languages? If not, what kind of contact phenomenon is it?

I attempt to answer these questions in my dissertation, relying on four primary sets of data:

1. **Corpus data** – I use the Lannang Corpus (LanCorp), a 375,000-word monitor corpus consisting of spontaneous speech (e.g., narratives, sociolinguistic interviews, casual conversations) recorded from 135 Manila Lannangs of diverse social backgrounds between 2016 and 2020 (Gonzales 2022a). It was transcribed by a team of trained transcribers. Approximately 85% of the data/sentences in the LanCorp are in Lánnang-uè. Other sentences are code-switches to Tagalog, Mandarin, and English.
2. **Elicitation data** – The elicitations were acquired between January 2019 and May 2020. Roughly half of the data was gathered via face-to-face sessions, while the other half was gathered online due to the pandemic. I used a combination of methods on thirteen consultants. My primary method for eliciting linguistic data was the “target language

interrogation elicitation” method (Chelliah and de Reuse 2011:368), where I verbally described a situation and asked what the consultant might say in that situation at a given moment. There were times when the interrogation approach was not successful, in which case I used “stimulus-driven elicitation” (Chelliah and de Reuse 2011:368). I presented an object or picture (in context) to the consultant and asked them to describe it, or to comment on it.

3. ***Judgment data*** – The judgments were acquired in the same time frame and the same mode as elicitation data. I used the “target language manipulation” method (Chelliah and de Reuse 2011:370), where I manipulated some word or structure of the target language and asked the consultant to react to it.
4. ***Experimental data*** – The data came from various experiments – a production experiment under the guise of game (see Chapter 6.5), a production experiment involving a wordlist (see Chapter 4.4), as well as an acceptability judgment experiment (see Chapter 6.5).

First, I contextualize the project by giving an overview of the Lannangs, Lánnang-uè, and its speakers (Chapter 2). I also briefly discuss notions of languageness in that section. Then, I describe the variety with the aim of uncovering evidence for languageness (Chapter 3). This is followed by a comprehensive investigation of seven features in three domains – prosody, lexicon, and syntax – that I anecdotally found to have higher rates of variation than other features that could potentially weaken the argument for a Lánnang-uè that is highly language-like.

In Chapter 4, I investigate one stress feature and three tone features in the variety; in Chapter 5, I examine patterns involving conjunctions and prepositions; in Chapter 6, I focus on matrix *wh*-questions. In these three chapters, I systematically investigate the features with respect to four properties of languageness: degree of spread within the community, degree of stability, systematicity, and linguistic independence (as outlined in Chapter 2.3). In the final chapter, Chapter 7, I summarize my findings and attempt to answer the questions posed in this chapter; I also discuss the implications of my dissertation and potential areas for further research.

If Lánnang-uè is similar to linguistic varieties reported to have high degrees of languageness such as Singlish, Baba Malay, Light Warlpiri, and Gurindji Kriol (Starr and Balasubramaniam 2019; Lee 2014; Meakins and O’Shannessy 2010), then many of the features examined in Lánnang-uè should be widespread and stable within the community. The patterns of variation, if any, should be conditioned or constrained by certain factors (e.g., linguistic, social).⁶ The patterns in Lánnang-uè should also not be dependent on (or greatly influenced by) the patterns of its source languages.

⁶ This is assuming that structured variation or “orderly heterogeneity” is a fundamental part of language (Weinreich et al. 1968:100).

Chapter 2 : General Background

2.1 The Lannangs

The Lannangs⁷ (derived from the Hokkien phrase *lán láng* ‘our people’) are individuals who have Southern Chinese ancestry and a mixed Chinese-Filipino cultural heritage.⁸ The group primarily consists of the late 19th to early 20th century Southern Chinese immigrants from the Fujian and Guangdong provinces to the Philippines, and their descendants (Doeppers 1986).

Some terms that group members use to refer to themselves within their group (in Lánnang-uè), based on ethnographic work conducted in summer 2018 and 2019 (Gonzales 2021a), include:

- *Fīl-Chì* [‘fī¹²²’tʃaj⁵¹]
- *Lánnáng* [la³⁵naŋ³⁵] or [lan³⁵naŋ³⁵]
- *Nánnáng* [nan³⁵naŋ³⁵]
- *Tsinôy/Chinôy* [tʃi²²’noj⁵⁵]
- *Huakiaú* [hwa²²kjaw³⁵]

Some terms that they use to refer to themselves to individuals outside their group are:

⁷ Within the Lannang community, the term ‘Lannang’ is usually used to refer to anyone with Chinese heritage (Uytanlet 2014:93). Sometimes, however, it is used more restrictively, only referring to subset of this population – individuals with a particular hybrid Chinese-Filipino heritage who are living in the Philippines (Chu 2021:3). Many Lannang community members assume that other individuals with (Southern) Chinese-heritage (e.g., Chinese Singaporeans, Peranakan) use ‘Lannang’ to refer to themselves; however, beyond the Lannang community, this term is not used to refer to Chinese-heritage individuals (2019-2020 fieldwork). Non-Lannang individuals with Chinese heritage (e.g., Mainland Chinese speakers temporarily residing in the Philippines) do not identify with this, and only the Lannangs do. Like Chu (2021), in this dissertation, I use the term ‘Lannang’ restrictively to refer to just the community with a particular hybrid heritage with Southern Chinese and Filipino influence (as defined in the paragraph) and not Chinese-heritage individuals in general.

⁸ It is impossible to investigate the actual situation of genetic mixing in the Lannang population due to lack of genetic data. However, based on the self-reported genealogical histories of 70 Lannangs, it appears that the majority of Lannangs are only culturally mixed; they are endogamous.

- *Chinese Filipino* (Uytanlet 2014)
- *Chinese in the Philippines* (Ang See 1990; Chua 2004)
- *Overseas Chinese* (Chua 2004)
- *Filipino-Chinese* (Chuaunsu 1989; Uytanlet 2014)
- *Lannang* (Chu 2021; Gonzales 2021a)
- *Pinsinos* (Ang See 1990)
- *Tsinoy/Chinoy* (Uytanlet 2014)

In terms of race, many Lannangs report themselves as having pure Chinese ancestry. There are also many who report having mixed Filipino and Chinese ancestry (*tshûtsi-â* ‘mixed blood’) who identify as Lannang. In terms of citizenship, most Lannangs are Filipinos by birth or by naturalization, but some do not have Filipino citizenship, even if they have lived in the Philippines their entire lives (Gonzales 2021a). In a sample of 37 Lannangs in Manila, roughly 16% report not having Filipino citizenship for various reasons, such as ineligibility for citizenship and preference for other citizenships. The Lannangs may seem like a heterogenous group from the perspective of self-reported Chinese ancestral lineage and citizenship. But most, if not all, Lannangs experience a common (hybrid) culture: they are educated in a Lannang school with a multilingual curriculum (English, Filipino, and Chinese) and engage in Filipino, Chinese, and Lannang cultural traditions and practices (Ang See 1990:108; Gonzales 2021a).

The Lannangs are also multilingual. They pride themselves on their ability to communicate in various languages (Gonzales 2018). The Lannangs of metropolitan Manila, the focus of this paper, generally claim knowledge of, if not proficiency in, at least three languages – Hokkien, Tagalog, and English. The first, Philippine Hokkien, is a dialect of Southern Min that has elements from Jinjiang Hokkien and Amoy Hokkien. Jinjiang and Amoy Hokkien were brought to the Philippines in different waves of migration from Jinjiang and Xiamen, respectively, beginning the 1850s (Doeppers 1986:382). The second, Tagalog, is an indigenous language in the Malayo-Polynesian branch of the Austronesian language family. The third is the local English or Philippine English.⁹ This variety (or more accurately, set of varieties) (Gonzales

⁹ By ‘Philippine English’, I am referring to a variety used in metropolitan Manila (Gonzales & Hiramoto 2020).

2017b; Gonzales and Hiramoto 2020; Gonzales 2022b) developed out of contact between Philippine languages (e.g., Tagalog in Manila) and American English, which was first introduced into mainstream Philippine society around the 1900s via educational reform during the American occupation (Bautista 2004).

Some of the Lannangs report being proficient in Mandarin, a Sinitic language that has been taught formally in most Lannang schools. Mandarin is rarely used in the community and is typically used only in domains where it was taught (e.g., school). The Lannangs of Cantonese heritage, forming around 10% of the total Lannang population of 1.5 million, consider themselves proficient in ‘standard’ Cantonese and/or Taishanese, a dialect of Cantonese spoken in the Taishan region of Southern China. There are Lannangs who also explicitly claim proficiency in “mixed Hokkien” (Lánnang-uè) (see Section 2.2.1 for terminology).

2.2 *Lánnang-uè*

2.2.1 *Terminology*

The term Lánnang-uè [³⁵la²²ʔwe⁵¹] was derived from the Hokkien phrase *lân lâng uè*, which literally means ‘our people speech.’ Other variants of this term that are used in the Lannang community include Lánlang-uè [³⁵lan²²ʔwe⁵¹] and Nánnang-uè [³⁵nan²²ʔwe⁵¹], although Lánnang-uè is the most common variant used, based on ethnographic fieldwork.

Within the community, many speakers claim that Lánnang-uè is equivalent to Hokkien or *Minnanhua* ‘Southern Min’ (Uytanlet 2014:161).¹⁰ However, looking at the actual practices of the community and how they correlate with terminology, the term is used to refer to a variety that blends Hokkien-, English-, Tagalog-, and Mandarin-derived elements (Gonzales 2021a). In my 2019 and 2020 fieldwork, when speakers are asked to use Lánnang-uè, most (135 of 142, 95%) of respondents did not use Hokkien (or a code with only Hokkien elements), but rather a mixed code with elements coming from the four languages mentioned earlier. As such, throughout this dissertation, I use the terms Lánnang-uè and Philippine Hybrid Hokkien (Gonzales 2018:1) to refer to the admixture instead of Philippine Hokkien or ‘a variety of Philippine Hokkien’. I distinguish between Lánnang-uè and Philippine Hokkien.

¹⁰ It is interesting to note that although ‘Lannang’ is used by most community members to refer to ‘Chinese’ in the general sense, ‘Lánnang-uè’ is not used to refer to ‘Chinese’ (language). Instead, speakers use ‘Hâmbún’ or ‘Huāgî’.

2.2.2 Investigated features in the scholarly literature on Lánnang-uè

There is little work on the features of Lánnang-uè. The first description of it, to my knowledge, is that of Ang See (1990:14), who anecdotally characterized the linguistic variety as an “adulterated” Chinese or “Chinese that is mixed with Filipino prefixes ... suffixes [and] syntax and spoken in Filipino tones” (14). Uytanlet (2014:139) has also described it as a mixture of English, Filipino, and Chinese – the result of “failure of transmission or mastery of the [Hokkien] language.”

The rest of the work on Lánnang-uè – conducted either by me alone or in collaboration with another scholar (Gonzales 2018; Gonzales and Starr 2020) – undermines this view of Lánnang-uè as an unsystematic, ad-hoc linguistic admixture. Our preliminary findings indicate high levels of languageness in Lánnang-uè.

In Gonzales (2017a), I documented ten types of question tags in the variety. I noted that eight of them were systematically derived from Hokkien, Tagalog, and English while two of them (i.e., *msibá* ‘is it not’ and *okáybo* ‘okay’) are innovative combinations from multiple source languages.

In Gonzales (2018), I found that all my Lánnang-uè-speaking participants tended to use and accept nominal derivational affixes if they were sourced from Tagalog and if they are simple prefixes (e.g., colleague prefix *ka-*, manner prefix *pag-*) or compound prefixes (e.g., state-of-being *pag+ ka-*). Compound Tagalog prefixes (e.g., *pagkakapag-*) and English derivational suffixes (e.g., *-tion*) are not used in Lánnang-uè. Constructions with them were also consistently given low ratings in acceptability judgments. My 2018 findings suggest that the derivational affixation feature of Lánnang-uè in the nominal domain is highly systematic, widespread in the community, and stable.

In Gonzales & Starr (2020), I explored the monophthongal system of Lánnang-uè. I found that monophthongs in English-, Tagalog-, or Hokkien-sourced Lánnang-uè words generally share the same vowel qualities. That is, all speakers were using a unified monophthong system instead of three systems. Some younger speakers were occasionally found to produce some vowels (e.g. [e] and [ʊ]) differently depending on the language. They would, for example, raise their [ʊ] in Hokkien- sourced Lánnang-uè words and lower their [ʊ] in English-sourced

ones (Gonzales and Starr 2020). Overall, our findings show acoustic and phonological evidence of high levels of consistency and structured variation in Lánnang-uè.

Overall, only three features of Lánnang-uè have been documented. The findings for these features hint at a high degree of languageness for Lánnang-uè.

2.2.3 Domains of use

In earlier work (Gonzales 2016; Gonzales 2018), I suggested that the use of Lánnang-uè is restricted only to informal domains. However, in my 2019 and 2020 fieldwork, I have also observed the use of Lánnang-uè in formal domains (e.g., business association meetings, church gatherings, weddings, teacher instruction in schools). This shows that Lánnang-uè is not only used colloquially; it is robustly used by the Lannang community in many domains of communication. As an in-group or community-based variety, Lánnang-uè is rarely used between Lannangs and non-Lannangs. It is sometimes used with non-Lannangs in the context of teaching, jokes, etc. It is rarely used by Lannangs when talking with ‘Mainlanders’ or Chinese from the People’s Republic of China located in Mainland China, although I have recently come across a small number of Lannangs using Lánnang-uè with (Mainlander) immigrant Chinese who have stayed in the Philippines for a long time.

2.2.4 Speakers and language background

Speakers of Lánnang-uè have different codes in their linguistic repertoire. In a 2017 survey, I found that these repertoires vary by age – older Lánnang-uè speakers, particularly those in their 80s, have Philippine Hokkien as their dominant code and have Lánnang-uè, Tagalog, English, and Mandarin as non-dominant ones (Gonzales 2017c). I observed that Lánnang-uè speakers in this age group have limited communication with non-Lánnang-uè speakers (e.g., communicating with their domestic helpers). They do not use Tagalog and English very frequently, although in some cases, these older Lánnang-uè speakers accommodate to their young Lánnang-uè-speaking relatives or peers by responding in a non-native variety of Tagalog or English. In most cases, however, the older Lánnang-uè speakers communicate with other Lánnang-uè speakers in Philippine Hokkien. It is important to note that they tend to interact more with their Lannang peers compared to non-Lannangs. Equally worth noting is that they rarely communicate with

their Lannang peers in Mandarin, even if they have knowledge of it from formal exposure in school and seem to be more proficient at it than the other age groups.

Lánnang-uè speakers in their 70s use Lánnang-uè and Philippine Hokkien almost equally, with a slight dominance in Lánnang-uè. Middle-aged speakers, those in their 40s to 60s, have Lánnang-uè as their dominant language. Both groups have Tagalog, English, and Mandarin in their linguistic repertoire but do not have them as dominant languages. On the other hand, younger speakers in their 20s and 30s have Tagalog as their dominant code (Gonzales 2017c:203) and Lánnang-uè, English, Hokkien, and Mandarin as non-dominant ones (ethnographic fieldwork 2017). I summarize this in Figure 1. The figure shows the mean relative dominance (percentage) of six Lannang codes used across domains (e.g., with parents, with friends) (Gonzales 2017c), while the table highlights the dominant codes.

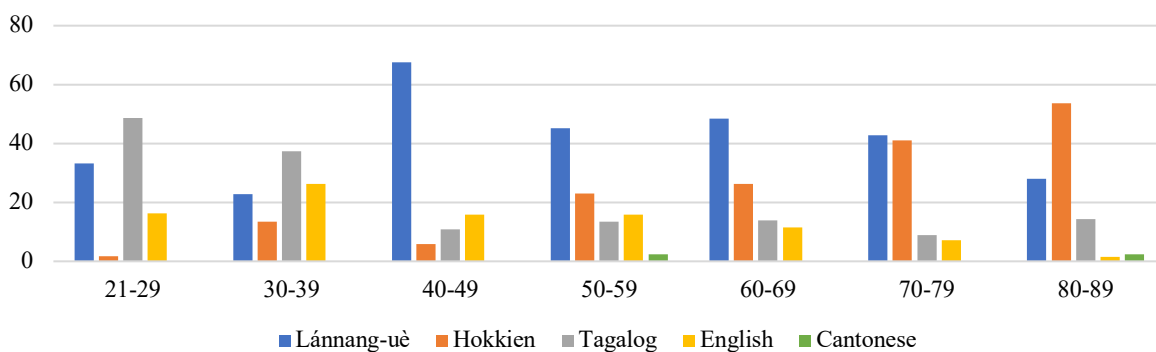


Figure 1. Summary of dominant codes across age groups (n = 65, in percentage) (Gonzales 2017c:203)

In terms of interactions with non-Lánnang-uè speaking peers, the middle-aged and young groups do not differ significantly. They interact with both Lannangs and non-Lannangs.

All age groups were formally taught Mandarin (explained using Lánnang-uè) but their degree of exposure to the language varies – there are differences between groups. Older speakers were not as exposed to Mandarin compared to younger speakers. Speakers in their 80s, for example, were taught using a Hokkien-dominant curriculum that had minimal or no Mandarin

whereas young speakers in their 20s were taught using a trilingual one (English-Filipino-Mandarin).¹¹

Apart from Mandarin exposure, another crucial difference between the age groups lies in their formal exposure to Hokkien, Tagalog, and English. The middle-aged group has been exposed to two curricula – they had English and Filipino classes in the morning and Chinese¹² classes in the afternoon. As far as I have been able to determine, they have been equally exposed to Tagalog, English, and Hokkien. The younger group, on the other hand, has only been minimally exposed to Hokkien. They have only taken Chinese subjects as electives (e.g., *Huāgí* ‘Chinese language’, *Suānsút* ‘Chinese math’) instead of a mandatory curriculum. Because of this, the young group had more formal exposure to English and Tagalog compared to Hokkien.

Another difference lies in their use of Lánnang-uè. Speakers in their 80s tend to use Lánnang-uè and Hokkien with Lannang speakers with the same age as them. When speaking to younger Lannangs, they use Lánnang-uè and sometimes English and Tagalog. With Lannangs older than them, they rarely use Lánnang-uè. The middle-aged speakers generally use Lánnang-uè with peers their age as well as individuals younger and older than them, sometimes switching to English, Tagalog, and Hokkien. Young Lánnang-uè speakers typically use Lánnang-uè only with Lannang speakers much older than them. Sometimes they use Tagalog and English. With Lannang individuals around their age range or younger than them, they almost always use Tagalog and/or English, occasionally using Mandarin, Hokkien, and Lánnang-uè.

2.2.5 *Speakers’ perceptions*

Many Lánnang-uè speakers do not recognize Lánnang-uè – as defined in Section 2.2.1 – as a distinct language (summer 2019 notes), perceiving it as “adulterated” Hokkien (Ang See 1990:14; Uytanlet 2014).¹³ There are others, however, who recognize it as a language that sets them apart from the non-Lannangs (e.g., mainland Chinese and Filipinos) (Gonzales 2021a).

¹¹ The exposure of Mandarin also depends partially on the curriculum of the Lannang school. This is because Lannang schools do not have a standardized curriculum, unlike in the case of Singapore or Taiwan.

¹² Based on my interviews with participants, the kind of ‘Chinese’ language taught depends on the curriculum of the Lannang school. Some schools have Chinese classes that are taught exclusively in Hokkien or Mandarin. The majority, however, teach in both: teachers use Mandarin textbooks but explain concepts in Mandarin and Hokkien. There are times when teachers would even use Lánnang-uè.

¹³ It is worth noting that Lánnang-uè is perceived as broken Hokkien and not broken English or Tagalog potentially because a large percentage of the Lánnang-uè lexicon is sourced from Hokkien relative to other languages, based on a preliminary lexical composition analysis conducted in 2019.

Many older Lánnang-uè speakers generally characterize it as a broken variety of Hokkien. Nevertheless, some of them view it as a code that young people have created, so while they are hesitant to accept ‘mixed Hokkien’, they find it necessary for the survival of the community’s heritage. Some of the middle-aged and young speakers view it as a failed attempt to acquire Hokkien (Ang See 1990; Uytanlet 2014), even if most are bilingual in both Philippine Hokkien and Lánnang-uè, particularly the middle-aged group. But there are also some speakers who have indicated that ‘mixed language’ or ‘code-mixing’ in the form of Lánnang-uè sets them apart from other groups with Chinese cultural heritage, acknowledging it as a distinct code that reflects their mixed Chinese and Filipino identity (Gonzales 2021a).

The uneven perceptions of many Lánnang-uè speakers toward Lánnang-uè pose a complication for the analyses of this dissertation, as participants might be using and making judgments with respect to Hokkien, even if the task requires them to use and make judgments with respect to Lánnang-uè, which, as stated, is generally perceived as ‘broken’ Hokkien. This is a necessary limitation of the dissertation, given the complexity of the Lannang language situation.

2.3 *Language and languageness*

‘Language’ is ironically one of the most difficult terms to define in linguistics. Scholars tend to have very different views on what constitutes a ‘language’ or how it should be defined. Those who view ‘languageness’ as a “matter of degree” (Görlach 2002:70) tend to disagree on the set of social and linguistic qualities or characteristics that define ‘languageness’.

In a short literature review, I noted that research discussing notions of ‘language’ and/or ‘languageness’ cluster in four groups.

The first group features studies that primarily use linguistic criteria to define languageness. One example is Meakins’ (Meakins 2012), who analyzed a contact variety called Gurindji Kriol using criteria of “language-hood”. She argued that Gurindji Kriol is a language because it has a high degree of interspeaker consistency. Her claim for languagehood also came from the finding of linguistic autonomy (i.e., the existence of source-language-independent forms and structures) as well as systematic (predictable) variation. In addition, she argues that Gurindji Kriol is a language because it has inflectional morphology from both Gurindji and Kriol – evidence of a single composite grammar and structural fusion. In other words, she heavily

relies on linguistic criteria to assess the languagehood of Gurindji Kriol, at least for this paper. Another scholar who adopted primarily linguistic criteria is Wichmann (2020), who used phonological distance coming from lexical data to distinguish languages from dialects. Specifically, he argues that the use of Normalized Levenshtein Distance – “the number of substitutions, insertions, or deletions required to transform one word into another” (Wichmann 2020:824) – between lexical sets is a universal criterion for distinguishing between language and dialect pairs. Tosco (2021) relied on the criterion of intelligibility (mutual unintelligibility) to measure languageness. He critically objects to a language-internal definition of ‘language’ (arguing that language is a purely social entity and the distinction between dialects and languages is artificial) and he argues languages exist because communication exists.

The second group of studies uses a socially-oriented criterion to determine languagehood. Comparing English and Irish Traveller Cant, for example, Rieder (2018) argues that Cant is a language because it has a community that has expressed ownership of the speech form. Cant also indexes solidarity within the community and is used as a performative practice to express social meaning. Rieder argues that Cant is language-like because it is deliberately used in situations where private content needs to be communicated. It is used as a tool to include those who are part of the community and exclude those they consider outsiders. In other words, Rieder viewed ownership, solidarity, activity (performance), and privacy/secretcy as essential criteria for languageness. Makoni and Pennycook (2007) characterize languages as social inventions – they advocate for an approach where there are no boundaries between languages and dialects/linguistic varieties (i.e., without diglossic functional separation). Language, for them, is essentially a collage of performative acts.

The third group of studies on languageness places equal emphasis on social and linguistic criteria. In one of the earliest works to discuss the issue of ‘language’, Haugen (1966) proposed four main criteria for languagehood. He argues that languages (as opposed to ‘dialects’) have four social and linguistic aspects: selection of norm and acceptance by community (social norms), and codification of form and elaboration of function (stylistic variation) (linguistic norms). Görlach (2002) has similar views and identified several linguistic criteria for languageness, such as structural distance and stylistic variation as well as social criteria like “self-perceptions of speakers and their will to be linguistically independent” (i.e., attitude) and

acquisition, or whether the variety was learned as a mother tongue and is dominant in the community (Görlach 2002:70).

The fourth group is composed of studies that take a more cognitivist and/or generativist approach to languageness (Balari and Lorenzo 2018). These works view language as an (inherent) property of an individual. The implication of such a theory is that everything that is produced by any individual is automatically considered language. The works of Chomsky (1986) are such examples. A comprehensive survey of such works is given by Balari and Lorenzo (2018).

In this dissertation, I do not view ‘language’ as something that is determined binarily (e.g., a variety being either a language or a variety involving code-switching between distinct languages). Instead, I view it as gradient (i.e., some varieties have higher degrees of ‘languageness’ than others), similarly to a number of other scholars (Görlach 2002; Rieder 2018; Tosco 2021). The variables I use to characterize ‘languageness’ are derived from works that utilize linguistic and social criteria. While languageness, as shown earlier, has many correlates, I focus on six primary criteria in this dissertation, all of which are tailored to the Lannang community. Like the first group of studies on languageness (just reviewed above), I adopt a primarily linguistic approach while acknowledging the social and cognitive underpinnings of language (e.g., identity-based entity, innateness). ‘Languageness’ here is operationalized by six diagnostic criteria:

1. *systematicity*, or the methodical use of linguistic elements, or existence of patterns (e.g., ‘rules’); the presence of structured variation or “orderly heterogeneity” (Weinreich et al. 1968:100),
2. degree of *spread*, or proportion of speakers using the features/patterns at all within the community,
3. degree of *stability*, or the consistency in the use of certain features/patterns,
4. degree of *linguistic independence*, or the extent to which the variety’s features and patterns are not influenced by (or positively correlated with) other varieties,
5. *clustering*, or whether different (socio)linguistic features/patterns in the variety are correlated with each other, or whether sociolinguistic factors uniformly explain variation across features, and

6. *attitudes*, or speakers' perception of the variety they are using (e.g., whether they regard it as linguistically independent from the source languages).

A variety can fall anywhere on the multi-dimensional continuum defined by these variables. Instead of being categorized as being a language or not, a variety may be more language-like or less so, that is, having a certain degree of 'linguageness.' In this dissertation, I adopt the view that Lánnang-uè falls on the spectrum of linguageness parametrized by these hallmarks of 'language'.

It should be noted that my operationalization of linguageness (specifically, the choice of parameters) in this dissertation has been adapted to the Lannang community. All six criteria proposed here are relevant for addressing the question of Lánnang-uè's linguageness in the Lannang community but may be of less relevance in examining linguageness of speech varieties in other communities. For example, the criteria of spread as defined above is appropriate for the Lannang community in Manila, where there are a sizeable number of Lánnang-uè users. However, the criterion seems inappropriate for speech varieties that only have one or two speakers (e.g., endangered languages). Caution must be taken for researchers hoping to use the criteria proposed for Lánnang-uè for their own analyses.

2.4 Spread and stability

In this dissertation, I distinguish between notions of spread and stability. A linguistic feature has high rates of spread if it is used at all by most individuals or groups within a particular population of interest (e.g., the community) (Trudgill 1974; D'Arcy 2005:334). Examples of linguistic features (Goldberg 2017:9–10; Schmid 2020:97) considered to be widespread are compounds containing a Size and Shape Specifier (SASS) in Al-Sayyid Bedouin Sign Language (Meir and Sandler 2019:348) and the absence of mouthing in Israeli Sign Language (Tkachman and Meir 2018). These features are widespread in the community of young individuals who have hearing disabilities because they are used by most of these individuals (at least once).

While 'spread' is associated with the notion of distribution over the majority of people within a community, 'stability' is associated with consistency – a notion that is commonly used by many (variationist) sociolinguists working on linguistic data from a language development perspective (Trudgill 1974; Gordon et al. 2004:79; D'Arcy 2005; Kossmann 2019; Buchstaller et

al. 2021). A linguistic feature, variant, or pattern is highly stable if it is consistently used/followed by a speaker (limited intraspeaker variation) and/or if the patterns of variation are similar across speakers in the community (limited interspeaker variation) (Gordon et al. 2004): the feature/pattern “does not change significantly” and is “replicated” in multiple events (Da Silva 2010:816). An example of a highly stable linguistic feature is [ɪŋ] as used by a particular group of speakers in an English variety in Northern England, where the difference between the rates of using the variant in 1971 and the rates in 2013 is found to be insignificant individually and among speakers (Mechler and Buchstaller 2019:4). Another example is the [a:]-[æ] alternation in early New Zealand English, where the authors found the consistent use of the feature within the community of speakers across time (Gordon et al. 2004:136).

Chapter 3 : Description of Lánnang-uè

3.1 Preliminaries

In this chapter, I describe Lánnang-uè or Philippine Hybrid Hokkien. I begin by providing a snapshot of its lexicon (Section 3.2) followed by a description of the variety's phonetics and phonology (Section 3.3). Then, I cover linguistic elements that fall within the domain of the noun phrase (Section 3.4) and the verb phrase (Section 3.5), as well as the clause-level constituent order of the variety (Section 3.6). I conclude the section by describing prepositions, conjunctions, interjections, and discourse particles in Lánnang-uè (Section 3.8 to 3.11).

I draw on three major data sources to inform my description: corpus data, elicitation data, and judgment data. For the bulk of the description, I use data from the Lánnang-uè component of the Lannang Corpus (LanCorp) – a 375,000-word, part-of-speech-tagged databank of transcribed natural Lánnang-uè speech (e.g., narratives, interviews) with occasional code-switches collected from 135 Lannangs in metropolitan Manila between 2018 and 2020 (Gonzales 2022a).

I also use elicitation and judgment data (Chapter 1) to complement my corpus data, especially when describing features that cannot be exemplified using data from the LanCorp.

Because Lánnang-uè historically does not have an established convention for writing, I use the broad phonetic transcription conventions of the International Phonetic Association (IPA) in my description. In addition, to facilitate the reading comprehension of readers who are not familiar with the IPA, I also used a self-developed phonetic orthography called Lannang Orthography (henceforth, LO), a writing system that is adapted from the way many Lannangs write in Lánnang-uè on social media platforms and applications (Gonzales 2018; The Lannang Archives 2020).

3.2 Lexicon

The lexicon of Lánnang-uè comprises words sourced from Hokkien, Tagalog, and English, according to data from sixteen Lannangs who completed a Swadesh list (see example in

Appendix A). Out of the 219 words that illustrate the culturally independent vocabulary of the code or words least likely to be borrowed, around 49% are exclusively derived from Hokkien, 5% from Tagalog, and 15% from English. The word for ‘wide’, for instance, is almost exclusively expressed using the English-sourced variant *wîde*, whereas the word for ‘to sew’ is expressed as *tahî*, from Tagalog (see Appendix B). However, not all words in Lánnang-uè originate from a single source language. Approximately 31% of the basic lexicon tend to be expressed using variants from multiple source languages. The Lánnang-uè word for ‘worm’, for example, can be Tagalog-sourced *uôd* or English-sourced *wòrm*. Overall, around 72% of all the basic vocabulary of Lánnang-uè – exclusively sourced or not – can be expressed using the Hokkien-sourced variant, 29% using the Tagalog-sourced variant, and 37% using the English-derived variant.

Based on the lexical distribution by source language, Hokkien is the primary lexifier of the variety. The distribution also shows that Lánnang-uè does not source its basic vocabulary from Mandarin. Some technical and culture-specific words in Lánnang-uè, however, are sourced from Mandarin (e.g., *siaukhaî* ‘paper used for composition in Chinese’).

3.3 Phonetics and phonology

3.3.1 Phoneme inventory

The phoneme inventory of Lánnang-uè consists of 29 consonants and 10 vowels; vocoids can be combined to form diphthongs and triphthongs. There are three series of stops and affricates: voiceless unaspirated, voiceless aspirated, and voiced. Fricatives in Lánnang-uè have a voicing contrast. It has the following consonants:

Table 1. Consonants of Lánnang-uè

		labial	labio-dental	dental	alveolar	post-alveolar	palatal	velar	glottal
plosives	voiceless unaspirated	p			t			k	ʔ
	voiceless aspirated	p ^h			t ^h			k ^h	
	voiced	b			d			g	
affricates	voiceless unaspirated				ts	tʃ			
	voiceless aspirated				ts ^h				
	voiced					dʒ			
fricatives	voiceless		f	θ	s	ʃ			h
	voiced		v	ð	z				
nasals		m			n				
laterals					l			ŋ	
approximant					ɹ		j	w	

I provide some minimal pairs to illustrate the phonemic status of some consonants in (1).

(1) /p/ : /b/
pá [pa³⁵] ‘father’ : *bá* [ba³⁵] ‘yes/no question marker’

/p/ : /p^h/
pāng- [paŋ³³] ‘RES’ : *phāng* [p^haŋ³³] ‘fragrant’

/t/ : /t^h/
tò [to⁵¹] ‘spill’ : *thò* [t^ho⁵¹] ‘vomit’

/t/ : /d/
tân [tan⁵⁵] ‘wait’ : *dân* [dan⁵⁵] ‘3.PL.INC’

/ts/ : /ts^h/
tsè [tse⁵¹] ‘now’ : *tshè* [ts^he⁵¹] ‘to find’

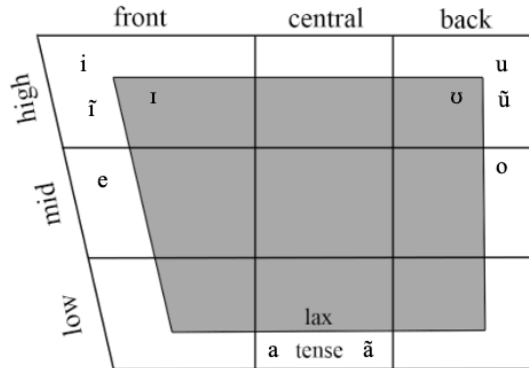
/s/ : /z/
sâp [sap⁵⁵] ‘sap’ : *zâp* [zap⁵⁵] ‘zap’

/ɹ/ : /l/
storè [stoɹ⁵¹] ‘store’ : *stàll* [stoɹ⁵¹] ‘stall’

/k/ : /g/
kôk [kok⁵⁵] ‘country’ : *côg* [kog⁵⁵] ‘cog’

Lánnang-uè vowel phonemes are given in Figure 2. Some minimal pairs can be found in (2). Many speakers do not have the nasalized vowels as part of their inventory.

Figure 2. Vowel phonemes of Lánnang-uè, adapted from Gonzales & Starr (2020)



(2) /i/ : /ɪ/
cheât [tʃiɪt⁵⁵] ‘cheat’ : *tshît* [tʃɪt⁵⁵] ‘wipe’; *seât* [sit⁵⁵] ‘seat’ : *sît* [sɪt⁵⁵] ‘things to do’

/i/ : /ĩ/
pí [pi³⁵] ‘compare’ : *pínn* [pĩ³⁵] ‘level (*pí* [pi³⁵] for many speakers)

/a/ : /ã/
kià [kja⁵¹] ‘send’ : *kiànn* [kjã⁵¹] ‘glass’ (*kià* [kja⁵¹] for many speakers)

/u/ : /ʊ/
suît [sut⁵⁵] ‘suit’ : *sût* [sʊt⁵⁵] ‘to whip’; *coôt* [kut⁵⁵] ‘coot’ : *kût* [kʊt⁵⁵] ‘bone’

There are 18 diphthongs and 2 triphthongs in Lánnang-uè formed by combining the vowel phonemes with offglides and/or onglides (i.e., /j/ and /w/). The lax vowels do not participate in such combinations. The diphthongs and triphthongs, along with examples, are given in Table 2 and (3), respectively.

Table 2. Diphthongs in Lánnang-uè

		i	ĩ	e	a	ã	o	u	ũ
j	onglide			/je/	/ja/		/jo/	/ju/	/jũ/
				<i>tshiēng</i> [ts ^h jeŋ ³³] 'clean'	<i>hiá</i> [hja ³⁵] 'there'		<i>siō</i> [sjo ³³] 'hot'	<i>siū</i> [sju ³³] 'fix'	<i>siūnn</i> [sjũ ³³] 'think'
	offglide			/ej/	/aj/		/oj/	/uj/	/ũj/
				<i>bày</i> [bej ⁵¹] 'bay'	<i>atây</i> [ʔa ³³ taj ⁵⁵] 'liver'		<i>sòy</i> [soj ⁵¹] 'soy'	<i>suí</i> [suj ⁵⁵] 'pretty'	<i>tsuinn</i> [tsũj ³⁵] 'before'
w	onglide	/wi/	/wĩ/	/we/	/wa/	/wã/			
		<i>bwisít</i> [bwi ³⁵ sit ⁵⁵] 'annoying'	<i>kuīnn</i> [kwĩ ³³] 'close'	<i>ué</i> [ʔwe ³⁵] 'shoe'	<i>huà</i> [hwa ⁵¹] 'hold'	<i>suānn</i> [swã ³³] 'mount'			
	offglide	/iw/			/aw/		/ow/		
		<i>sisíw</i> [si ³³ siw ⁵⁵] 'chick'			<i>hikâw</i> [hi ³³ kaw ⁵⁵] 'earrings'		<i>sòw</i> [sow ⁵¹] 'sow'		

(3) /jaw/
siaú
[sjaw⁵⁵]
'crazy'

/waj/
huái
[hwaj³⁵]
'those'

3.3.2 Phonotactics

The general syllable structure of Lánnang-uè is (C)(C)(C)V(C)(C)(C). The maximal syllable type is CCCVCC while the minimal is V, where V can be a syllabic contoid.¹⁴ Table 3 shows the possible ways in which the segments can be arranged, with at least one example for each permutation.

¹⁴ 'Syllabic' here means 'functions as a vowel'. 'Contoid' refers to a speech sound with significant obstruction in the vocal tract above the glottis.

Table 3. Syllable permutations found in Lánnang-uè with examples

Syllable type		Word	IPA	Gloss
Open syllables	V	<i>m sī</i>	[m ³³ si ³³]	is not
	CV	<i>writèr</i>	[.iaj ³³ tɿ ⁵¹]	writer
	CCV	<i>ué</i>	[ʔwe ³⁵]	shoe
	CCCV	<i>straw</i>	[st.ɿo ⁵¹]	straw
Closed syllables	CVC	<i>tahî</i>	[ta ³³ hi ⁵⁵]	to sew
		<i>enjòy</i>	[ʔen ³³ dʒoj ⁵¹]	to enjoy
		<i>um sī</i>	[ʔom ³³ si ³³]	is not
	CVCC	<i>coûrt</i>	[ko.ɿt ⁵⁵]	court
	CCVC	<i>stalk</i>	[stok ⁵⁵]	stalk
		<i>plòw</i>	[plow ⁵¹]	plow
		<i>tshiēng</i>	[ts ^h jeŋ ³³]	clean/thousand
	CCVCC	<i>stârt</i>	[stɑ.ɿt ⁵⁵]	start
<i>plânt</i>		[plant ⁵⁵]	plant	
CCCVC	<i>stròng</i>	[st.ɿoŋ ⁵¹]	strong	
	<i>splattèr</i>	[spla ³³ tɿ ⁵¹]	splatter	
CCCVCC	<i>strând</i>	[st.ɿand ⁵⁵]	strand	
CVCCC	<i>síxths</i>	[sɿksts ⁵⁵]	sixths	

It is possible that there are CCCVCCC syllables in Lánnang-uè, but there are no such examples in my database.

Some phonotactic rules of Lánnang-uè are as follows:

1. All syllables must have a nucleus. In the case of syllables with a syllabic contoid, the syllabic contoid – functionally the vowel – is the nucleus (e.g., [ʔm³³ tsaj³³] ‘do not know’).
2. While codas are optional, onsets are obligatory. The only exception is when the syllable has a contoid for a nucleus (e.g., [m³³ si³³]).
3. The only contoids that appear as nuclei of one-segment syllables are /m/ and /ŋ/.
4. The only contoids that appear as the nucleus of CV syllables are /m/ and /ŋ/.
5. In syllables with a CC onset, a sonorant cannot precede an obstruent (e.g., [wk], [np], [lt]).

6. In syllables with a CCC onset, the first consonant must be /s/. The second consonant must be one of the following voiceless unaspirated stops: /p/, /t/, /k/. The third must be either /l/ or /ɹ/.
7. All consonants except aspirated plosives, aspirated affricates, and /h/ can appear in codas.

Lánnang-uè words with the V structure originate from Hokkien. Syllables with CCC onsets or offsets are found in English-sourced words.

3.3.3 *Suprasegmentals*

The suprasegmental system of Lánnang-uè comprises a tone subsystem and a stress subsystem. The former (applying to all Lánnang-uè words regardless of word origin) relies on pitch to distinguish meaning between words, while the latter (applying to English- and Tagalog-sourced words)¹⁵ utilizes duration.

3.3.3.1 Tone

Lánnang-uè has eight phonemic tones: mid, high-I, high-II, high-III, falling-I, falling-II, rising, and neutral.¹⁶ Each of these phonemes is associated with a distinct set of allophones. This is shown in Table 4 along with some examples. The high-I tone (second column), for instance, is only phonetically realized in the variety as a high tone. In contrast, the high-II tone can be phonetically realized as a high tone, a low tone, or a rising tone, depending on the tone of syllable after it. As illustrated in the last row of Table 4, the seven phonemic tones can be used to distinguish meaning in words: *kaū* [kaw³³] with a mid tone, for example, means ‘submit’ whereas *kaú* [kaw³⁵] with a rising tone means ‘monkey’.

¹⁵ It is possible that Hokkien- and Mandarin-derived words may have stress that is primarily cued by duration, but this is something that needs to be investigated.

¹⁶ A neutral tone is a tone that does not have a fixed pitch or pitch range, hence ‘neutral’. Its exact pitch is dependent on the tone that came before it, unlike other tones that are not.

Table 4. Tones in Lánnang-uè

Tone	Mid (M)	High-I (H1)	High-II (H2)	High-III (H3)	Falling-I (F1)	Falling-II (F2)	Rising (R)	Neutral (N)
Phoneme	/33/	/55/	/55/	/55/	/51/	/51/	/35/	/X1/
Allophone(s)	[33] [22]	[55]	[55] [22] [35]	[55] [22] [33]	[51] [55]	[51] [22] [33]	[35] [22] [33]	[11] [31] [51]
Example(s) (Lannang Orthography, IPA, gloss)	<i>siōsiō</i> [sjo ³³ sjo ³³] 'fever' <i>sio-â</i> [sjo ²² a ⁵⁵] 'roasted duck'	<i>tshîttshît</i> [ts ^h it ⁵⁵ ts ^h it ⁵⁵] 'wipe'	<i>tê</i> [te ⁵⁵] 'short' <i>tetê</i> [te ²² te ⁵⁵] 'really short' <i>té thaumúng</i> [te ³⁵ h ^{aw} 22 muŋ ³⁵] 'short hair'	<i>dôg</i> [dog ⁵⁵] 'dog' <i>doghoûse</i> [dog ²² haws ⁵⁵] 'doghouse' <i>dôggy</i> [dog ³³ gi ³⁵] 'doggy'	<i>sì</i> [si ⁵¹] 'four' <i>sîpâh</i> [si ⁵⁵ pa ²⁵⁵] 'four hundred'	<i>phî</i> [p ^h i ⁵¹] 'nose' <i>phîsaî</i> [p ^h i ²² saj ⁵⁵] 'nose' <i>phîsaî</i> [p ^h i ³³ saj ⁵⁵] 'nose'	<i>tít</i> [tit ³⁵] 'straight' <i>títít</i> [tit ²² tit ³⁵] 'really straight' <i>Huîdîpîn</i> [hwi ³³ di ³³ pin ³³] 'Philippines'	<i>lo</i> [lo ¹¹] 'PFV' <i>lo</i> [lo ³¹] 'PFV' <i>lo</i> [lo ⁵¹] 'PFV'
Minimal pair/set example(s)	<i>kaû</i> [kaw ³³] 'submit'		<i>kaû</i> [kaw ⁵⁵] 'dog' <i>kausaî</i> [kaw ²² saj ⁵⁵] 'dog feces' <i>kaú</i> [kaw ³⁵ məŋ ³⁵] 'dog hair'	<i>hikâw</i> [hi ²² kaw ⁵⁵] 'earrings'	<i>kaú</i> [kaw ⁵¹] 'arrive' <i>kaû gûn</i> [kaw ⁵⁵ gun ⁵⁵] 'arrive to us'	<i>còw</i> [kaw ⁵¹] 'cow' <i>còwbôys</i> [kaw ²² bojs ⁵⁵] 'cowboys' <i>còwbôys</i> [kaw ³³ bojs ⁵⁵] 'cowboys'	<i>kaú</i> [kaw ³⁵] 'monkey' <i>kausaî</i> [kaw ²² saj ⁵⁵] 'monkey feces'	

Lánnang-uè words or syllables sourced from Hokkien¹⁷ can have any of the eight tones except the high-III tone. Those derived from Tagalog or English,¹⁸ however, can only have either the high-III tone or the falling-II tone (and in rare cases, the rising tone). If the syllable is closed and in a word that is derived from Tagalog, that syllable has the high-III tone (e.g., *sampâl* /sam⁵⁵ pal⁵⁵/ ‘slap’). If the syllable is open and in a Tagalog-origin word, it has the falling-II tone (e.g., *basù* /ba⁵¹ su⁵¹/ ‘cup’). On the other hand, if the syllable is found in a word that is sourced from English, that syllable generally has the falling-II tone (e.g., *buttòn* /ba⁵¹ ton⁵¹/ ‘button’) – the exception is if this syllable ends with an obstruent. Then, the syllable has the high-III tone (e.g., *toothpîck* /tut⁵⁵ pik⁵⁵/ ‘toothpick’).

Variation in tone assignment appears to be minimal in Lánnang-uè. Based on preliminary analyses, only a few speakers occasionally vary their use of the tone conventions, either producing non-conventional lexical tones for the word or producing them with English- or Tagalog-like (non-lexical) pitch. Some speakers, for example, use high tones for syllables in Tagalog- and English-derived words that are conventionally produced with a falling tone (e.g., *basû* /ba⁵⁵ su⁵⁵/ ‘cup’ instead of *basù* /ba⁵¹ su⁵¹/).

Lánnang-uè has right-dominant tone sandhi. The tone of a word or syllable can change depending on the tones of the surrounding words or syllables. In general, the final syllable or word in the tonal phrase (e.g., the NP, the VP) preserves its tone. The exception is if the final syllable in the phrase has a neutral tone. All non-final syllables that fit the conditions specified in the tone sandhi rules undergo tonal changes. Here are the rules pertaining to tone sandhi in Lánnang-uè:

1. HI /55/ → H [55]

High-I tones are exempt from tone sandhi changes.

(4)	/pweʔ⁵⁵/	+	/ts^hjeŋ³³/	→	[pweʔ⁵⁵ ts^hjeŋ³³]	<i>puêh tshiēng</i>
	HI		M		H M	‘8000’
	/pweʔ⁵⁵/	+	/paʔ⁵⁵/	→	[pweʔ⁵⁵ paʔ⁵⁵]	<i>puêh pâh</i>
	HI		HI		H HI	‘800’

¹⁷ Hokkien does not have the high-III tone.

¹⁸ Tagalog and English are not tone languages.

2. M /33/ → L [22] / {H, F}
 → M [33] /elsewhere

Mid tones typically become low tones before high and falling tones within the tonal phrase; they are realized as mid tones elsewhere.

(5)	/tsin³³/ M	+	/suj⁵⁵/ HII	→	[tsin²² suj ⁵⁵] L H	<i>tsin suí</i> 'really pretty'
	/tsin³³/ M	+	/gej⁵¹/ FII	→	[tsin²² gej ⁵¹] L F	<i>tsin gày</i> 'really gay'
	/tsin³³/ M	+	/pjeŋ³³/ M	→	[tsin³³ pjeŋ ³³] M M	<i>tsīn piēng</i> 'really cold'
	/tsin³³/ M	+	/təŋ³⁵/ R	→	[tsin³³ təŋ ³⁵] M R	<i>tsīn túng</i> 'really long'

Some speakers do not follow this rule and produce a syllable with /33/ as [33] even if the following syllable has a high or falling tone.

3. R /35/ → L [22] / _ {H, F}
 → M [33] / _ {M, R}
 → R [35] / elsewhere

Rising tones typically become low tones before high or falling tones within the tonal phrase.

(6)	/bo³⁵/ R	+	/soŋ⁵⁵/ HII	→	[bo²² soŋ ⁵⁵] L H	<i>bo sôŋ</i> 'not comfortable'
	/bo³⁵/ R	+	/gej⁵¹/ FII	→	[bo²² gej ⁵¹] L F	<i>bo gày</i> 'not gay'

They become mid tones before mid or rising tones within the tonal phrase.

- (7) $/bo^{35}/$ + $/pjeŋ^{33}/$ → $[bo^{33} pjeŋ^{33}]$ *bō piēng*
R M **M** M 'not cold'
- $/bo^{35}/$ + $/gaw^{35}/$ → $[bo^{33} gaw^{35}]$ *bō gau*
R R **M** R 'not smart'

They are still realized as a rising tone in all other cases (i.e., at the end of the phrase, in isolation).

- (8) $/bo^{35}/$ → $[bo^{35}]$ *bó*
R **R** 'none'

4. HII /55/ → R [35] / _ {M, R}
→ L [22] / _ {H, F}
→ H [55] / elsewhere

High-II tones become rising tones before mid tones or rising tones within the tonal phrase.

- (9) $/ja^{55}/$ + $/pieŋ^{33}/$ → $[ja^{35} pieŋ^{33}]$ *yá piēng*
HII M **R** M 'very cold'
- $/ja^{55}/$ + $/təŋ^{35}/$ → $[ja^{35} təŋ^{35}]$ *yá túng*
HII R **R** R 'very long'

They become low tones before high tones or falling tones within the tonal phrase.

- (10) $/ja^{55}/$ + $/kuj^{51}/$ → $[ja^{22} kuj^{51}]$ *ya kuì*
HII FII **L** F 'very expensive'
- $/ja^{55}/$ + $/suj^{55}/$ → $[ja^{22} suj^{55}]$ *ya suí*
HII HII **L** H 'very beautiful'

High-II tones are realized as a high tone in all other cases (i.e., before neutral tones).

(11) /**si**⁵⁵ lo^{X1}/ → [**si**⁵⁵ lo⁵¹]
HIII N **H** F

Sí lò.
 ‘Die./That’s it.’

5. **HIII** /55/ → L [22] or M [33] / _ tone
 → H [55] / _ elsewhere

High-III tones become low or mid tones before any tone within the tonal phrase. Based on my analyses, the low and mid phonetic realizations are in free variation.

(12) /**bejs**⁵⁵/ + /bo⁵¹/ → [**bejs**^{22/33} bo⁵¹] *bāsebàll/bāsebàll*
HIII FII **L/M** F ‘baseball’
 /**sam**⁵⁵/ + /pa⁵⁵/ → [**sam**^{22/33} pa⁵⁵] *sāmpâl/sāmpâl*
HIII HIII **L/M** H ‘slap’
 /**tut**⁵⁵/ + /pik⁵⁵/ → [**tut**^{22/33} pik⁵⁵] *toōthpîck/toōthpîck*
HIII HIII **L/M** H ‘toothpick’

They are realized as a high tone in all other cases (e.g., at the end of the phrase, in isolation).

(13) /**tut**⁵⁵/ → [**tut**⁵⁵] *toōth*
HIII **H** ‘tooth’

6. **FI** /51/ → H [55] / _ tone
 → F [51] / elsewhere

Falling-I tones become high tones before any tone within the tonal phrase.

(14) /**si**⁵¹/ + /ts^hjeŋ³³/ → [**si**⁵⁵ ts^hjeŋ³³] *sî tshiēng*
FI M **H** M ‘4000’
 /**si**⁵¹/ + /pa²⁵⁵/ → [**si**⁵⁵ pa²⁵⁵] *sî pâh*
FI HI **H** H ‘400’

They are realized as a falling tone in all other cases (e.g., at the end of the phrase, in isolation).

(15) /si⁵¹/ → [si⁵¹] si
FI **F** '4'

7. FII /51/ → L [22] or [33] / _ tone
→ F [51] / elsewhere

Falling-II tones become low or mid tones before any tone within the tonal phrase. Based on my data, the low and mid phonetic realizations are in free variation.

(16)	/si ⁵¹ /	+	/pa ⁵⁵ /	→	[si ^{22/33}	pa ⁵⁵]	<i>sipâ/sîpâ</i>
	FII		HIII		L/M	H	'kick'
	/twa ⁵¹ /	+	/sja ³³ /	→	[twa ^{22/33}	sja ³³]	<i>tuasiâ/tuâsiâ</i>
	FII		M		L/M	M	'loud sound'
	/ba ⁵¹ /	+	/su ⁵¹ /	→	[ba ^{22/33}	su ⁵¹]	<i>basù/bâsù</i>
	FII		FII		L/M	F	'cup'
	/ba ⁵¹ /	+	/ton ⁵¹ /	→	[ba ^{22/33}	ton ⁵¹]	<i>batòn/bâtòn</i>
	FII		FII		L/M	F	'baton'

They are realized as a falling tone at the end of the phrase or in isolation.

(17) /ja⁵⁵/ + /twa⁵¹/ → [ja^{22/33} twa⁵¹] tuà
HIII **FII** **H** **F** 'big'

(18) /twa⁵¹/ → [twa⁵¹] tuà
FII **F** 'big'

8. The tone in monosyllabic personal pronouns (e.g., *dî* '2.SG', *gûn* '1.PL.EXC') located in a simple noun phrase at the end of the sentence changes to neutral tone N [11].

- (19) /bok³⁵ su³³ p^haɿ⁵⁵ di⁵⁵/ →
 R M HI HI
- bok³³ su³³ p^haɿ⁵⁵ di⁵⁵ →
 M M H H
- bok³³ su³³ p^haɿ⁵⁵ di^{X1}
 M M H N

Bōksū phâh dî.
 ‘Pastor hit you.’

9. N /X1/ → F [51] / {H, R} _ ||
 → F_{low} [31] / M _ ||
 → L_{ultra} [11] / elsewhere

After rule 8 has been applied (if needed), the neutral tones change based on the final pitch of the preceding tone. If a high or rising tone precedes the neutral tone and if a phonological phrase boundary succeeds it, the tone changes to a (high) falling tone [51].

- (20) /pa⁵¹taj⁵⁵/ + /lo^{X1}/ → [pa³³taj⁵⁵ lo⁵¹] *Patây lò.*
 FII HIII N M H F ‘Dead. / That’s it.’

- (21) bok³³ su³³ p^haɿ⁵⁵ di^{X1} →
 M M HI N
- [bok³³ su³³ p^haɿ⁵⁵ di⁵¹]
 M M HI F

Bōksū phâh dî.
 ‘Pastor hit you.’

If the neutral tone is preceded by a mid tone and if a phonological phrase boundary succeeds it, the neutral tone changes to a (low) falling tone [31].

- (22) /si⁵⁵ kwe³³/ + /lo^{X1}/ → [si³⁵ kwe³³ lo³¹]
 HII M N R M F_{low}
Sí kuē lò.
 ‘Dead chicken. / That’s it.’

Elsewhere, such as in cases where the neutral tone is preceded by a falling tone, the tone changes to an ultra-low tone [11].

- (23) /ts^hju⁵¹/ + /lo^{X1}/ → [ts^hju⁵¹ lo¹¹]
 FI N R **L**ultra
Tshiù lo.
 ‘(They) sang already.’

Unlike Hokkien-, English-, and Tagalog-derived words, words sourced from Mandarin do not follow the eight-tone contrast or sandhi rules in Lánnang-uè and instead follow the five-tone contrast (Table 5) and rules found in Mandarin.¹⁹

Table 5. Tones in Mandarin and Mandarin-sourced words in Lánnang-uè (words are in Pinyin orthography)

Tones of Mandarin-origin words in Lánnang-uè	1	2	3	4	5
Phoneme	/55/	/35/	/313/	/51/	/33/
Allophone(s)	[55]	[35]	[31] [13] [11] [313]	[51]	[33] [11]
Example	<i>wěibā</i> [wei ³¹ pa ⁵⁵] ‘tail’	<i>bá</i> [pa ³⁵] ‘pull out’	<i>bǔ</i> [pa ³¹³] ‘put’	<i>bà</i> [pa ⁵¹] ‘father’	<i>bā</i> [pa ³³] ‘PRT’

3.3.3.2 Stress

Lánnang-uè words of Tagalog and English origin have lexical stress (henceforth, stress), meaning that stress or the absence of it can distinguish the meaning of a word. Take, for example, *sandâl* [san³³ dal⁵⁵]. Placing the stress on the first syllable, as in *sandâl* [ˈsan³³ dal⁵⁵], produces a word meaning ‘sandal’, while placing the stress on the final syllable like in

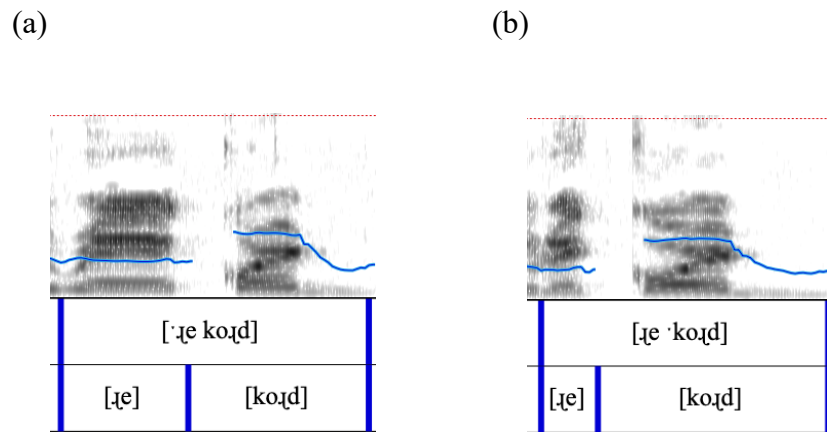
¹⁹ The data shows the existence of two distinct tone systems in Lánnang-uè, Lánnang-uè tone and Mandarin tone, which could be support for the hypothesis that Lánnang-uè is low on the languageness scale. However, I argue that Mandarin words are unassimilated (non-nativized) loanwords and not an integral part of the Lánnang-uè lexicon (see 3.2). This argument is warranted, given that Mandarin in the Lannang community is almost always introduced via formal schooling.

[san³³ dal⁵⁵] produces a word meaning ‘to lean on’. Other minimal pairs that illustrate the stress contrast are shown in (24).

(24)	<i>bakà</i>	[‘ba ³³ ka ⁵²] ‘cow’	[ba ³³ ‘ka ⁵²] ‘maybe’
	<i>lutô</i>	[‘lu ³³ toʔ ⁵⁵] ‘cook’	[lu ³³ toʔ ⁵⁵] ‘cooked’
	<i>recôrd</i>	[‘ɿe ³³ koɿd ⁵⁵] ‘document/music’	[ɿe ³³ ‘koɿd ⁵⁵] ‘to record (action)’

Lánnang-uè stress is exclusively cued by duration (see Chapter 4). I illustrate this for the *recôrd* pair in Figure 3. Comparing between syllables across the two words, the stressed [ɿe] (left) is 0.162 seconds longer than the unstressed [ɿe] (right). The stressed [koɿd] (right) is 0.145 seconds longer than the unstressed [koɿd] (left).

Figure 3. Acoustic comparison of duration between a minimal pair in Lánnang-uè: (a) *recôrd* [‘ɿe³³ koɿd⁵⁵] vs. (b) *recôrd* [ɿe³³ ‘koɿd⁵⁵]



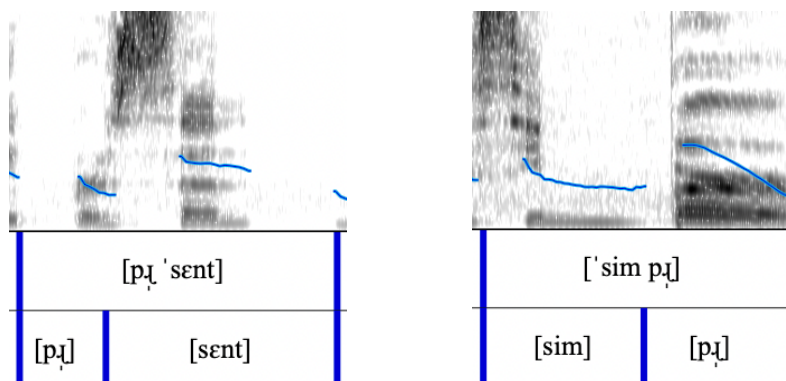
Based on preliminary analyses, variation in the use of stress is present in the variety (see also Chapter 4). There are a couple of speakers, for example, who sometimes do not only assign stress to Tagalog- and English-sourced words, but also appear to assign stress to certain Hokkien-origin words (e.g. *chāmā* [tʃa³³ ‘ma⁵⁵] ‘fork’, *tiēnau* [tje³³ ‘naw⁵⁵] ‘computer’, *shammih* [ʃa²² miʔ³⁵]). Some speakers also occasionally appear not to follow the duration-cued stress patterns – for example, they sometimes do not stress (lengthen) syllables that are meant to be stressed or unstress (shorten) syllables that are unstressed.

3.3.3.3 Phrase-final lengthening

The final syllable of a phrase is lengthened regardless of the origin of the word. If the phrase only consists of one word, as is the case for the two words in the *recôrd* pair in Figure 3, then the final syllable of the word that comprises the phrase is lengthened. For example, in Figure 3a, the second syllable in *recôrd* [³³ʁe³³ ko.ɔd⁵⁵] is expected to be short because of lack of stress. It is, however, long, due to phrase-final lengthening. In the Figure 3b, the second syllable in *recôrd* [³³ʁe³³ ko.ɔd⁵⁵] is long(er) than the second syllable in Figure 3a because it has both phrase-final lengthening and stress (lengthening).

A syllable that has neither phrasal lengthening (located in a non-phrase final position) nor lexical stress is produced exceptionally short as opposed to just short. For example, the non-phrasally lengthened and lexically unstressed [pɪ] in *percênt* is much shorter (0.179) than the [pɪ] in *simpèr* (0.36), which only lacks lexical stress (Figure 4). A phrase-final stressed syllable is also much longer than a non-phrase-final stressed syllable.

Figure 4. Acoustic comparison of duration between *percênt* [pɪ'sent] and *simpèr* ['sim pɪ]



3.4 Noun phrases

In this section, I identify and describe linguistic elements within the noun phrase (NP): determiners, pluralization, interrogatives, adjectives, relative clause modification, and pronouns. The section concludes with a description of affixes associated with the NP.

3.4.1 The determiner system

Lánnang-uè determiners include demonstratives and articles. Four morphemes function as demonstratives in Lánnang-uè, all of them sourced from Hokkien: *tsí* ‘this’ /tsi⁵⁵/, *tsuai* ‘these’ /tswaj³⁵/, *hí* ‘that’ /hi⁵⁵/, and *huai* ‘those’ /hwaj³⁵/ . As demonstratives, they specify the proximity of the noun in relation to the speaker. The first two are proximal while the other two are distal. Lánnang-uè demonstrative determiners have a number contrast. *Tsí* and *hí* are used with singular referents; *tsuai* and *huai* are used with plural referents. The determiners precede the head noun, as in examples (25) to (28).

- (25) tsi⁵⁵ ge²² laŋ³⁵ ja³⁵ 'di³³.ii⁷⁵⁵ e⁵⁵
Tsí ge lánɡ yá dīrī ēh.
 this CLS person very disgusting PRT
 ‘This person is very disgusting.’
 (CFH-001)²⁰
- (26) ba³³.ka³³ tswaj³⁵ loŋ³⁵tsoŋ⁵¹ kan²².ja²²kan²².ja⁵¹
Bākā *tsuai* lóngtsòŋ kanyakanyà.
 maybe these all on.their.own
 ‘Maybe all of these will do things by themselves.’
 (CD-001)
- (27) hi⁵⁵ gi³³na⁵⁵ tsa⁵⁵ pe⁷⁵⁵ khi⁵⁵ tsi³³ge³³ twa³³e³³ 'rak⁵⁵
Hí gīnā tsá pēh -khī tsīgē tuā=ē rôck.
 that kid then climb -DIR DET big=MOD rock
 ‘That kid then climbed on top of a big rock.’
 (PC0009-FRST19)
- (28) hwaj³⁵ ma²².ŋa²² 'ra²²bit⁵⁵ le⁵⁵ ts^hoŋ⁵¹ ja²² ba²²
Huai mɡa rabbit lê tshòŋ sha ba?
 those PL rabbit PROG do what PRT
 ‘What are those rabbits doing?’
 (elicitation, PC0068)

Some speakers occasionally use Tagalog-sourced *yung* ‘that/those’ /juŋ⁵⁵/ instead of *hí* ‘that’ or *huai* ‘those’.

²⁰ Unless I specify that the example is elicited, enclosed in parentheses is the file name of the recording/transcription in which the example is found.

- (29) juŋ²² ma³³ŋa³³ le⁵⁵ laj²² tsja³⁵ o³⁵ si²² an⁵⁵ 'pja²²vins⁵⁵
Yūng mgā lē lai tsiá ó si ân province
 DEM PL PROG come here PRT COP PREP province

la²²je⁵¹
 lai=è.
 come =MOD

‘Those who are coming here are from the province.’
 (C-001)

They sometimes use English-sourced *this* ‘this’ /dis⁵⁵/ and *thât* ‘that’ /dat⁵⁵/ idiomatically:

- (30) 'pe²²ro²² dis²² tajm⁵¹ k^ha⁵⁵laŋ⁵⁵ kon²²vɿ²²sej²²fon⁵¹ laŋ⁵⁵ ta²²la²²ga⁵¹
 Pero **this** time khâlâng conversation lâng talagà.
 but DEM time like conversation only really
 ‘But this time, it is really only like a conversation.’
 (PC0001-FRST18)

Lánnang-uè has two articles – definite and indefinite. They indicate whether the following noun is specific or non-specific. In the definite class, we have singular *hîgé* ‘the’ /hi⁵⁵ge³⁵/ (31) and plural *huai* ‘the’ /hwaj³⁵/ (32). They indicate that the speaker is referring to a particular entity.

- (31) din⁵⁵ tio²² te⁵⁵ hi⁵⁵ge²² 'sta²²di²² 'tuɿ⁵¹
 Dîn tìoh tē **hîge** study tòur.
 2.PL should follow ART.DEF.SG study tour
 ‘You should join the study tour group.’
 (E-001)

- (32) k^ha⁵⁵nan⁵⁵ ts^hju³³ lan⁵⁵ k^ha⁵⁵ ʔu²² le⁵⁵ tʃjaw⁵⁵
 Khânân tshīū lân khâ u lê chiaú
 as like 1.PL.INC CMPV PF PROG follow
 hwaj²² 'ruls⁵⁵
huai rûle-s.
 ART.DEF.PL rule-PL

‘It is like we are better at following the rules.’
 (PC0009-FRST19)

Lánnang-uè has *tsîgé* ‘a’ /tsi³³ge³⁵/ as an indefinite article, derived from Hokkien *tsit ge* ‘one CLF’. It is used when the speaker wants to refer to any entity in a set and does not care which one it is. For example, in (33), the speaker is asking the listener whether the entity (i.e., Lánnang-

uè) can be considered a ‘variety’ in general, not whether the entity is a specific variety of a certain language.

- (33) tsi⁵⁵ ge³⁵ si³³ tsi³³ge³³ va²²ra²²je²²ti⁵¹ ba³⁵
Tsí gé sī tsīgē variety bá?
 DEM CLS COP ART.INDEF variety Q
 ‘Is this a variety?’
 (PC0019-CLIN18)

Some speakers occasionally use the English-derived definite article *thè* ‘the’ /di⁵¹/ and indefinite article *à* ‘a’ /ʔa⁵¹/, but only if the head noun is derived from English as well.

- (34) gwa⁵⁵ tio^{ʔ22} pas⁵⁵ ʔa²²kios²² da²² 'rowd⁵⁵ ko⁵⁵
Guâ tìoh pàss across the roâd kô.
 1.SG NEC pass PREP ART.DEF road PRT
 ‘I should pass across the road.’
 (CLIN-19-68:40519)

To refer to general concepts (e.g., *lôve* ‘love’, *láng* ‘humans’), articles are not used. They are used with mass nouns like *wine* ‘wine’ and *tsuí* ‘water’ if the speaker wants to express specificity. Some older speakers occasionally do not use articles with mass nouns at all. Lánnang-uè articles cannot be analyzed as demonstrative determiners because they do not indicate degree of proximity. For example, in (31), the speaker is asking the listener to join a specific tour group, not a tour group physically distant from or near to them.

3.4.2 Classifiers

Classifiers in Lánnang-uè – sourced from Hokkien – categorize or sort a noun based on its “formal or semantic class” (e.g., animals, long objects, flat objects) (Crystal 2008:78). For example, in (35), the classifier *tsiâh* /tsja^{ʔ55}/ categorizes the noun *deèr* ‘deer’ as an animal. In (36), the classifier *é* /ʔe³⁵/ or *gé* /ge³⁵/ indicates that *láng* ‘person’ is an unspecified entity. A complete list²¹ of classifiers used in Lánnang-uè alongside the semantic classes they are associated with and some examples is given in Table 6.

²¹ Some of the classifiers from this list were acquired from the corpus database. The rest were elicited from five native speakers of Lánnang-uè. This list is, to my knowledge, complete.

Table 6. The classifiers of Lánnang-uè

Classifier	IPA	Semantic class(es)	Example
<i>bê</i>	/be ⁵⁵ /	seafood	<i>tsi bê hipôn</i> 'one shrimp (that is seafood)'
<i>diáp</i>	/djap ³⁵ /	small, round entities	<i>sa diáp yemà</i> 'three yema candies (that are small and round entities)'
<i>é/gé</i>	/ʔe ³⁵ / /ge ³⁵ /	entity (general, unspecified)	<i>nūng ē rabbît</i> 'two rabbits (that are unspecified entities)' <i>nūng ē spirít</i> 'two spirits (that are unspecified entities)'
<i>hāng</i>	/haŋ ³³ /	abstract entities	<i>sā hāng mīngkiā</i> 'three things (that are abstract entities)'
<i>khō</i>	/k ^h o ³³ /	money	<i>sa khō pesò</i> 'three pesos (that is money)'
<i>khuân</i>	/k ^h wan ⁵⁵ /	kind	<i>nūng khuân hospitàl</i> 'two hospitals (that is of a certain kind)'
<i>kī</i>	/ki ³³ /	long, rigid entities	<i>tsi kī ballpèn</i> 'one pen (that is a long and rigid entity)' <i>sā kī cellphone</i> 'three mobile phones (that is a long and rigid entity)'
<i>kiēng</i>	/kjeŋ ³³ /	buildings	<i>nūng kiēng hospitàl</i> 'two hospitals (that are buildings)'
<i>kô</i>	/ko ⁵⁵ /	months	<i>sa kô gé</i> 'three moons (that are months)'
<i>niâ</i>	/nja ⁵⁵ /	clothing covering human body	<i>sā niâ sā</i> 'three shirts (that are clothing for humans)'
<i>pùn</i>	/pun ⁵⁵ /	entities that have pages	<i>sa pùn dictionary</i> 'three dictionaries (that are entities with pages)' <i>sa pùn journal</i> 'three journals (that are entities with pages)'
<i>tâh</i>	/ta ^h ⁵⁵ /	place	<i>tsi tâh sótsai</i> 'one place (that is a place)'
<i>tè</i>	/te ⁵¹ /	lumpy, crumbly, chunk- or piece-like entities	<i>sā tê meteorîte</i> 'three meteorites (that look like chunks)' <i>sā tê cāke</i> 'three cakes (that are crumbly entities)'
<i>tiaú</i>	/tjaw ³⁵ /	strand-like entities that are flexible, wavy, or can be bent	<i>sa tiaú miswà</i> 'three wheat vermicelli noodles (that are strand-like entities)'
<i>tiū</i>	/tju ³³ /	flat entities	<i>sā tiū banana leáf</i> 'three banana leaves (that are flat entities)'
<i>tsāng</i>	/tsaŋ ³³ /	trees and shrubs	<i>tsī tsāng Narrà chù</i> 'one Narra tree (that is a tree)' <i>sā tsāng bamboò</i> 'three bamboo plants (that are like trees)'
<i>tsiâh</i>	/tsja ^h ⁵⁵ /	animals; animal-like entities	<i>nūng tsiâh rabbît</i> 'two rabbits (that are animals)' <i>hī tsiâh ginâ</i> 'that kid (who is animal-like)'
<i>tūng/tiēng</i>	/təŋ ⁵⁵ / /tjeŋ ⁵⁵ /	vehicles	<i>sa tūng/tiēng bicycle</i> 'three bicycles (that are vehicles)'

The default and most commonly used classifier in Lánnang-uè is *é* /eʔ³⁵/ or *gé* /ge³⁵/. *É* is the general form, while *gé* is only used after the numeral one *tsít* or *tsí*. Classifiers are used before the noun and after numerals or demonstratives, as shown in Table 6, (35), and (36).

- (35) tsi⁵⁵ tsja^{ʔ55} 'diɾ⁵¹ tsju³³ kwa³³ kin⁵⁵ tsaw⁵⁵ lo⁵¹
Tsí tsiáh deèr tsiū kuā kîn tsaú lò.
 this CLS deer at.once rush fast run PRT
 ‘This deer at once quickly ran away.’
 (PC0001-FRST18)

- (36) m̄²² tsa³³ʔja⁵⁵ kuŋ²² si²² tsi²² ge²² laŋ³⁵ ʔoɾ²²
M tsā-iâ kung si tsi ge lánŋ or...
 NEG know if COP one CLS person or
 ‘I do not know if it is just one person or...’
 (CFH-001)

3.4.3 Quantifiers

The ten quantifiers of Lánnang-uè are: *longtsòŋ* /loŋ²²tsòŋ⁵¹/ ‘all (general)’, *kê* /ke⁵⁵/ ‘all (general)’, *tagé* /ta²²ge³⁵/ ‘all (human)’, *tsikuá* /tsi²²kwa³⁵/ ‘some’, *uwé* /ʔu²²we³⁵/ ‘some’, *tampóh* /tam²²po^{ʔ35}/ ‘little (non-countable)’, *tsiô* /tsjo⁵⁵/ ‘few’, *tsuè* /tswe⁵¹/ ‘many’, *bó* /bo³⁵/ ‘none’, and *mui* /muj³⁵/ ‘each/every’. All are derived from Hokkien. They “indicate [the] quantity or scope” of the head noun of a noun phrase (Schachter and Shopen 2007:37; Crystal 2008). Only *longtsòŋ* and *kê* are placed after the head noun; all other quantifiers are placed before the head noun. I have yet to find conditioning factors (e.g., semantic, pragmatic, social) for the use of the two ‘all (general)’ quantifiers or the two ‘some’ quantifiers.

The following examples illustrate how the quantifiers are used:

- (37) din⁵⁵ e³³ pa²²pa³⁵ ʔa⁵⁵si³³ma²²ma³⁵ loŋ²²tsòŋ⁵¹ si²² ʔan⁵⁵
Dîn =ē papá âsī mamá longtsòŋ si ân
 2.PL GEN dad or mom all COP PREP
 tsja³⁵ ts^hut⁵⁵si⁵¹
tsiá tshùtsi?
 here born

‘Are all of your moms or dads born here?’
 (PC0097-CLIN19)

(38) ʔin⁵⁵ nəŋ²² e³⁵ ke⁵⁵ saŋ²² ʔim²² poi²² tant⁵⁵ dɪn⁵⁵
În nung é kê sang importânt dîn.
 3.PL two CLS all same important also
 ‘The two of them are all similarly important.’
 (PC0009-CLIN19)

(39) ta²² ge²² la ŋ³⁵ ʔu²² tsa²² ʔja⁵⁵ ʔu²² ja³⁵ tswe²²
Tage láng u tsa-iâ u yá tsue
 all person PF know have very many

vo²² ka²² bju²² la²² iis⁵⁵
vocabularies.
 vocabulary.PL

‘All people have known... have many vocabularies.’
 (PC0070-CLIN18)

(40) tsi²² kwa³⁵ ʔin²² das²² tui⁵¹ na²² ʔu²² hwat⁵⁵ tʰaŋ²²
tsikuá industry na uhuatthang
 some industry that can

pa ŋ³³ tsan²² ʔin⁵⁵ e³³ faj²² nan²² ʃal²² nids⁵⁵
pangtsan îñ =ē financial need-s
 help 3.PL GEN financial need-PL

‘some lines of work that can help their financial situation’
 (DE-001)

(41) ʔu²² we²² sjen³³ si³³ koŋ⁵⁵ ʔiŋ²² bun²² ʔwe⁵¹
Uwe siēnsī kōng Ingbun-uè
 some teacher speak English
 ‘Some teachers speak English.’
 (CLIN-19-132:30698)

(42) pe²² ro²² k^ha⁵⁵ ho⁵⁵ si³³ tam²² po^ʔ22 la³⁵ naŋ²² ʔwe⁵¹
Pero khâ hô sī tampoh Lánnang-uè.
 but CMPV good COP little Lánnang-uè
 ‘But a little Lánnang-uè is better.’
 (CLIN-18-68:10647)

(43) ja³⁵ tsjo³⁵ laŋ³⁵
Yá tsió láng.
 very few people
 ‘Very few people.’
 (CLIN-19-134:31400)

- (44) bo²² la²²ŋe²² k^hi⁵⁵ bi²²'non²²do⁵¹ lo²² e⁵⁵
Bo lánɡ=e kĥi Binondò lo éh.
 NEG person will go Binondo PFV PRT
 ‘No one / no person will go to Binondo.’
 (PC0005-CLIN18)
- (45) muj³⁵ tsi²² ge²² so³⁵tsaj³³ bo²² saŋ³⁵
Muí tsi ge sótsai bo sánɡ.
 each one CLS place not same
 ‘Each/every one of the places is not the same.’
 (CLIN-19-68:27453)

3.4.4 Pluralization

There are four markers in Lánnang-uè that are used to pluralize nouns, which are unspecified for number in Lánnang-uè by default (e.g., Hokkien-derived *kau* means ‘dog’ but could also mean ‘dogs in general’):

1. plural numeral
2. plural determiner
3. plural -s/-es suffix
4. plural particle²² *mgà*

The first marker, the plural numeral (in a plural numeral + classifier + N construction), is used to pluralize nouns regardless of their origin. Unlike other markers, it is also used in contexts that require a speaker to be specific about the noun quantity, such as answers to questions about quantity (e.g., Q: *Kuí e kau à?* ‘How many dog(s)?’ A: *Sā e kau.* ‘Three dogs.’). It is used on its own (46 to 48) or with other markers (49) except the marker *mgà* /ma⁵¹ŋa51/ (e.g., **sa e mga lánɡ* ‘three people’). There are, to my knowledge, no other distributional patterns or constraints. Some examples are as follows:

²² I use the term ‘particle’ to refer to any invariable linguistic item with grammatical function that does not readily fit into a standard classification of parts of speech (e.g., adverb, noun) (Crystal 2008:352; Busmann 2006:867). I consider it as a subset of ‘words’ and not a distinct “level of grammatical units between the clitic and the word levels” (Zwicky 1985:290). Particles/words, unlike clitics, are not phonologically dependent on a host word.

- (46) nəŋ³³ tsja²⁵⁵ kaw⁵⁵
nūng tsiâh kau
 two CLS dog
 ‘two dogs’
 (FRST-19-52:9266)
- (47) gwa⁵⁵ ʔu³³ sa³³ e³³ 'ja²²ja³⁵
Guâ ū sã e yayá.
 1.SG have three CLS helper
 ‘I have three domestic helpers.’
 (elicitation, PC0068)
- (48) gwa⁵⁵ ʔu³³ sa³³ tju³³ sjaw²² k^haj⁵⁵
Guâ ū sã tiū siaukhaî
 1.SG have three CLS calligraphy-paper
 ‘I have three pages of calligraphy paper.’
 (elicitation, PC0068)
- (49) dan⁵⁵ ʔu²² nəŋ³³ e³³ 'tɿms⁵⁵ ko⁵⁵
Dân u nūng ē term -s kô.
 1.PL.INC have two CLS term-PL PRT
 ‘We have two terms.’
 (PC0007-CLIN18)

The second marker is the plural determiner (in a plural determiner + N construction). This marker is used when the speaker, in addition to wanting to pluralize the noun, also wants to specify the accompanying noun and restrict its reference without providing information about the exact quantity. The noun can be derived from any language, as shown below:

Plural demonstrative + Hokkien-derived noun

- (50) tswaj³⁵ gi²²na⁵⁵
tsuai ginâ
 DEM.PL kid
 ‘these kids’
 (CLIN-18-4:1210)

Plural quantifier + Hokkien-derived noun

- (51) ʔu²²we²² sjen³³si³³ koŋ⁵⁵ ʔiŋ²²bun²²ʔwe⁵¹
Uwe siēnsī kông Ingbun-uè
 some teacher speak English
 ‘Some teachers speak English.’
 (CLIN-19-132:30698)

Plural demonstrative + English-derived noun

- (52) tswaj³⁵ ho²²mo²²'sek²²ʃwal⁵¹ lan⁵⁵ bo³³ beʔ⁵⁵
Tsuái *homosexual*, *lân* *bō* *béh.*
 DEM.PL homosexual 1.PL.INC NEG want
 'These homosexuals, we don't want [them].'
 (CLIN-18-4:1440)

Plural definite article + Tagalog-derived noun; plural demonstrative + Hokkien-derived noun

- (53) hwaj³⁵ pu²²'lis⁵⁵ kaj²²hu²²liʔ⁵⁵ hwaj³⁵ lan³⁵
Huai *pulís* *kay-hulí* **huái** *láng.*
 ART.PL police CAUS-capture DEM.PL person
 'The police captured those people.'
 (elicitation, PC0068)

The second marker is used with the three other markers, including the third marker – exemplified below.

- (54) k^ha⁵⁵nan⁵⁵ ts^hju³³ lan⁵⁵ k^ha⁵⁵ ʔu²² le⁵⁵ tʃjaw⁵⁵
Khânân *tshiu lân* *khâ* *u* *lê* *chiaû*
 as like 1.PL.INC CMPV PF PROG follow
 hwaj²² 'iuls⁵⁵
huai *rûle-s.*
 DEM.PL rule-PL

'It is like we are better at following the rules.'
 (PC0009-FRST19)

The third marker is the English-derived plural suffix (e.g., *-s*, *-es*), which is placed after the noun. The suffix *-s* [s] is attached after the noun except when the word ends with /s/, /z/, /tʃ/, /ʃ/, or /dʒ/, in which case *-es* [ɛs] is attached. It is used when the speaker only wants to pluralize the noun and nothing else. The suffix only attaches to English-origin nouns, but not mandatorily. It is generally used with other markers (55) but can occasionally be used on its own (56). No other usage patterns or constraints exist, as far as my investigation has uncovered.

- (55) hja³⁵ ʔu³³ ma²²ŋa²² 'bis⁵⁵
Hiá *ū* ***mga*** *beê-s.*
 there have PL bee-PL
 'There are bees there.'
 (PC0071-FRST18)

(56) tsju³³ si³³ 'pows²²tɿs⁵⁵ ʔeŋ²² la³⁵naŋ²²ʔwe⁵¹ ʔa⁵⁵ni³³ k^hwan⁵⁵
Tsiū *sī* *poster-s* *ieng* *Lánnang-uè* *ânī* *khuân.*
at.once COP poster-PL use Lánnang-uè like.that kind
‘So, it is posters that use Lánnang-uè and the like.’
(PC0092-CLIN19)

(57) 'tʃ^hɿtʃes bwe²² sjoŋ³³sin⁵¹ la¹¹
Church-êś *bo=e* *siōngsìn* *la*
church-PL NEG=POS believe PRT
‘Churches won’t believe.’
(elicitation, PC0068)

The fourth marker is the Tagalog-sourced plural particle *mgà* /ma⁵¹ŋa⁵¹/ (i.e., *mgà* + N construction). Like the first two markers, it is also used to pluralize words regardless of the words’ origin: *mgà* is placed before nouns derived from Hokkien (59 and 60), Tagalog (61), English (63), and Mandarin (64). Unlike the plural numerals and plural determiners but similar to the -s marker, the *mgà* marker is used by speakers who want to pluralize a noun and nothing else. It is most often used when the noun is not already pluralized using other strategies. It sometimes co-occurs with plural determiners (59) and the -s suffix (58), but never with the plural numeral marker. To my knowledge, there are no other usage conditions or constraints for this marker.

(59) 'pe²²ro²² hi⁵⁵ ge³⁵ bo²² səŋ⁵⁵ hwaj³⁵
Pero *hī* *gé* *bo* *sūng* *huái*
but that CLS NEG count ART.PL

ma²²ŋa²² sin³³kjaw³⁵ tse²²tsun⁵¹ dik³³ laj⁵¹ la¹¹
mgà *sīnkiaú* *tsetsùn* *dik* *lài* *la.*
PL new.immigrants now enter DIR PRT
‘But that doesn’t count the new immigrants who have entered.’
(PC0005-CLIN18.eaf)

(60) 'pe²²ro²² ma²²ŋa²² laŋ³⁵ o³⁵ ma²²ŋa²² laŋ³⁵
Pero *mgà* *láng* *o,* *mgà* *láng*
but PL person PRT PL person

ʔu²² kjo⁵⁵ gwa⁵⁵ a⁵⁵ni³³
u *kiô* *guà* *ânī.*
PF call 1.SG like.that

‘But people, people have called me that.’
(PC0095-CLIN19.eaf)

- (61) tsi⁵⁵ ge²² 'grup⁵⁵ laŋ³⁵ tu²²si²² koŋ³⁵ tʃ^haj²²'nis⁵⁵ ma²²ŋa²²
Tsí ge group, láng tusi kóng Chinése mga
 DEM CLS group person then say Chinese PL
 tʃ^haj²²'nis⁵⁵ ma²²ŋa²² ʔin²²'tʃik⁵⁵
Chinése mga Intsik.
 Chinese PL Chinese
 'So this group, people say 'Chinese' (to refer to) the Chinese people, the Intsik
 [derogatory term to refer to Chinese] people'
 (PC0006-CLIN18.eaf)

- (62) hi⁵⁵ge²² ma²²ŋa²² 'tʃaj²²nɪs²² fi²²li²²'pi²²no⁵¹
Hige mga Chinese Filipino
 the PL Chinese Filipino
 'The Chinese Filipinos'
 (PC0071-CLIN18)

- (63) di⁵⁵ k^hwa⁵¹ ma²²ŋa²² 'fa²²mi²²li⁵¹ kha⁵⁵ tswe⁵¹ si²² koŋ³⁵
Dî khuà mga family khâ tsuè si kóng
 2.SG look PL family CMPV many COP speak
 hok²²kjen²²ʔwe⁵¹ pa²² na²²man⁵⁵
Hokkiênuè pa namân.
 Hokkien PRT PRT

'Look, more families speak Hokkien (than any other language), I think.'
 (PC0019-CLIN18.eaf)

- (64) di⁵⁵ kaj²² t^hwe^{ʔ22} hwaj³⁵ ma²²ŋa²²
Dî kay- thueh huai mga
 2.SG CAUS-take DEM.PL PL
 sjaw²²k^haj⁵⁵ ŋa⁵⁵
siaukhai ngá.
 calligraphy.paper PRT

'Take those calligraphy papers.'
 (elicitation, PC0068)

The maximum number of plural markers that can co-occur in a single noun phrase is three. For example, the construction below has the plural determiner (i.e., plural definite article), the plural particle *mga*, and the plural suffix *-s*. The simultaneous use of three markers is less common than the use of one or two markers.

- (65) 'da²²pat⁵⁵ ʔin⁵⁵ tjoʔ²² ʔob²²'sɿv⁵⁵ hwaj³⁵ ma²²ŋa²² 'ɿuls⁵⁵
Dapât *în* *tioh* *obsêrve* *huái* *mga* *rule-s.*
 NEC 3.PL should observe ART.PL PL rule-PL
 'It is imperative that they should observe the rules.'
 (CLIN-19-118:20620)

3.4.5 Numerals

Cardinal numbers in Lánnang-uè can be formed using the following: the number zero, the basic numbers one to nine, and place markers (e.g., tens, thousands). The English-derived *zerò* /zi⁵¹row⁵¹/ is the default form for the number zero, while the Hokkien-sourced *khòng* /k^hoŋ⁵¹/ 'space' is used only when speakers are reciting a series of numbers (e.g., *di khòng khòng* '2-0-0'). The basic numbers of Lánnang-uè, sourced from Hokkien, are given in (66).

- (66)
- | | | | |
|---|----------------|-----------------------|-----------------------|
| 1 | <i>î/ťsít</i> | /ʔit ⁵⁵ / | /ťsit ³⁵ / |
| 2 | <i>dì/nūng</i> | /dì ⁵¹ / | /nŋ ³³ / |
| 3 | <i>sā</i> | /sa ³³ / | |
| 4 | <i>sì</i> | /sì ⁵¹ / | |
| 5 | <i>gō</i> | /go ³³ / | |
| 6 | <i>lák</i> | /lak ³⁵ / | |
| 7 | <i>tshît</i> | /ťjit ⁵⁵ / | |
| 8 | <i>puéh</i> | /pweʔ ⁵⁵ / | |
| 9 | <i>kaû</i> | /kaw ⁵⁵ / | |

The numbers 1 and 2 behave differently from the other basic numbers: they have two forms. By default, speakers use Hokkien-derived *î* and *dì*. However, when speakers count with these numbers²³ (67) or use these before measure words (i.e., classifiers, all place markers) (68), they use Hokkien-derived *ťsít* and *nūng*, respectively. The only exception is when the measure word is the place marker *ťsap* 'tens' /ťsap³⁵/, in which case *dì* is used instead of *nūng* (i.e., *di ťsap* 'twenty'). *ťsít* is shortened to *ťsi* when used before measure words (e.g., *ťsi pāh* 'one hundred', *ťsi pūn* 'one book-like object').

- (67)
- | | | | | | | |
|--|------------------|------------------|------------------|------------------|-------------------|--------------------|
| ťsit ³⁵ | nŋ ³³ | sa ³³ | sì ⁵¹ | go ³³ | lak ³⁵ | ťjit ⁵⁵ |
| ťsít, | nūng, | <i>sā,</i> | <i>sì,</i> | <i>gō,</i> | <i>lák,</i> | <i>tshît...</i> |
| one | two | three | four | five | six | seven |
| 'one, two, three, four, five, six, seven...' | | | | | | |
| (elicitation, PC0068) | | | | | | |

²³ I mean just the numbers 1 and 2, not any number involving 1 and 2 (e.g., 21, 52, 101).

(68)	dan ⁵⁵	ʔu ²²	nŋ ²²	e ³³	'tɿms ⁵⁵	ko ⁵⁵
	<i>Dán</i>	<i>u</i>	<i>nūng</i>	<i>ē</i>	<i>têrm-s</i>	<i>kô.</i>
	1.PL.INC	have	two	CLS	term-PL	PRT
	'We have two terms.'					
	(PC0007-CLIN18)					

Lánnang-uè numbers greater than nine use place markers, given in (69). Place markers beyond the thousands are derived from English (e.g., *milliôn* 'million') (70). All other place markers are sourced from Hokkien.

(69)	tens	<i>tsáp</i>	/tsap ³⁵ /
	hundreds	<i>pâh</i>	/pa ⁵⁵ /
	thousands	<i>tshiēng</i>	/ts ^h jeŋ ³³ /
	millions	<i>milliôn</i>	/'mil ²² jon ⁵⁵ /
	billions	<i>billiôn</i>	/'bil ²² jon ⁵⁵ /

(70)	so ²²	dí ⁵⁵	koŋ ⁵⁵	tsi ²²	'mil ²² jon ⁵⁵	si ³³	bo ³³
	<i>So</i>	<i>dí</i>	<i>kông</i>	<i>tsi</i>	<i>milliôn</i>	<i>sī</i>	<i>bō</i>
	so	2.SG	say	one	million	COP	not

ʔa ²² kju ²² .iej ⁵⁵ te ⁵¹	ba ³⁵
<i>accurâte=ē</i>	<i>bá?</i>
accurate=MOD	Q

'So, are you saying the 1 million is not accurate?'
(PC0068)

Numbers that involve place markers in Lánnang-uè must be preceded by a basic number that specifies the quantity of the place, except when the number is between 10 and 19 (e.g., *tsáp* '10', *tsap dì* '12', other examples in 71). For example, the number 100, *tsi pâh*, has place-quantifier *tsít* 'one' before the place marker *pâh* 'hundreds'. The number 22, *di tsap dì*, has place-quantifier *dì* 'two' before the place marker *tsáp* 'tens'. Basic numbers can follow place markers for some numbers (e.g., *tsap sī* '14' in 72).

(71)	10	tsap ³⁵ <i>tsáp</i> tens				
	11	tsap ²² it ⁵⁵ <i>tsap</i> <i>ít</i> tens one				
	13	tsap ³³ sa ³³ <i>tsāp</i> <i>sā</i> tens three				
	19	tsap ²² kaw ⁵⁵ <i>tsap</i> <i>kaú</i> tens nine				
	20	dí ²² tsap ³⁵ <i>dī</i> <i>tsáp</i> two tens				
	35	sa ³³ tsab ³³ go ³³ <i>sā</i> <i>tsāp</i> <i>gō</i> three tens five				
(72)	si ²² gu ²² io ⁵¹ <i>Sigurò</i> maybe	ma ²² ŋa ²² <i>mga</i> APPROX	tsap ²² <i>tsap</i> ten	si ⁵⁵ <i>sī</i> four	e ³⁵ <i>é</i> CLS	
	‘Maybe around fourteen of them.’ (CFH-002)					

The number 100 is formed by combining the quantifier *tsít* and the place marker *pâh* ‘hundreds’. Other combinations can be formed for higher numbers by using higher-order place markers (73).

(73)	100	tsi ²² <i>tsi</i> one	paʔ ⁵⁵ <i>pâh</i> hundreds
	200	nəŋ ²² <i>nung</i> two	paʔ ⁵⁵ <i>pâh</i> hundreds
	400	si ⁵⁵ <i>sî</i> four	paʔ ⁵⁵ <i>pâh</i> hundreds
	3000	sa ³³ <i>sā</i> three	tʃ ^h jeŋ ³³ <i>tshiēng</i> thousands
	5000000	go ³³ <i>gō</i> five	'mi ^l 22jon ⁵⁵ <i>milliòn</i> millions

If the number involves multiple place markers (e.g., hundreds, tens), the numerical string associated with the higher-order place (e.g., hundreds) is placed before the lower-order place (tens), as in (74).

(74)	401	si ⁵⁵ <i>sî</i> four	paʔ⁵⁵ <i>pâh</i> hundreds	ʔit ⁵⁵ <i>î</i> one
	625	lak ²² <i>lak</i> six	paʔ⁵⁵ <i>pâh</i> hundreds	di ²² tsab²² go ³³ <i>di tsap gō</i> two tens five
	1022	tsi ³³ <i>tsi</i> one	ts^hjeŋ³³ <i>tshiēng</i> thousands	di ²² tsab²² di ⁵¹ <i>di tsap dî</i> two tens two
	4256	si ⁵⁵ <i>sî</i> four	tʃ^hjeŋ³³ <i>tshiēng</i> thousands	nɿ ³³ paʔ⁵⁵ go ²² tsab²² lak ³⁵ <i>nung pâh go tsap lāk</i> two hundreds five tens six

Numbers that involve the date and time or decimals originate from English. For instance, ‘2020’ is not expressed in Lánngang-uè as Hokkien-sourced *nūng tshiēng di tsáp* ‘2020’ or *dī khòng dī khòng* ‘two zero two zero’, but as English-sourced *twenty twenty* ‘year 2020’

[ˈtʷen²²ti²²ˈtʷen²²ti⁵¹]. *Nine-thirty* ‘9:30’ [ˈnajn²²ˈtɿ²²ti⁵¹] is used instead of *kaú tiam sa tsap hūn* ‘nine hour thirty minutes’. Example (75) shows a Lánnang-uè speaker using an English-origin decimal number. English-sourced *poînt* ‘point’ is used for the decimal point.

- (75) hi⁵⁵ge³³ˈwan²²ˈpojnt²²ˈfajv²²ˈmɪl²²jon²² ˈʔeθ²²nɪk²² tʃ^haj²²ˈnis⁵⁵
Hîgē one point five million ethnic Chinêse?
 ART one point five million ethnic Chinese
 ‘(Do you mean) the 1.5 million ethnic Chinese?’
 (PC0005-CLIN18)

Ordinal numbers are formed by placing Hokkien-sourced *tē* /te³³/ before the cardinal number. To form ‘fifteenth’, for example, a speaker uses *tē* ‘ORD’ before *tsāp gō* ‘fifteen’, as in *tē tsāp gō* ‘fifteenth’. Another example is shown in (76).

- (76) te³³ go³³ dit³⁵ to⁵¹ laj¹¹
Tē gō dít tò -lai.
 ORD five day return -DIR
 ‘They returned on the fifth day.’
 (CFH-001)

The Tagalog-sourced *mga*²⁴ /ma⁵¹ˈŋa⁵¹/ ‘around’ is used when approximating a number, as in (77) and (78).

- (77) ma²²ˈŋa²² tsab²²di²² tiam²² pwa⁵¹
Mga tsapdi tiam puà
 APPROX twelve time half
 ‘Around twelve thirty’
 (DE-001)

- (78) tuj⁵⁵ si²²gu²²ˈio⁵¹ ma²²ˈŋa²² lak²²tsap²² he⁵¹ ʔi²²sjoŋ⁵¹
Tuí sigurò mga laktsap hè isìòng.
 toward maybe APPROX sixty year above
 ‘Maybe towards those around sixty years old and above.’
 (PC0001-CLIN18)

²⁴ ‘Mga’ is used in the orthography instead of ‘manga’ in accordance with the preferences of my native speaker informants.

3.4.6 Interrogatives

The five interrogatives in Lánnang-uè that are associated with the noun phrase are *shangá* /ʃa²²ŋa³⁵/ ‘who’, *shammîh* /ʃa²²mɪʔ³⁵/ ‘what’, *kuí* /kuj⁵⁵/ ‘how many’, *tó* /to³⁵/ ‘which’, and *dua* /dwa³³/ ‘how much’. *Kuí* and *tó* can only be used before a classifier (e.g., *Kuí nià?* ‘How many (clothes)?’). *Dua* can only be used before adjectives (e.g., *Dua suí?* ‘How beautiful?’) and is used in interrogatives that involve degree (e.g., degree of beauty).

There is variation in the pronunciation of the ‘who’ and ‘what’ interrogatives. Most speakers of Lánnang-uè consistently use *shangá* ‘who’ [ʃa³³ ŋa³⁵] and *shammîh* ‘what’ [ʃa³³ mɪʔ⁵⁵], while the rest consistently use *siangá* [sja³³ ŋa³⁵] and *siammîh* [sja³³ mɪʔ⁵⁵]. I have not systematically investigated whether these pronunciation differences are associated with particular social groups. Examples are given in (79) through (83). More examples can be found in Section 3.6, where *wh*-questions will be discussed.

- (79) kap⁵⁵ ʃa²²ŋa³⁵ bo²² saŋ³⁵
Kâp **shangá** *bo* *sáng?*
 with who not same
 ‘Who are they not identical with?’
 (PC0072-CLIN18)

- (80) k^hwā⁵⁵ gwa⁵⁵ k^hwā⁵⁵ tjoʔ²² ʃa³³ mɪʔ⁵⁵ o⁵¹
Khuânn *guâ* *khuânn* -*tioh* **shammîh** ò?
 see 1.SG see -PFV what PRT
 ‘(Does it) depend on what I have seen?’
 (PC0052-FRST19)

- (81) di⁵¹ ʔu³³ kuj³⁵ e²² ba⁵¹
Dî *ū* **kuí** *e* *bà?*
 2.SG have how.many CLS PRT
 ‘How many do you actually have?’
 (PC0068, elicitation)

- (82) di⁵⁵ sju³³ dwa²² tswe⁵¹ a¹¹
Dî *siū* **dua** *tsuè* *a?*
 2.SG receive how many PRT
 ‘How much did you receive?’
 (CFH-002)

- (83) di⁵⁵ p^haŋ⁵⁵si⁵⁵ to³⁵ tsi²² ge³⁵ a⁵¹
Dì phâhsî tó tsi gé à?
 2.SG kill which one CLS PRT
 ‘Which one did you kill?’
 (PC0068, elicitation)

3.4.7 *Adjectivals*²⁵

3.4.7.1 Modification

There are two ways to modify a noun using adjectivals: pre- and post-nominal modification. The variety has four adjectival modification strategies:

1. Adj + =*ē*²⁶ + N
2. Adj + N
3. Adj + *nà* + N
4. N + *nà* + Adj

The default pre-modification strategy is the Adj + Hokkien-derived modifier clitic²⁷ =*ē* /ŋe^{X1}/ + N construction. This strategy is generally used with adjectives.

²⁵ The term ‘adjective’ refers to a word whose most characteristic role is as the modifier of a noun phrase. The term ‘adjectival’, on the other hand, is a general term that comprises adjectives and constituents that have the role of adjectives (Matthews 2007:47).

²⁶ The modifier element *ē* is analyzed as a clitic because it has relatively few combinatory restrictions (Zwicky and Pullum 1983). For example, in addition to attaching to adjectives, *ē* can also attach to full sentences (e.g., [*Jasôn alís*]=*ē sîkân* ‘time when Jason left/was leaving or literally Jason-leaving time’), verb phrases (e.g., [*Hîge [lêh shutsuí]*]=*ē pagông* ‘the turtle that is currently swimming/literally, the currently-swimming turtle’.)

²⁷ I use the term ‘clitic’ to refer to grammatical units that are “not straightforwardly either an affix or a word on its own” (Matthews 2007:163). I characterize a linguistic element as a clitic if it exhibits a low degree of selection with respect to the host (i.e., it has relatively low combinatory constraints) (Zwicky and Pullum 1983:503; Mintz and Byrd 2010). This behavior is different from that of affixes, which have a high degree of combinatory restrictions (e.g., *taga-* can only attach to noun words, not noun phrases or verb words) (Mintz and Byrd 2010).

(84) di⁵⁵ tuj⁵⁵ mikst⁵⁵ e¹¹ la³⁵ naŋ²² ʔue⁵¹ ʔu²²
Dî tuí mixêd =e Lánnang-uè u
 2.SG to mixed =MOD Lánnang-uè have

ʃa²² mi^ʔ55 kam²² kak⁵⁵ a⁵¹
siammîh kamkak a?
 what feel PRT

‘What do you feel about mixed Lánnang-uè?’
 (PC0002-CLIN18)

(85) ʃem²² pɿe⁵¹ di⁵⁵ na³³ si³³ ‘ha²² lo²² ‘ha²² lo^ʔ55 e¹¹ laŋ²² gwedʒ⁵⁵
Syemprè dî nā sī halohalô =e languagê...
 sure 2.SG if COP hybrid =MOD language

‘Of course, if we are talking about a hybrid language...’
 (PC0002-CLIN18)

(86) ja³⁵ twa¹¹ e¹¹ ‘di²² fɿ²² .ɿens⁵⁵
Yá tua =e differênce?
 very big =MOD difference

‘Very big difference?’
 (PC0071-CLIN18)

Occasionally, some speakers use the default strategy with non-adjective adjectivals.

(87) ku²² kɿaw³⁵ k^ha⁵⁵ tsa⁵⁵ le⁵⁵ tswe⁵⁵ səŋ²² di⁵⁵ e¹¹ si²² tsun⁵¹
kukiaú khâ tsâ lê tsuê sungdî =e sitsùn
 old.immigrants CMPV early PROG make business=MOD time

‘a time when old immigrants were making businesses earlier’
 (CLIN-19-119:21178)

The second pre-modification strategy is the Adj + N strategy. This strategy is only used with English-derived adjectives, adjectival demonyms (i.e., nouns used to denote the natives or inhabitants of a particular country, state, city, etc.) derived from English and Hokkien and adjectives that indicate a location (e.g., *Irish* ‘Irish’).

(88) ʔin⁵⁵ ʔi³³kjeŋ³³ ʔa²²'si²²mi²²let⁵⁵ nan⁵⁵e³³ 'kul³³ tʃuɾ⁵¹
În ikiēng assimilate nân=ē culturè
 3.PL already assimilate 1.PL.INC=GEN culture PRT

lo²² ti³³ hi⁵⁵ge²² fi²²li²² 'pi²²no²²'tʃaj²²nis²² 'kul²²tʃuɾ⁵¹
lo tī hīge Filipino-Chinese culturè.
 PFV PREP ART Filipino.Chinese culture
 'They already assimilated to our culture, to the Filipino-Chinese culture.'
 (PC0002-CLIN18)

(89) lan⁵⁵ maŋ³³ k^hwa⁵⁵k^hin³³ hwi³³di³³pin³³ laŋ³⁵ diŋ⁵⁵ la⁵¹
Lân māng khuâkhîn Huīdīpīn lāng dīn là.
 1.PL.INC NEG look.down Philippines person also PRT
 'Let's not look down on the Filipinos.'
 (PC0019-CLIN18)

The least commonly and most inconsistently used pre-nominal modification strategy is the Adj + Tagalog-derived *na* /na⁵¹/ morpheme + N strategy. Here, the adjectival must be a simple adjective or adjectival demonym as well. But unlike the second strategy, the adjectivals are not restricted to those derived from English.

(90) hi⁵⁵ge²² 'blak⁵⁵ na²² 'sel²²fown⁵¹ ti³³ to²²loʔ³⁵ a⁵⁵
Hīge blāk na cellphòne tī tolòh á?
 ART black MOD cellphone PREP where PRT
 'Where is the black mobile phone?'
 (elicitation, PC0068)

(91) pa²²ra²²ʔu²²hwa⁵⁵na ŋ²² ju²²'najt⁵⁵ hwaj³⁵ fi²²li²² 'pi²²nos⁵⁵ kap⁵⁵
Para uhuatnang unite huái Filipinôs and
 so can unite ART Filipinos and
 hwaj³⁵ 'pjuɾ⁵¹ na²² 'tʃaj²² nis⁵⁵ ta²²la²² ga⁵¹
huái pùre na Chinese talaga.
 ART pure MOD Chinese really

'So that we can really unite the Filipinos and the pure Chinese.'
 (PC0094-CLIN19)

(92) bas²²tos⁵⁵ na²² laŋ³⁵
bastôs na lāng
 rude MOD person
 'rude person'
 (elicitation, PC0068)

Simultaneously using the first and third pre-nominal modification strategies to modify a noun with an adjective is ungrammatical (e.g., **blâck-e na bâg* ‘black bag’).

Post-nominal modification is another way to modify a noun in the variety. It is used to modify nouns using a clause or a verb phrase. The only post-nominal strategy is the N + Tagalog-derived *nà* /na⁵¹/ morpheme + Adj construction.

Nà has multiple functions. It is very occasionally used as a modifier particle²² after the adjectival in pre-modification (see earlier description). In the context of post-modification, its main function is that of a relativizer (see Section 3.9). It is placed before a clause or a verb phrase.

- (93) gwa⁵⁵ kaj²²tan⁵⁵ hi⁵⁵ge²² 'plant⁵⁵na³³ jo²²be⁵¹ ho²² gwa⁵¹
Guâ kay-tân hîge plânt nā shobè ho guâ.
 1.SG CAUS-throw ART plant REL little.sister give me
 ‘I caused the plant that my little sister gave me to be thrown.’
 (elicitation, PC0068)

- (94) gwa⁵⁵ ʔu²² 'fiends⁵⁵ na²² ʔu²² le⁵⁵ ʃu³³tsuj⁵⁵
Guâ u friênds na u lê shutsuî
 1.SG have friends MOD PF PROG swim
 ‘I have swimmer friends/friends who swim habitually.’
 (elicitation, PC0068)

- (95) hi⁵⁵ge²² laŋ³⁵ na²² be²² pa³⁵la³⁵ ʔi⁵⁵e³³ ts^hja³³
Hîge lāng na be pālā î=ē tshia
 ART person REL NEG pay 3.SG=GEN car
 ti³³ti³³ ko⁵¹ i¹¹ ba³⁵
tīī kò i bá?
 ITER guard 3.SG Q

‘Is the person that hasn’t paid for their car continuing to guard them?’
 (PROT-16-NA:38436)

Its secondary function – one that is occasionally used, and only by some speakers – is that of a post-modification particle. If this strategy (i.e., N + *nà* + Adj) is used, the adjectival must be a simple adjective or an adjectival demonym.

(96) gwa⁵⁵ ʔu²² 'fiends⁵⁵ na²² la³⁵naŋ³⁵
Guâ u friênds na Lánnáng.
 1.SG have friends MOD Lannang
 'I have Lannang friends.'
 (PC0095-CLIN19)

(97) tsa⁵⁵ ʔin²² tsa⁵⁵ ja³⁵ tswe²² 'sti²².tjo²²tajps⁵⁵
Tsâ in tsâ yá tsue stereotype-s
 then 3.PL then very many stereotype-PL

 na²² ʔa²²ni²²t^he⁵⁵ din⁵⁵ la⁵¹
na anithê dîn là.
 MOD similar also PRT

'And then, they also have many similar stereotypes.'
 (PC0009-CLIN19)

(98) so²² di⁵⁵ kak²²tít⁵⁵ la³⁵naŋ²²ʔue⁵¹ na²² 'ʔo²².ial⁵¹ ʔo²²'kej⁵¹
So dí kaktít Lánnang-uè na oràl okày
 so 2.SG feel Lánnang-uè MOD oral okay

 laŋ⁵⁵ ba³⁵
láng bá?
 only Q

'So, do you feel that oral Lánnang-uè is only okay?'
 (PC0001-CLIN18)

To modify a noun with multiple adjectivals, the adjectival series must first be coordinated by placing the cumulative conjunction (e.g., *kâp* 'and', *kiaū* 'and/with') (see Section 3.9) in between the final adjectival and the penultimate one. Either the Hokkien-derived =*e* clitic or the Tagalog-derived *nà* morpheme must then be placed between the coordinated adjectival series and the noun to be modified. Examples involving both pre-nominal adjectival strategies can be found below:

(99) hi⁵⁵ge²² ja³⁵ ts^haw⁵¹ kap⁵⁵ ja²² ts^ha⁵⁵e¹¹ da²²'ga^ʔ⁵⁵
hîge yá tshau kâp ya tshâ =e dagâ
 ART very stinky and very noisy=MOD rat
 'the very stinky, very noisy rat'
 (elicitation, PC0068)

- (100) hi⁵⁵ge²² pu³⁵ ta²²'mad⁵⁵ kap⁵⁵ 'ɲawnd⁵⁵ na²² 'pan²²da⁵¹
hîge *puí,* *tamâd,* ***kâp*** *roûnd* *na* *pandà*
 ART fat lazy and round MOD panda
 'the fat, lazy, round panda'
 (elicitation, PC0068)

Not having a modifier clitic or particle in the (pre-modified) noun phrase is ungrammatical.

- (101)* hi⁵⁵ge²² 'flim²²si⁵¹ kap⁵⁵ kuj³⁵ ts^hju⁵¹
hîge *flimsy* ***kâp*** *kuí* *tshiù*
 ART flimsy and high tree
 'the flimsy and tall tree'
 (male speaker 25, judgment)

There is a post-nominal alternative, where the Tagalog-derived particle *nà* is placed before the adjectival series and after the noun, as in:

- (102) hi⁵⁵ge²² mjaw³³ na²² 'fɿ²²ɿi⁵¹ swe⁵¹ kap⁵⁵ tɲ³⁵
hîge *miaū* *na* *furry* *suè* ***kâp*** *túng*
 ART cat MOD furry small and long
 'the furry, small, long cat'
 (elicitation, PC0068)

3.4.7.2 Comparative constructions

Lánnang-uè has five basic comparative constructions. The first four are comparatives of superiority. The last one is a comparative of equality. The variety does not have comparatives of inferiority (e.g., *less*, *least* in English).

1. NP₁ + *pí* + NP₂ + *khâ* + Adj
2. NP₁ + *khâ* + Adj + *pí/kaysà* + NP₂
3. NP₁ + *khâ/mâs* + Adj
4. NP₁ + *mâs* + *khâ* + Adj
5. NP + *sáng* + Adj

The comparatives of superiority specify that an entity (NP₁) exceeds another entity (NP₂) with respect to a gradable property (Adj). The first two constructions are used when the speaker wants to be explicit about what the superior entity is being compared to. In the first, the Hokkien-derived comparative particle *pí* /pi³⁵/ is placed in between the superior entity NP₁ (e.g., *managèr*

‘manager’ in 103) and the inferior one (e.g., *supervisòr* ‘supervisor’ in 103). The string is then followed by the Hokkien-derived particle *khâ* /k^ha⁵⁵/ and a gradable adjective (e.g., *kui* ‘high’).

- (103) 'ma²²ne²²dʒ_ɿ⁵¹ pi³⁵ su²²p_ɿ²²'vaj²²so_ɿ⁵¹ k^ha⁵⁵ kuj³⁵ ba³⁵
Managèr ***pí*** *supervisòr* ***khâ*** *kui* *bá?*
 manager CMPV supervisor CMPV high Q
 ‘Is the manager higher in rank compared to the supervisor?’
 (CE-001)

In the second, the constituent containing the gradable property (i.e., *khâ* + Adj) is not placed after NP₁-CMPV-NP₂ construction but rather in between – specifically, after the superior entity NP₁ (e.g., *Hâmbún* ‘Chinese’).

- (104) ham⁵⁵bun³⁵ k^ha⁵⁵ ʔad²²vans⁵⁵ kay²²sa²² fi²²li²²'pi²²no⁵¹
Hâmbún ***khâ*** *advancê* ***kaysa*** *Filipinò.*
 Chinese CMPV advanced CMPV Filipino
 ‘The Chinese language is more advanced than Filipino.’
 (elicitation, PC0068)

- (105) ʔin⁵⁵ k^ha⁵⁵ ʔe²² t^hja³³ gwa³¹ pi²² di⁵⁵
În ***khâ*** *e* *thiã* *guà* ***pí*** *dí.*
 3.SG CMPV ABI listen 1.SG CMPV 2.SG
 ‘They (SG) naturally listen to me more compared to you.’
 (elicitation, PC0068)

In the first construction, only *pí* is used; in the second, either the *pí* comparative or the *kaysà* comparative particle derived from Tagalog is used. I have not found factors that condition the use of one particle over the other. The first construction is used by default; the second is used when the speaker wants to emphasize the gradable attribute over the entity of comparison.

The third and fourth comparatives of superiority are different from the first two in that they are used when the speaker finds no need to be explicit about the entity of comparison. The phrase containing the entity of comparison does not exist in either construction.

In the third construction, the superior entity NP₁ is directly followed by the Hokkien-derived *khâ* or the Tagalog-derived *mâs*²⁸ /mas⁵⁵/ particle, and then the adjective. The particle *khâ* is used more frequently than *mâs* in this construction. There are, to my knowledge, no factors that condition the use of one particle over the other.

²⁸ *Mâs* has Spanish origins but is regarded as a Tagalog word borrowed from Spanish (Schachter and Otones 1972). Here I simply regard it as a Tagalog-origin word.

- (106) ʔin⁵⁵ k^ha⁵⁵ .i²²'fajnd⁵⁵ ko⁵⁵
În khâ refinêd kô.
 3.SG CMPV refine PRT
 'They (SG) are more refined (than others).'
 (PC0002-CLIN18)
- (107) ʔu²² 'mas²² hok²²kjen⁵⁵ lo⁵¹
Uh, mas Hokkiên lò.
 Uh CMPV Hokkien PFV
 'Uhm, [they have] already [become] more Hokkien.'
 (PC0071)
- (108) 'pe²²ro²² ba²²'ka⁵¹ ʔu³³ 'mas²² gaw³⁵e³³
Pero baka ū mas gaw³⁵e³³
 B ut maybe have CMPV smart=MOD
pa i²²'ti²²si²²pants⁵⁵
participant-s...
 participant-PL
 'But maybe there are smarter participants...'
 (PC0169)

The fourth construction (i.e., NP₁ + *mâs* + *khâ* + Adj), unlike the third, involves both *mâs* and *khâ* particles, where *mâs* must come before *khâ*. The simultaneous occurrence of these two particles does not make the comparison more emphatic than if just one of them were present, as shown below:

- (109) 'mas²² k^ha⁵⁵ hon²²pjen⁵¹ ba³⁵ tuj⁵⁵ din⁵⁵ ʔa⁵⁵si³³
Mas khâ hongpiên bá tuī dîn âsī...?
 CMPV CMPV convenient Q PREP 2.PL or
 'Is it more convenient for you or...?'
 (CFH-002)

This construction is only occasionally used by some speakers. Based on my analyses, the third and fourth constructions are in free variation; they are used interchangeably.

To express a comparison of equality between two entities, the comparative construction NP + Hokkien-sourced word *sáng* /saŋ35/ 'same' + Adj is used, where NP must denote two entities. This noun phrase can be a coordinated pair of nouns joined by a conjunction (e.g., *î kap guâ* 'they and I') or a word/phrase that references two entities. In (110), for example, the two entities in the phrase *în nung-é* 'two of them' are equally important.

(110) ʔin⁵⁵ nəŋ²² e³⁵ ke⁵⁵ saŋ³³ ʔim²² poi²² tant⁵⁵ din⁵⁵
În nung é kê sãng important dîn.
 3.PL two CLS all same important also
 ‘The two of them are both similarly important.’
 (PC0009-CLIN19)

(111) dʒan⁵¹ kap⁵⁵ me²² ɿ⁵¹ saŋ³³ gaw³⁵
Jòhn kâp Mary sãng gáu.
 John and Mary same smart
 ‘John is as smart as Mary.’
 (elicitation, PC0068)

3.4.7.3 Superlative constructions

Lánnang-uè has two known superlative constructions. To indicate that an entity has qualities or attributes that set them above other entities, two constructions are used:

1. *té* + Adj + modifier + N
2. *pinakà-* + Adj + modifier + N

The first construction involves placing a Hokkien-derived particle *té* /te³⁵/ before the adjectival whereas the second involves attaching a Tagalog-derived prefix *pinakà-* /'pi²²na²²ka⁵¹/ to the adjectival. In both constructions, a modifier (i.e., Tagalog-derived *na* or Hokkien-derived =*ē*) is placed between the adjectival and the noun. In (112), for example, *té* is placed before the gradable adjective *kuí* ‘high’ to indicate that the *managèr* ‘manager’ (N) is above the rest in terms of rank (height).

(112) te³⁵ kuj³⁵ na²² 'ma²²ne²²dʒɿ⁵¹
té kuí na managèr
 SUP high MOD manager
 ‘the highest manager’
 (elicitation, PC0068)

(113) te³⁵ klow⁵⁵e¹¹ si³³ kjo⁵⁵ tswe⁵⁵ praj⁵⁵
Té closê=e sī kiô tsuê pride.
 SUP close=e COP call RSLT pride
 ‘The closest is referred to as pride.’
 (CLIN-19-130:29052)

- (114) te³⁵ ʔu³³ tsi³⁵e³³ hjo⁵⁵ŋe⁵⁵
té *ū* *tsí=ē* *hiôngê*
 SUP have money=MOD DEM
 ‘that [people] with the most money/ the richest [people]’
 (CLIN-19-130:28836)
- (115) hi⁵⁵ge²² 'pi²²na²²ka²² ts^ha⁵⁵e³³ gi²²na⁵⁵ si³³ ʔi⁵⁵
Hîge *pinaka-* *tshâ=ē* *ginâ* *sī* *î.*
 ART SUP noisy=MOD kid COP 3.SG
 ‘They (SG) are the loudest kid.’
 (elicitation, PC0068)

The two superlative constructions are both frequently used in the variety. To my knowledge, there are no distributional patterns.

3.4.7.4 Approximation

There are two strategies in Lánnang-uè to encode an approximative meaning (‘sort of X quality’): (1) the use of Tagalog-derived *medyò/mejò* /me⁵¹ dʒo⁵¹/ or Hokkien-derived *tampóh* /tam³³ poʔ³⁵/ before the adjectival and (2) reduplicating the adjectival.

The first strategy is the default. The approximation particle is placed before the adjective, as in the following:

- (116) ʔm³³ʔuj³³ hi⁵⁵ djap²² tsjoʔ²²t^haw³⁵ 'me²²dʒo²² ma²²la²²'ki⁵¹
īn-uī *hī* *diap* *tsiohthau* *medyo* *malakì.*
 because DEM CLS rock APPROX big
 ‘Because that rock is sort of big.’
 (FRST-18-1:505)
- (117) 'pa³³raŋ³³ 'me²²dʒo²² 'rej²²sist⁵⁵ ʔa⁵⁵ni³³
Pārāng *medyo* *racîst* *ânī.*
 PRT APPROX racist like.that
 ‘It appears that (they are) sort of racist like that.’
 (CLIN-19-95:14102)
- (118) 'pa³³raŋ³³ 'me²²dʒo²² ʔu³³ tsi³⁵
Pārāng *mejo* *ū* *tsí*
 PRT APPROX have money
 ‘It appears that (they are) sort of rich/have money.’
 (elicitation, PC0068)

- (119) 'so²² tam²²po²² k^ha⁵⁵laŋ⁵⁵ ʔan²²kom²² foɿ²²ta²²bol⁵¹
*So **tampoh**, khâlâng, uncomfortablè.*
 so APPROX like uncomfortable
 'So it was sort of, like, uncomfortable.'
 (CLIN-18-68:6932)

To my knowledge, there are no factors that condition the use of the more common variant *medyò/mejò* over *tampóh*. Both are used interchangeably.

The second strategy, reduplication, is used more restrictively. It is only used with adjectives that are monosyllabic and derived from Hokkien. An identical copy of the adjective is placed adjacent to the original adjective, as in:

- (120) pɛ^ʔ⁵⁵ k^hi⁵⁵ tsi²²ge²² twa²² twa²² tsjo^ʔ³⁵ tjeŋ³⁵k^ha³³
*pêh khî tsige **tua tua** tsióh tiéngkhā.*
 climb DIR ART big big rock top
 'climbed on top of a rock that is sort of big'
 (FRST-19-99:15158)

- (121) ʔom²² tsaj²² ʔja⁵⁵ goŋ³³ goŋ⁵¹ ho³⁵
*Um tsai-iá. **Gōng gòng** hó.*
 NEG know stupid stupid PRT
 'Don't know. Sort of stupid, right?'
 (CLIN-19-107:17211)

3.4.7.5 Negation and adjectives

The particle *bó* /bo³⁵/, derived from Hokkien, is used to express the opposite or negated version of the adjectival.

- (122) k^ha⁵⁵laŋ⁵⁵ bo²² suj⁵⁵ ko⁵⁵ ho³⁵
*Khâlâng **bo** suí kô hó.*
 like NEG pretty PRT PRT
 'It is like, not pretty (or ugly).'
 (CLIN-18-5:1837)

- (123) hi⁵⁵ge²² bo²² ʔan⁵⁵ tsja³⁵ t^hak²²ts^he^ʔ⁵⁵e³³ gi²²na⁵⁵
*hîge **bo** ân tsiá thaktshêh=ē ginâ*
 ART NEG PREP here study=MOD kid
 'the kid that did not study here/ the not-study-here kid'
 (elicitation, PC0068)

3.4.7.6 Adverbial modification

To modify an adjective or an adjectival demonym, the adverb is placed before the adjectival, as in:

- (124) lan⁵⁵ ʔi⁵⁵kjeŋ³³ ʔan⁵⁵ tsja³⁵ ja²² ku⁵⁵ lo²² e⁵⁵
Lân *ikiēng* *ân* *tsiá* *ya* *kû* *lo* *êh*.
 1.PL.INC already PREP DEM very long PFV PRT
 ‘We were here a very long time already.’
 (CLIN-19-94:13983)

The intensive adverbial *napakà* /'na⁵¹pa⁵¹ka⁵¹/ ‘completely’, derived from Tagalog, denotes a more forceful or stronger attribute relative to the root on which the intensive is built. It is used to modify adjectives, adjectival demonyms, adjectival verb phrases that are headed by ability modality markers, and adjective phrases headed by a negative particle. It is placed before adjectivals that are derived from Hokkien, Tagalog, and English.

- (125) 'na²²pa²²ka²² bas²²'tos⁵⁵e³³ sjen³³si³³
napaka *bastôs=ē* *siēnsī*
 completely rude=MOD teacher
 ‘thoroughly rude teacher’
 (elicitation, PC0068)

- (126) 'na²²pa²²ka²² ʔi²²dʒan⁵¹e³³ laŋ³⁵
napaka *Indiàn=ē* *láng*
 Completely Indian=MOD person
 ‘thoroughly Indian person’
 (elicitation, PC0068)

- (127) 'na²²pa²²ka²² ʔe²²hjaw³⁵ səŋ⁵⁵ 'nam²²bɿ⁵⁵se³³ gi³³na⁵⁵
napaka *ehiaú* *súng* *number-s=ē* *ginâ*
 completely ABI count number-PL=MOD kid
 ‘kid that is thoroughly able to count numbers’
 (elicitation, PC0068)

- (128) 'na²²pa²²ka²² bo²² gaw³⁵e³³ kaw⁵⁵
napaka *bo* *gáu=ē* *kaú*
 completely NEG smart=MOD dog
 ‘thoroughly stupid dog’
 (elicitation, PC0068)

3.4.8 The pronominal system

3.4.8.1 Personal pronouns

Lánng-uè has seven personal pronouns, all of which are derived from Hokkien (Table 7). The variety has a first person plural inclusive/exclusive distinction. The pronoun *gún* refers only to the speaker or writer and their associates but not to the addressee(s); *lân*, *dân*, or *nân* refers to the speaker and the addressee(s). The three first person inclusive plural pronouns appear to be phonetic or phonological variants of a single morpheme. I have yet to find conditioning factors for the use of one variant over the others. Based on my preliminary analyses, they appear to be in free variation.

Table 7. Personal pronouns

	Singular	Plural
first person	<i>guâ</i> /gwa ⁵⁵ /	<u>Exclusive</u> <i>gún</i> /gun ⁵⁵ / <u>Inclusive</u> <i>lân</i> /lan ⁵⁵ / <i>dân</i> /dan ⁵⁵ / <i>nân</i> /nan ⁵⁵ /
second person	<i>dî</i> /di ⁵⁵ /	<i>dîn</i> /din ⁵⁵ /
third person	<i>î</i> /ʔi ⁵⁵ /	<i>în</i> /ʔin ⁵⁵ /

I provide examples for each of the seven personal pronouns below.

- (129) gwa⁵⁵ u²² tsi³³ ge³³ 'mi²²dɛl²² 'nejm⁵¹
Guâ u tsī gē middle nàme.
 1.SG have one CLS middle name
 'I have one middle name.'
 (CLIN-18-19:5860)

- (130) ka²²si⁵⁵ gun⁵⁵ 'ta²²la²²ga⁵¹ pju⁵¹
Kasī **gún** talagà pùre.
 because 1.PL.EXC really pure
 'Because we (excluding you) are really pure.'
 (CLIN-18-19:5914)

- (131) tan³³si³³ dan⁵⁵ tsi²²tsun⁵¹ ʔom³³ si³³ ti³³ ʔaj²² di²²jal⁵¹
Tānsī dān tsitsùn ūm sī tī ideəl
 but 1.PL.INC now NEG COP PREP ideal
- e³³ ko⁵⁵
 =*ē* *kô*.
 MOD PRT

‘But I am telling you that we (including you) are now not in an ideal one.’
 (CLIN-19-50:8747)

- (132) di⁵⁵ ta²²la²²ga⁵¹ di⁵⁵ tjo²² k^hwa⁵⁵ di⁵⁵e³³ ʔo²²dʒens⁵⁵
Dī talagà ... dī tloh khuâ dī=ē audiēnce.
 2.SG really ... 2.SG should look 2.SG=GEN audience
 ‘You really... you should look at your audience.’
 (CLIN-19-50:8752)

- (133) ka⁵⁵na⁵⁵ din⁵⁵ bo²² le⁵⁵ ʔjen³³ la³⁵na²²ʔue⁵¹ a¹¹
Kānā dīn bo lê iēng Lánnang-uè a?
 why 2.PL NEG PROG use Lánnang-uè PRT
 ‘Why aren’t you using Lánnang-uè?’
 (CLIN-18-68:10445)

- (134) ʔi⁵⁵ le⁵⁵ taw⁵⁵ ts^he²² hi⁵⁵ge²² ʔiog⁵⁵
Ī lê taū tshē hīge frōg.
 3.SG PROG help find ART frog
 ‘They (singular) are helping find the frog.’
 (FRST-18-70:10504)

- (135) ta⁵⁵ne⁵¹ ʔin⁵⁵ k^hwa⁵¹ tsi²²ge²² twa³³ e³³ tsjo²²t^haw³⁵
Tānè ĩn khuà tsige tuā =ē tsiohthau.
 later 3.PL look ART big MOD rock
 ‘Later, they (plural) found a big rock.’
 (FRST-18-70:10515)

3.4.8.2 Reflexive pronouns

The two reflexive pronouns of Lánnang-uè are the Hokkien-sourced *kaīki* and Tagalog-origin *mismò*.²⁹ Both can be placed directly after a personal pronoun to form an emphatic reflexive (136 to 138). *Kaīki* is the general form and can also be used in non-emphatic contexts, after the verb (139 and 140); *mismò* can only be used in emphatic contexts (137 and 138).

²⁹ *Mismò* originates from Spanish *mismo* ‘same’. It was presumably borrowed into Tagalog and then incorporated into Lánnang-uè. Here, I treat *mismò* as a Tagalog word.

- (136) di⁵⁵ kaj³³ki³³ ts^hoŋ⁵⁵ko⁵⁵si⁵¹ ko⁵⁵
*Dî **kaĩkĩ** tshông kôsi kô.*
 2.SG self make story PRT
 ‘You yourself should make a story.’
 (PC0020-FRST18)
- (137) gwa⁵⁵ mis³³mo⁵¹ ja³⁵ kaj²²k^ho⁵⁵ ‘pe²²ne²²tuejt⁵⁵
*Guá **mĩsmò** yá kangkhô penetrâte*
 1.SG self very hard penetrate

 ‘in²²tu²² ‘dej²² ko²²‘mju²²ni²²‘ti⁵¹
into their community.
 into their community.

 ‘I myself find it hard to penetrate into their community.’
 (PC0071)
- (138) di⁵⁵ mis³³mo⁵¹ mas²² ‘stia³³gel⁵¹ na³³ ‘fo²² ‘ʔol²²dɿ⁵¹
*Dî **mĩsmò** mas strugglè nā for oldèr.*
 2.SG self CMPV struggle CND for older
 ‘You will find you yourself struggling more if [you are talking] to older [people].’
 (PC0095)
- (139) tjoŋ³³kok⁵⁵ laŋ³⁵ k^ha⁵⁵ ko⁵⁵ kaj²²ki⁵¹
*Tiōngkòk láng khâ ko **kaĩkĩ.***
 China people more care self
 ‘The Chinese care for themselves more.’
 (PC0004-CLIN18)
- (140) gwa⁵⁵ ʔm²² si²² se^ʔ⁵⁵ kaj³³ki⁵¹
*Guá um si sêh **kaĩkĩ.***
 1.SG NEG COP say self
 ‘I’m not talking about myself.’
 (PC0099-CLIN19)

3.4.8.3 Idiomatic constructions

The variety has idiomatic constructions that involve pronouns derived from Tagalog (i.e., *kò* ‘1.SG’, *natîn* ‘1.PL.INC’, *namîn* ‘1.PL.EXC’, *mò* ‘2.SG’, *nyò* ‘2.PL’, *niyà* ‘3.SG’, *nilà* ‘3.PL’). These constructions are *sabi* ‘say’ + ____ ‘__ said’ and *akalà* ‘presumption’ + ____ ‘____ presumed’ (141).

- (141) 'pe²²ro²² 'ka²²hit²²na²² ja³⁵ ku⁵⁵ ʔa²²'ka²²la²² mo⁵¹
Pero *kahitna* *yá* *kû* *akala* **mò**
 but even.if very long presumption 2.SG

ʔu²² tjam³⁵tsjeŋ³³

u *tiámtsieng*.

have time

‘But even if a long time has passed, you presume you have time.’

(PC0005-CLIN18)

3.4.9 Inflectional affixes and clitics

The only inflectional affixes³⁰ in the domain of the noun phrase are the pluralizing *-s* [s] and *-es* [ɛs] suffixes. They have already been described in Section 3.4.4.

The only two clitics in the same domains are genitive case marker =é /ʔe³⁵/ and modification marker =e /ʔe^{X1}/,³¹ both derived from Hokkien. The clitic /ʔe³⁵/ indicates that the phrase is a possessor (e.g., *guâ*=é [gwa⁵⁵e³⁵] ‘1.SG GEN’ or ‘mine’). In the examples below, =é (phonemically =é) is attached to the pronoun *guâ* ‘1.SG’, the pronoun *în* ‘3.PL’, and the noun *ginnâ* ‘kid’ in their respective sentences to make them possessors of the noun phrases that follow it (e.g., =é makes *guâ* the possessor of *pârt* in 142).

- (142) gwa⁵⁵e²² 'part⁵⁵ si²² 'mis²²mo⁵¹ hi⁵⁵ge²²'ʔin²²tɿ²²vju⁵¹ ko⁵⁵
Guâ=é *pârt* *si* *mismò* *hîge* *interview* *kô*.
 1.SG=GEN part COP itself DEM interview PRT
 ‘My part is the interview itself.’
 (PC0004-CLIN18)

- (143) di⁵⁵ pat⁵⁵ khî⁵⁵ ʔin⁵⁵e³³ mju²²'si²²jum⁵¹ bo¹¹
Dî *pât* *khî* *în*=é *museùm* *bo?*
 2.SG PF go 3.PL=GEN museum Q
 ‘Have you ever experienced going to their museum?’
 (PC0005-CLIN18)

- (144) ʔi⁵⁵ tjoʔ²² ʔwat²²djam⁵¹ gun⁵⁵ lon²²tsɔŋ⁵¹ gi²²na⁵⁵e²² mja³⁵
Î *tioh* *uatdiàm* *gûn* *longtsòng* *ginâ*=é *miá*.
 3.SG should recite 1.PL.EXC all kid=GEN name
 ‘They should recite our children’s names.’
 (PC0019-CLIN18)

³⁰ These are affixes that “signal grammatical relationships” (Crystal 2008:243) or, specifically, affixes that distinguish “different grammatical forms of the same lexical unit” (Matthews 2007:465).

³¹ Both *e*’s are clitics because they attach to phrases and words instead of just words (e.g., the genitive =é: [gûn=e yayá]=é “our helper’s”).

- (145) hi⁵⁵ge²² 'kwɪn²² 'ʔof²² 'ʔɪŋɡlan⁵⁵de²² sa³³ ya²² suj⁵⁵
Hige queen of Englând=ē sā yá suí.
 ART queen of England=GEN attire very pretty
 'The queen of England's dress is very pretty.'
 (PC0068, elicited)

The clitic =e /ʔe^{X1}/ functions as a modification marker. It indicates that the adjectival modifies the phrase following it (e.g., *suí-è* [suj⁵⁵e⁵¹] 'beautiful one'). I discussed this in detail in Section 3.4.7.

- (146) ʃem²²pɾe⁵¹ di⁵⁵ na³³ si³³ 'ha²²lo²²'ha²²loʔ⁵⁵e³³ laŋ²²'gweɖʒ⁵⁵
Syemprè dí nā sī halohalô=ē languagê...
 sure 2.SG if COP mixed=MOD language
 'Of course, if we are talking about a language that is mixed...'
 (PC0002-CLIN18)

- (147) ku²²kjaw³⁵ k^ha⁵⁵ tsa⁵⁵ le⁵⁵ tswɛ⁵⁵ səŋ²²dj⁵⁵ e³³ si²²tsun⁵¹
kukiaú khâ tsâ lê tsuê sungdî =ē sitsùn
 old.immigrants CPMV early PROG make business =MOD time
 'a time when old immigrants were making businesses earlier'
 (CLIN-19-119:21178)

3.4.10 Derivational prefixes

Lánngang-uè has Tagalog-origin prefixes that can change the part of speech (and meaning) of the word to which they are added.

The prefix *pang-* /paŋ⁵⁵/ combines with a verb (148, 149) or noun (150) to form an adjective that describes what an entity is reserved for. For instance, *pâng-* prefixes the verb *airfry* in (148) to form the adjective *pang-airfry* 'for-airfrying'. Before *machine*, this adjective indicates that the machine is reserved for air frying.

- (148) pa²²pa³⁵ tan⁵⁵sak⁵⁵ hi⁵⁵ge²² paŋ²² 'ʔɪ²²fɾaj⁵¹e³³ ma²²'ʃɪn⁵¹
Papá tânsák hige pang- airfry=ē machìne.
 father throw ART RES- airfry=MOD machine
 'Dad threw away the machine that is reserved for airfrying.'
 (PC0068, elicited)

(149) A: ʔin⁵⁵ bwe²²tswe⁵⁵ kaj²² ga²²law⁵⁵
În bo=etsuê kay- galâw ...
 3.PL no=PER CAUS- move
 ‘They cannot cause [it] to be moved.’

B: paŋ²² di²²po²²sit⁵⁵ 'laŋ⁵⁵ o⁵¹
Pang- deposit lâng o?
 RES- deposit only PRT
 ‘Is it only reserved for depositing?’

(PC0005-CLIN18)

(150) sjen²²si³³ kaj²² swe⁵⁵ hi⁵⁵ge²² paŋ²² ʔo²²to³⁵ŋe³³ sa³³
Siensî kay- suê hîge pang- ohtúng=ē sā.
 teacher CAUS-wash ART RES- school=MOD cloth
 ‘Teacher washed the clothes meant for school (thoroughly).’
 (PC0068, elicited)

The *tagà-* /ta⁵¹ga⁵¹/ prefix can denote origin or function, depending on the root it attaches to. If used with a verb, it denotes function: it attaches to the verb to form an adjective that specifies the function of an entity. In (151), *tagà-* attaches to *tsítsiáh* to form the adjective *taga-tsítsiáh* ‘cook-er/cook’. In the clause, this adjective modifies *î* ‘3.SG’, indicating that the entity functioned as a cook.

(151) 'ta²²ga²² tsi³⁵tsja³⁵ laŋ⁵⁵ la⁵¹ ʔi⁵⁵
Taga- tsítsiáh lâng la î.
 FUNC- cook only PRT 3.SG
 ‘He/she is only a cook.’
 (PC0068, elicited)

The same prefix can combine with a noun, specifically a location, to form an adjective that describes the origin of a person. For example, if a person is from China, a Lánnang-uè speaker can say that they are *taga-Taidiók* ‘Mainland Chinese’. Another example is given in (152).

(152) ʔi⁵⁵ si²² 'ta²²ga²² san²² 'pab²²lo⁵¹ ba³⁵
Î si taga- San Pablò bá?
 3.SG COP ORIG- San Pablo Q
 ‘Are they from San Pablo?’
 (PC0068, elicited)

The prefix *ka-* /ka⁵¹/ attaches to a Tagalog-, English-, or Hokkien-origin noun or a verb to form a noun that roughly means ‘colleague’. The noun is a location for gatherings (e.g., church, university, office) or a member of a group that has meetings or gatherings (e.g., churchmate, schoolmate, or officemate). Example (153) shows this: *ka-* prefixes the noun *trabahadòr* ‘laborer’ (e.g., construction worker) to form the noun *katrabahadòr* ‘laborer colleague’. It can also be attached to a location noun like *office* ‘office’ and a verb like *trabahò* ‘work’ to form the nouns *kaoffice* ‘office colleague’ and *katrabahò* ‘work colleague’, respectively.

- (153) kjaw³³ ʔin⁵⁵ hwaj³⁵ ʔin⁵⁵e³³ 'ka²²tɾa²²ba²²ha²²' do.i⁵⁵
kiaũ *în* *hwaí* ... *în=ē* ***ka-trabahadòr***.
 with 3.PL DEM.PL ... 3.PL=GEN COLL-worker
 ‘with those, their fellow workers’
 (PC0012-CLIN19)

Mag- /mag⁵⁵/ is a prefix that can mark/denote a relationship to a group. It derives an adjective by attaching to a Tagalog-, English-, or Hokkien-sourced noun that indicates a collective relationship (e.g., colleagues, classmates, lovers, spouses, friends). For example, the *mag-* prefixes *pieng-iũ* ‘friends’ in (154) to form the adjective *magpieng-iũ* that describes the noun being modified, *în nung é* ‘them both’, as having a relationship of friends.

- (154) hi⁵⁵ge²² ʔin⁵⁵ nəŋ²² e³⁵ si³³ 'mag²² pjeŋ²²ʔju⁵⁵
Hìge *în* *nung* *é* *sī* ***mag-*** *pieng-iũ*.
 ART 3.PL two CLS COP RLM- friend
 ‘The two of them are characterized as having a friendly relationship.
 Both are friends’
 (PC0068, elicited)

The prefix *pampà-* /pam⁵¹pa⁵¹/ can be attached to any verb or adjective, regardless of its source language, to form an adjective that describes a noun as cause of something. *Pampà-* prefixes to *gaú* ‘smart’ to form the adjective *pampagaú* ‘intelligence-causing’ in (155). In the sentence, this adjective modifies the noun phrase *hìge gamòt* ‘the medicine’, describing the phrase as the cause of intelligence.

- (155) hi⁵⁵ge²² ʔjo ʔ³⁵ si²² pam²²pa²² gaw³⁵e⁵¹
Hige *ióh* *si* ***pampa-*** *gaú* =*ē*.
 ART medicine COP CAUS- intelligent=MOD
 ‘The medicine is one that makes you intelligent.’
 (PC0068, elicited)

The prefix *pagka-* /pag⁵⁵ka⁵¹/ has two functions in Lánnang-uè: one indicates manner while the other indicates a state of being. When attached to a Tagalog-, Hokkien-, or English-origin verb, *pagka-* forms a noun that captures the manner of the verb. For example, in (156), the prefix is attached to the verb *tshām* ‘mix’ to create the noun phrase *pagkatshām* ‘manner of mixing’. It may occasionally be shortened to *pag-* by some speakers without change in meaning (e.g., *pagtshām* ‘manner of mixing’). When shortened, the first syllable of the verb root is sometimes reduplicated without perceptible consequences for meaning (e.g., example 157; *pagtshatshām* ‘manner of mixing’).

- (156) di⁵⁵ tuj⁵⁵ la³⁵naŋ²²ʔue⁵¹e²² pag²²ka²² ts^ham³³
Dî *tuí* *Lánnang-uè*=*ē* ***pagka-*** *tshām*
 2.SG towards Lánnang-uè=GEN MAN- mix

 ʔu²² ʃa²²miʔ⁵⁵ kam²²kak⁵⁵ bo⁵¹
u *shammîh* *kamkâk* *bò?*
 have what feel NEG

‘Do you feel anything towards Lánnang-uè’s manner of mixing?’
 (PC0012-CLIN19)

- (157) hi⁵⁵ge³⁵ si²² di⁵⁵ e³³ wej⁵¹ naŋ pag²² ʔi²² ʔi²²sip⁵⁵
Higé *si* *dî* =*ē* *wây* *ng* ***pag-*** *iisîp*
 DEM COP 2.SG =GEN way of MAN- think
 ‘That is your manner/way of thinking.’
 (PC0094-CLIN19)

When attached to a Tagalog-, Hokkien-, or English-sourced noun, *pagka-* /pag⁵⁵ka⁵¹/ creates a noun denoting a state. Attaching *pagkà-* to the noun *doctòr* ‘doctor’ or *Lánnáng* ‘Lannang’ creates the nouns *pagkadoctòr* ‘the state of being a doctor’ and *pagkalánnang* ‘the state of being a Lannang’.

Naka- /na⁵¹ka⁵¹/ can function as a prefix that denotes a resulting state or a state of wearing an article, depending on the root it prefixes. When attached to a Tagalog-, Hokkien-, or English-sourced verb, *naka-* forms an adjective that describes an entity’s position or state

resulting from performing the action designated by the verb. In (159), for example, *naka-* prefixes the Tagalog-derived *dikî* ‘stick’ to form the adjective *nakadikî*, which roughly describes the noun, *āntsām* ‘dirt’, as being stuck or, specifically, as being in a position or state directly resulting from the action of sticking.

- (158) so³³ hi³³ge⁵⁵ gin³³na⁵⁵ tsju³³ si³³ na²²ka²² pa²²toŋ⁵⁵
Sō, hīgê gīnnâ tsiū sī naka- patōng
 So, ART.SG kid then COP RSLT- place
 ti³³ di⁵⁵e³³ t^haw³⁵ lo⁵¹
tī deêr=ē thaú lô.
 PREP deer=GEN head PFV

‘So the kid is placed at the deer’s head.’
 (PC0103-FRST19)

- (159) gwa⁵⁵ kaj²² skiab⁵⁵ hwaj³⁵ na²²ka²² di²²kit⁵⁵e³³
Guâ kay- scrûb huai naka- dikî=ē
 1.SG CAUS- scrub DEM.PL RSLT- stick=MOD
 ʔan³³tsam³³ la³¹
āntsām la.
 dirt PRT

‘I scrubbed the stuck dirt.’
 (PC0072, spontaneous conversation)

When attached to a Tagalog-, English-, or Hokkien-derived noun that denotes an article (e.g., trinkets, clothes, weapons, mobile phones), the prefix *naka-* /na⁵¹ka⁵¹/ forms an adjective that describes an entity that wears or holds the article designated by the noun. For example, *naka-* can attach to *bakkia* ‘eyeglasses’ to form the adjective *nakabakkia*, which can be used to describe a person or animal that wears glasses.

The prefixes are summarized in Table 8.

Table 8. Derivational prefixes in the noun phrase domain

Prefix	Denotation	Attaches to	Derives	Example
<i>kà-</i>	colleague	Vb, N (location for gathering, person that has gatherings)	N	<i>ka-trabahò</i> 'workmate/colleague'
<i>mâg-</i>	relationship	N	Adj	<i>mag-pieng-iû</i> 'characterized as having a friendly relationship'
<i>nakà-</i>	resulting state	Vb	Adj	<i>naka-patông</i> 'being in a position or state directly resulting from the action of placing'
	wearing an article	N (article)	Adj	<i>naka-bakkà</i> 'characterized as wearing glasses'
<i>pag(kà)-</i>	manner	Vb	N	<i>pag(ka)-tshām</i> 'manner of mixing'
<i>pagkà-</i>	state-of-being	N	N	<i>pagka-lánnang</i> 'the state of being a Lannang'
<i>pampà-</i>	cause	Vb, Adj	Adj	<i>pampa-gaú</i> 'intelligence-causing'
<i>pâng-</i>	reservation	Vb, N	Adj	<i>pang-airfry</i> 'machine that is reserved for air-frying'
<i>tagà-</i>	function	Vb	Adj	<i>taga-tsítsiáh</i> 'cook-er/cook'
	origin	N (location)	Adj	<i>taga-San-Pablò</i> 'San Pablo-er/ originating from San Pablo'

3.5 Verb phrases

In this section, I identify and describe linguistic elements within the verb phrase (VP). I discuss modality, aspect, the ability suffix, causative markers, benefactive markers, directional markers, the resultative marker, negation, adverbial modification, and the copula. I discuss transitivity in Section 3.6.

3.5.1 Modality

In Lánnang-uè, expressing modality – be it possibility, desire, ability, permission, necessity, or prohibition – is done with particles and clitics derived from Hokkien and Tagalog. These auxiliaries are placed before the verb phrase.

To mark possibility, the Hokkien-derived \bar{e} /ʔe³³/ ³² or *bêh*³³ /beʔ⁵⁵/ particle is used. Both are used to express the high certainty of an action occurring in the future. While *bêh* is used particularly to indicate events or actions that are bound to happen soon, \bar{e} is used to emphasize that an event or action that will happen at an unspecified time.

- (160) ʔe³³ pjen⁵⁵tswe⁵⁵ 'miks⁵⁵ k^hi⁵¹ a¹¹
 \bar{E} piêntsue⁵⁵ mîx khi a.
 POS become mix DIR PRT
 '[It] will certainly become mixed (away from being pure) sometime in the future.'
 (PC0090-CLIN18)

- (161) gwa⁵⁵ ʔe³³ 'tekst⁵⁵ ʔi⁵¹
Guâ \bar{e} têxt ì.
 1.SG POS message 3.SG
 'I will certainly message them sometime in the future.'
 (elicitation, PC0068)

- (162) tsi⁵⁵ge²² gi²²na⁵⁵ bεʔ⁵⁵ k^hon⁵¹ ʔi³⁵tsuj³⁵
tsîge *ginnâ* **bêh** khùn itsuí...
 ART kid POS sleep before
 'before the kid will sleep soon...'
 (PC0020-FRST18)

- (163) gun⁵⁵ bεʔ⁵⁵ to⁵¹ khi¹¹ 'mu²²na⁵⁵
Gûn **bêh** tò khi munâ...
 1.SG POS return DIR first
 'We will soon return (to somewhere) first'
 (elicitation, PC0068)

When the possibility modal \bar{e} / ʔe³³/ is preceded by the negation particle *bo* or *be*, the clitic form of \bar{e} is used: = \bar{e} [e³³]. It is appended after the particle, as in:

³² This particle, /ʔe³³/, is not the same as the genitive-marking *e* /ʔe³⁵/ or the modifier *e* /ʔe^{X1}/. The three linguistic elements are distinct, non-homophonous morphemes.

³³ *Bêh* is multifunctional – it functions as a lexical verb before a noun phrase (e.g., *Guâ bêh hîge*. 'I want that') and functions as either a modal of possibility or modal of want/desire before a verb. See discussion in 3.5.1.

(164)	ka ^{22'} si ⁵⁵	ja ³⁵	tswe ²²	taj ²² djok ³⁵	laj ³⁵
	<i>Kasî</i>	<i>yá</i>	<i>tsue</i>	<i>Taidiok</i>	<i>lai</i>
	because	very	many	China	come
	bwe ³³	sî ³³	a ³¹		
	<i>bo=ē</i>	<i>sīnn</i>	<i>a.</i>		
	NEG=POS	bear.child	PRT		

‘Because very many China-comers won’t (certainly) bear children.’
(CLIN-19-122:23997)

The two modals in Lánnang-uè that express desire are *bêh* /be^{ʔ55}/ and *ai* /ʔaj⁵¹/. Both morphemes are multifunctional. They function as lexical verbs before nouns (e.g., *Guâ bêh tsige applè* ‘I want an apple’, *Guâ ai mi* ‘I love noodles’). Before verbs, however, they function as modal auxiliaries. As a modal, *bêh* has two functions: the first is expressing possibility, as discussed in the previous paragraph, and the second is want and/or desire.

(165)	gwa ⁵⁵	na ²²	bε ^{ʔ55}	mejn ^{22'} tejn ⁵¹	hok ²² kjen ⁵¹
	<i>Guâ</i>	<i>na</i>	<i>bêh</i>	<i>maintain</i>	<i>Hokkièn...</i>
	1.SG	CND	DES	maintain	Hokkien
	‘If I want to maintain Hokkien...’				
	(CLIN-19-105:16517)				

(166)	ʔin	na ³³	bε ^{ʔ55}	'tʃhendʒ ⁵⁵	la ⁵¹
	<i>În</i>	<i>nā</i>	<i>bêh</i>	<i>chânge</i>	<i>là...</i>
	3.PL	if	DES	change	PRT
	‘If they want to change ...’				
	(PC0012-CLIN19)				

There are some cases where the function of *bêh* is ambiguous. For instance, without context, the *bêh* in the earlier example be interpreted as a modal of possibility, as in:

(167)	ʔin	na ³³	bε ^{ʔ55}	'tʃhendʒ ⁵⁵	la ⁵¹
	<i>În</i>	<i>nā</i>	<i>bêh</i>	<i>chânge</i>	<i>là...</i>
	3.PL	if	POS	change	PRT
	‘If they are about to soon change...’				
	(elicitation, PC0068)				

In situations where the speaker wants to avoid this ambiguity, the Hokkien-derived desiderative modal *ai* /ʔaj⁵¹/ is used, as this modal does not express possibility at all. For example, if one

replaces the *bêh* with an *ai* modal from earlier example, the utterance will only have the desiderative reading.

- (168) ʔin⁵⁵ na³³ ʔaj⁵⁵ 'tʃejndʒ⁵⁵ la⁵¹
În nā aî chānge là...
 3.PL if DES change PRT
 'If they want to change...' NOT 'If they are about to soon change...'
 (judgment, PC0068)

Other examples involving the *ai* modal are as follows:

- (169) di⁵⁵ ʔu ʔaj⁵⁵ ʔwat⁵⁵ khi⁵¹ ʔa⁵⁵si²²bo²² lo⁵¹ a¹¹
Dî u aî uât khi âsi bo lò a?
 2.SG PF DES return DIR or NEG PFV PRT
 'Do you want to return [to somewhere] or not anymore?'
 (CLIN-19-68:17094)

- (170) gwa⁵⁵ ja³⁵ sjoŋ²² ʔaj⁵⁵ t^hja³³ tsi³³ga³⁵ tjoŋ²²kok⁵⁵e³³ ko⁵⁵si⁵¹ko⁵⁵
Guâ yá siong aî thia tsigá Tiongkok=ē kôsi kô.
 1.SG very often DES listen some China=GEN story PRT
 'I often desire to listen to some of China's stories.'
 (CLIN-19-108:17419)

To indicate one's ability to finish something given the circumstances, the Hokkien-derived particles *uhuâtlāng*, *ē*, and *ehiaû* are used. The first, *uhuâtlāng* [ʔu²²hwat⁵⁵laŋ³³/], indicates general ability and does not seem to have semantic restrictions. The particle [ʔu²²hwat⁵⁵laŋ³³] is used interchangeably with two other variants, reflecting pronunciation differences – *uhuâtnāng* [ʔu²²hwat⁵⁵naŋ³³] and *uhuâtthang* [ʔu²²hwat⁵⁵t^haŋ³³].

- (171) di⁵⁵ ʔu²²hwat⁵⁵laŋ³³ ʔeks²²p^ɿɛs⁵⁵ di⁵⁵ ka³³ki⁵¹ ko⁵⁵
Dî uhuâtlāng exprêss dî kaiki kô.
 2.SG ABI express 2.SG self PRT
 'You can express yourself.'
 (PC0009-CLIN19)

When preceded by the negative particle *bo* /bo³⁵/, the clitic forms of *uhuâtlāng*, *uhuâtnāng*, and *uhuâtthang* are used: =*huâtlāng* [hwat⁵⁵laŋ³³], =*huâtnāng* [hwat⁵⁵naŋ³³], and =*huâtlāng* [hwat⁵⁵t^haŋ³³]. They attach to the right of the negation particle.

- (172) gwa⁵⁵ bo²² hwat⁵⁵naŋ³³ tsjap⁵⁵fju³³ 'pa³¹
Guâ bo=huâtnang tsiâpshū pà ...
 1.SG NEG=ABI accept yet
 'I am not able to accept this yet...'
 (PC0091-CLIN19)

The multifunctional morpheme \bar{e} /ʔe³³/, which was earlier described as a modality of possibility, is also used as a modal of ability. However, unlike *uhuâtlāng* and its variants, the particle is specifically used to express an ability that can be characterized as innate or natural.

- (173) di⁵⁵ ʔe³³ tsa²²ʔjã⁵⁵ bo⁵¹
Dî ē tsa-iānn bô?
 2.SG ABI know Q
 'Are you able to know (naturally)?'
 (CLIN-19-68:23849)

When preceded by the negation particle *bo* /bo³⁵/ or *be* /be⁵¹/, the clitic form of \bar{e} is used: = \bar{e} [e³³]. It is appended after the negation particle, as in:

- (174) di⁵⁵ bwe³³ t^hja³³tjo^{ʔ31} gwa⁵⁵ le⁵⁵ koŋ⁵⁵
Dî bo=ē thiā-tioh guâ lê kông.
 2.SG NEG=ABI hear-PFV 1.SG PROG speak
 'You won't be able to successfully hear what I am saying.'
 (CLIN-19-68:24147)

The third ability modal, multifunctional *ehiaú* /ʔe³³hjaw⁵⁵/, on the other hand, is used to describe an ability that has been acquired through learning or training. Before a noun, the Hokkien-derived morpheme functions as a lexical verb meaning 'know'. Before a verb, the morpheme functions as a particle that expresses this modality of learned ability. In (175), the particle *ehiaú* is used before *kông* 'speak' to indicate that the subject can speak it by learning the linguistic variety, whereas \bar{e} is used before *biungpiák* 'understand' to indicate that *în* '3.PL' has a natural ability to understand Lánnang-uè.

(175) pe³³ro³³ ʔin⁵⁵ ʔe³³ bjeŋ²²pjak³⁵ kap⁵⁵ ʔe²²hjaw²²
Pēō *în* *ē* *biengpiák* *kâp* *ehiau*
 but 3.PL ABI understand and ABI

koŋ⁵⁵ la³⁵naŋ²²ʔue⁵¹ ko⁵⁵
kông *Lánnang-uè* *kô*.
 speak Lánnang-uè PRT

‘But they can (naturally) understand and speak (from learning) Lánnang-uè.’
 (PC0012-CLIN19)

The clitic form =*ehiaú* [e³³hjaw⁵⁵] is used when this ability modal is negated. It attaches to the right of the negation particle (e.g., *bō=ehiaū kông* [bwe³³hjaw³³koŋ⁵⁵] ‘unable to speak’).

In contrast with *ē* and *ehiaú*, Hokkien-derived *uhuâtnang* / *uhuâtlang* / *uhuâtthang* is also used as a modal of permission if the speaker wants to either give consent to the listener(s) or ask for consent from them.

(176) di⁵⁵ ʔu²²hwat⁵⁵thaj³³ laj²² lo⁵¹
Dî *uhuâtthang* *lai* *lò*.
 2.SG PER come PRT
 ‘You can/are permitted to come now.’
 (PC0068, elicitation)

The particles *ē-iêng* [ʔe³³ʔjeŋ⁵⁵], *ē-iûng* [ʔe³³ʔjəŋ⁵⁵], and *ētsuè* [ʔe³³tswe⁵¹] are interchangeably also used to indicate this modality of permission. Unlike *uhuâtnang* / *uhuâtlang* / *uhuâtthang*, however, these particles exclusively express permission.

(177) lan⁵⁵ ʔe³³ʔjeŋ⁵⁵ se^ʔ⁵⁵
Lân *ē-iêng* *sêh*.
 1.PL.INC PER say
 ‘We are permitted to say.’
 (CLIN-19-136:32619)

(178)	pag ²²	dī ⁵⁵	ʔaj ⁵⁵	ʔjeŋ ³³	hwa ²² na ³⁵ ʔwe ⁵¹
	<i>Pag</i>	<i>dī</i>	<i>aī</i>	<i>iēng</i>	<i>Huaná-uè...</i>
	if	2.SG	DES	use	Filipino
	ʔa ³³ si ³³		ʔe ³³ tswe ⁵¹	ʔjeŋ ⁵¹	
	<i>āsī</i>		<i>ētsuē</i>	<i>ièng.</i>	
	also		PER	use	

‘If you want to use Filipino, [you] are also permitted to use it.’
(CLIN-19-16:4934)

The clitic form =*ē-iēng* [e³³ʔjeŋ⁵⁵], =*ē-iûng* [e³³ʔjəŋ⁵⁵], or =*ētsuè* [e³³tswe⁵¹] is used in the context of negation. It is appended after the negation particle (e.g., *bo=ē-iēng sêh* ‘not permitted to say’).

To express the modality of prohibition (to instruct or warn the listeners not to do an action), the negative particles *māng* /maŋ³³/ and *methāng* /m̩³³ t^haŋ³³/ or *ūmethāng* /ʔom³³ t^haŋ³³/ are used. Both variants of the latter are interchangeable, with the *māng* variant more commonly used.

(179)	lan ⁵⁵	maŋ ³³	k ^h wa ⁵⁵ k ^h in ³³	hwi ³³ dī ³³ pin ³³ laŋ ³⁵	dīn ⁵⁵	la ⁵¹
	<i>Lân</i>	<i>māng</i>	<i>khuâkhîn</i>	<i>Huīdīpīn</i>	<i>láng</i>	<i>dîn là.</i>
	1.PL.INC	NEG.PROH	look.down	Philippines	person also	PRT
	‘Let’s not look down on the Filipinos.’					
	(PC0019-CLIN18)					

(180)	dī ⁵⁵	m̩ ³³ t ^h aŋ ³³	koŋ ⁵⁵
	<i>Dī</i>	<i>methāng</i>	<i>kông.</i>
	2.SG	NEG.PROH	speak
	‘Don’t say it.’		
	(CFH-001)		

To express necessity, the multifunctional Hokkien-sourced particle *tiôh* /tjo³⁵/ is used before the verb. Aside from being used to express completive aspect (see section 3.5.2), it is also used to express internal (e.g., physiological) or external (e.g., rules, laws, ethics) needs. In (181), for example, the speaker is questioning the need to sign a document based on ethics rules set by the research review board.

- (181) ka⁵⁵na⁵⁵ tjoʔ²² ts^hjam³³ mja³⁵ pa⁵¹
Kânâ **tióh** *tshiām* *miá* *pà?*
 why NEC sign name PRT
 ‘Why should I sign this? / Why do I need to sign this?’
 (PC0002-FRST18)

The modal particle *dapât* /'da⁵¹pat⁵⁵/ also has the same function as *tióh* /tjoʔ³⁵/. It is placed pre-verbally.

- (182) di⁵⁵ khi⁵⁵ le²²paj⁵¹ da²²pat⁵⁵ ʔu²² k^hwa⁵⁵tjoʔ²²
Dî *khî* *lepai* **dapât** *u* *khuâ-tioh*
 2.SG go worship NEC PF look-PFV PRT

 'tʃhejndʒ⁵⁵ la⁵¹
changê *là.*
 change PRT
 ‘(If) you go worshipping, you should have witnessed a successful change.’
 (PROT-16-NA:37896)

- (183) gwa⁵⁵ 'da²²pat⁵⁵ m̩³³ si³³ kjo⁵⁵ tswe⁵⁵ hwa³³kjaw³⁵
Guâ **dapât** *m* *sī* *kiô* *tsuê* *Huakiaú*
 1.SG NEC NEG COP call RSLT Chinese.immigrant

 lo²² e⁵⁵
lo *êh.*
 PFV PRT
 loosely: ‘The reason is that I should not be called a Chinese immigrant.’
 (CLIN-19-130:28638)

However, unlike *tióh*, it is also placed before and after clauses to mark that the clause itself is necessary. To the extent of my knowledge and research, there are no factors that condition the use of clause-initial *dapât* over clause-final *dapât*.

- (184) 'da²²pat⁵⁵ gwa⁵⁵ ʔe³³ hjaw⁵⁵ 'stiejt⁵⁵ hok²²kjen⁵¹
Dapât *guâ* *ē* *hiaú* *straight* *Hokkièn.*
 NEC 1.SG ABI know straight Hokkien
 ‘It ought to be that I am able to know straight Hokkien.’
 (CLIN-19-115:19600)

- (185) di⁵⁵ k^hi⁵⁵ le²²paj⁵¹ 'da²²pat⁵⁵
*Dî khî lepai **dapât**.*
 2.SG go worship NEC
 'It ought to be that you go to worship.'
 (PROT-16-NA:37896)

Dapât and *tióh* can co-occur in the same verb phrase to increase the degree of necessity and emphasis. When they do, *dâpat* must be placed before *tióh*.

- (186) di⁵⁵ 'da²²pat⁵⁵ tjo²² laj³⁵
*Dî **dapât** **tióh** lai*
 2.SG NEC NEC come
 'You really should come.'
 (elicitation, PC0068)

The Hokkien-derived negation particle *mién* is used to negate an expression of necessity (i.e., 'you do not need to V').

- (187) di⁵⁵ mjɛn³⁵ ʔin⁵¹ ʔi¹¹ a¹¹
*Dî **mién** ìn i a.*
 2.SG NEG.NEC answer 3.SG PRT
 'You need not answer them.'
 (CLIN-19-129:28130)

I summarize the modals of Lánnang-uè in Table 9.

Table 9. Modals in Lánngang-uè

Modal	Function	Modality	Source	Distribution
<i>ai</i>	indicates that the event or action is done out of the agent's desire or want; the general modal of desire, used to avoid <i>bêh</i> ambiguity	desiderative	Hokkien	__ VP
<i>bêh</i>	indicates that the event or action will happen soon	proximal possibility	Hokkien	__ VP
	indicates that the event or action is done out of the agent's desire or want	desiderative	Hokkien	__ VP
<i>dapât</i>	indicates internal (e.g., physiological) or external (e.g., rules, laws, ethics) needs; increases degree of obligation when with <i>tióh</i>	necessity	Tagalog	__ VP/S S __
<i>ē</i>	indicates certainty of action or event in an unspecified time	possibility	Hokkien	__ VP
	indicates innate or natural skill or ability to finish something given the circumstances	ability	Hokkien	__ VP
<i>ēhiaû</i>	indicates learned skill or ability to finish something given the circumstances	ability	Hokkien	__ VP
<i>ē-iêng</i> <i>ē-iûng</i> <i>ētsuè</i>	indicates that the action or event is something that has been permitted by the speaker	permission	Hokkien	__ VP
<i>māng</i> , <i>mthāng</i> , <i>ūmthāng</i>	indicates that the action is forbidden and should not be completed	prohibition	Hokkien	__ VP
<i>mién</i>	indicates negation of necessity	necessity	Hokkien	__ VP
<i>tióh</i>	indicates internal (e.g., physiological) or external (e.g., rules, laws, ethics) needs	necessity	Hokkien	__ VP
<i>uhuâtlāng</i> , <i>uhuâtthāng</i> , <i>uhuâtnāng</i>	indicates that the action or event is something that has been permitted by the speaker	permission	Hokkien	__ VP
	indicates general skill or ability to finish something given the circumstances; the general modal of ability	ability	Hokkien	__ VP

3.5.2 Aspect

Lánng-uè has grammatical aspect. It uses particles, adverbs, and other linguistic elements from Hokkien and Tagalog to mark the “duration or type of temporal activity” denoted by the verb or verb phrase (Crystal 2008:38; Matthews 2007). It has seven aspects, which are as follows:

1. Perfective
2. Perfect
3. Inchoative
4. Durative
5. Progressive
6. Iterative
7. Habitual

To indicate that a situation is seen as perfective or a completed whole (Crystal 2008:356; Matthews 2007:184, 686), four markers are used: Hokkien-derived *lò*, Hokkien-derived *diaû*, Hokkien-derived *-tioh*, and Tagalog-derived *kakà-*. When not used as an inchoative aspect marker, the multifunctional particle *lò* /lo^{X1}/ is used to mark perfective aspect. It is the default or ‘unmarked’ perfective marker used by speakers of Lánng-uè. It is placed after the verb phrase (e.g., *sī tuâ tsiá tuahàn* ‘grew up here’, *feèl yá kuí* ‘feel very high’), as shown in the examples below:

(188) gun⁵⁵ si³³ twa⁵⁵ tsja³⁵ twa²²han⁵¹ lo¹¹ 11
 Gûn *sī* *tuâ* *tsiá* *tuahàn* **lo** *a*
 1.PL.EXC COP PREP here grow.up PFV PRT
 ‘We grew up here.’
 (CLIN-19-132:30596)

(189) ‘ka²²si⁵⁵ ʔi⁵⁵ ‘fi⁵¹ ja³⁵ kuj³⁵ lo⁵¹ la
 Kasī *î* *feèl* *yá* *kuí* **lò** *la.*
 because 3.SG feel very high PFV PRT
 ‘Because he felt very proud (or above others).’
 (CLIN-19-132:30686)

- (190) ʔin⁵⁵ ʔi⁵⁵kjeŋ ʔa²²'si²²mi²²lejt⁵⁵ nan⁵⁵e³³ 'kul²²tʃuɿ⁵¹ lo¹¹
În ikiēng assimilâtê nân=ē culturè lo.
 3.PL already assimilate 1.PL=GEN culture PFV
 'They already assimilated to our culture.'
 (PC0002-CLIN18)

To emphasize the completion of an action or process as opposed to just marking an action or process as a whole event located at an undivided moment in time, the perfective particle *diaû* /djaw⁵⁵/ is used. The contrast between the *diaû* perfective and the general *lò* perfective is illustrated in the elicited examples (191) and (192) below. The verb phrases in both examples are viewed as action units without internal composition (e.g., continuous action, repetition). However, in (191), the verb phrase *hugâs huaí platò* 'wash the plates' is only implied to be complete while in (192) the completion of the verb phrase, washing the plates, is explicit.

- (191) gwa⁵⁵ 'hu²²gas⁵⁵ hwaj³⁵ 'pla²²to⁵¹ lo¹¹
Guâ hugâs huaí platò lo.
 1.SG wash ART.PL plate PFV
 'I [already] washed the plates.'
 (elicitation, PC0068)

- (192) gwa⁵⁵ 'hu²²gas⁵⁵ hwaj³⁵ 'pla²²to⁵¹ djaw⁵⁵
Guâ hugâs huaí platò diaû.
 1.SG wash ART.PL plate PFV
 'I finished washing the plates.'
 (elicitation, PC0068)

Two examples containing *diaû* from spontaneous conversations are as follows:

- (193) 'bas²²ta⁵¹ 'ʔin²²tɿ²²vju²² 'ʔin²²tɿ²²vju²² djaw⁵⁵ kjo⁵¹ ʔi¹¹
Bastâ interview interview³⁴ diaû kiò i
 enough interview interview PFV call 3.SG
 'Enough/stop and believe me, (they) called him as soon as they finished interviewing them for a certain amount of time.'
 (CE-001)

- (194) tsiah²² djaw⁵⁵ lo⁵¹
Tsiah diaû lò!
 eat PFV PFV
 '(I already) finished eating!'
 (E-004)

³⁴ The verb is reduplicated to code for durative aspect, which will be discussed later.

The third perfective marker is the Hokkien-derived suffix *-tioh* /tjoʔ^{X1}/.³⁵ When the multifunctional *tioh* morpheme is used before a verb phrase, it functions as a modal of possibility. However, when used after the verb, it functions as a perfective suffix that explicitly indicates that the completed action or process is successful (e.g., *ia-tioh* ‘successfully won’, *discover-tioh* ‘successfully discovered’, *fail-tioh* ‘successfully failed’, *tsiah-tioh* ‘successfully ate’, *lakâd-tioh* ‘successfully walked’). An example of the *-tioh* suffix in an utterance with no other perfective marker is as follows:

- (195) di⁵⁵ k^hwa⁵¹ tjoʔ¹¹ tswe⁵⁵ suŋ²²di³⁵ laŋ³⁵ la⁵¹
Dî khuâ -tioh tsuê sungdí lánŋ là.
 2.SG see -PFV make business person PRT
 ‘You (successfully) found business people.’
 (CLIN-19-134:31338)

The suffix can also co-occur with perfective particles. When it does, it adds a ‘successful completion’ reading to the utterance, as in the following:

- (196) hi⁵⁵ge²² gin²²na⁵⁵ k^hwa⁵¹tjoʔ¹¹ hi⁵⁵ge²² bu²²tj⁵¹
Híge ginnâ khuâ-tioh híge butì
 ART kid see-PFV ART bottle

 bo²² la²²man⁵⁵ lo⁵¹
bo lamân lò.
 no matter PFV
 ‘The kid was successful in seeing that the bottle had nothing.’
 (PC0071-FRST18)

The fourth is the Tagalog-derived prefix *kakâ-* /ka²²ka⁵¹/. This prefix indicates that the action is a recently completed action (e.g., *hiusiâk* ‘rest’, *kaka-hiusiâk* ‘just completed resting’). It attaches to Tagalog-, English-, and Hokkien-derived verbs. Some examples from spontaneous speech are as follows:

- (197) den²² hi⁵⁵ge²²kaw⁵⁵ o³⁵ ‘ka²²ka²²t^hi⁵⁵t^ho²²hi⁵⁵ge²² hajv⁵⁵lak⁵⁵ loʔ²² laj²²
Then híge kau ó kaka-thîtho híge hîve lâk loh lai.
 then ART dog PRT PFV-play ART hive drop fall DIR
 ‘Then just after the dog completed playing, the hive fell down.’
 (FRST-20-175:37052)

³⁵ I analyzed this as a suffix and not a clitic because *tioh* cannot only be attached to verbs and not verb phrases.

- (198) 'pe²²ro²² hi⁵⁵ge²²gin²²na⁵⁵ na²²'man⁵⁵ o³⁵ 'ka²²ka²²ts^he²²
Pero *hîge ginnâ* *namân* *ó* *kaka- tshe*
 but ART kid PRT PRT PFV- search
- hwaj³⁵ 'pa²²la²²ka[?]o³⁵
huai *palakâ* *ó...*
 DEM.PL frog PRT

‘But those kids, having just finished searching for the frogs...’
 (FRST-20-175:37056)

To indicate that the verb has perfect aspect, four markers are used: the *ũ* /ʔu³³/, *bè* /be⁵¹/, *pât* /pat⁵⁵/, and *lè* /le⁵¹/ particles. The multifunctional Hokkien-derived *ũ* morpheme is used as a general perfect marker with no known restrictions. Before a noun, it functions as a lexical verb meaning ‘have’ (e.g., *Guâ ũ hepatitis* ‘I have hepatitis’). Before a verb phrase, it is a particle that indicates a “state” resulting from a past action (Shopen 2007:304).

- (199) gwa⁵⁵ ʔu³³ dim³³ tsi²²ge²² 'kan⁵¹ naŋ²² 'bi⁵¹
Guâ ũ dîm tsiŋe *càn ng beèr.*
 1.SG PF drink ART can of beer
 ‘I have drunk a can of beer.’
 (elicitation, PC0068)

To negate the verb phrase with perfect aspect, Hokkien-derived *bè* /be⁵¹/ – a negative marker that also encodes perfect aspect – is used before the phrase, as in the following:

- (200) gwa⁵⁵ be³³ dim³³ tsi²²ge²² 'kan⁵¹ naŋ²² 'bi⁵¹
Guâ bē dîm tsiŋe *càn ng beèr.*
 1.SG NEG.PF drink ART can of beer
 ‘I have not drunk a can of beer.’
 (elicitation, PC0068)

- (201) di⁵⁵ be³³ t^hak²²ts^he?
Dî bē thaktshêh.
 2.SG NEG.PF study
 ‘You have not studied.’
 (CLIN-19-116:19717)

It is possible to have *ũ* co-occur with the plain perfective markers. In constructions with these markers, the utterances have the perfect reading on top of the whole, undivided action/process, explicit completion, and/or the successful completion reading:

- (202) gwa⁵⁵ ʔu³³ dim³³ tsi²²ge²² 'kan⁵¹ naŋ²² 'biɿ⁵¹ lo¹¹
Guâ ũ dīm tsige càn ng beer lo.
 1.SG PF drink ART can of beer PFV
 'I have drunk a can of beer.'
 (elicitation, PC0068)
- (203) gwa⁵⁵ ʔu³³ dim³³ djaw⁵⁵ tsi²²ge²² 'kan⁵¹ naŋ²² 'biɿ⁵¹
Guâ ũ dīm diaũ tsige càn ng beer.
 1.SG PF drink PFV ART can of beer
 'I have finished drinking a can of beer.'
 (elicitation, PC0068)
- (204) gwa⁵⁵ ʔu³³ dim³³ tioh³¹ tsi²²ge²² 'kan⁵¹ naŋ²² 'biɿ⁵¹
Guâ ũ dīm tioh tsige càn ng beer.
 1.SG PF drink PFV ART can of beer
 'I have successfully drunk a can of beer.'
 (elicitation, PC0068)

Examples containing *ũ* from spontaneous conversations are as follows:

- (205) gwa⁵⁵ tsa³³ma⁵⁵ ʔu³³ tap⁵⁵ di⁵¹ lo¹¹
Guâ tsāmâ ũ tâp dì lo.
 1.SG earlier PF answer 2.SG PFV
 'Earlier, I have already answered you.'
 (PC0004-CLIN18)
- (206) di⁵⁵ 'ta²²la²²ga⁵¹ ʔu²² ts^hoŋ⁵¹
Dî talagà u tshông ...
 2.SG really PF do
 'You've really done it...'
 (CLIN-19-133:30945)
- (207) nan⁵⁵ ʔu³³ k^hwa⁵¹tjo^ʔ¹¹ a¹¹
Nân ũ khuà-tioh a.
 1.PL.INC PF see-PFV PRT
 'We have successfully seen it.'
 (CLIN-19-134:31187)

Like *ũ*, the Hokkien-derived *pât* /pat⁵⁵/ particle is also multifunctional: before a noun, it is a lexical verb that means 'recognize' (e.g., *pât mamá* 'recognize mother'); before a verb, it functions as a perfect aspect marker. Unlike the perfect *ũ* particle, *pât* additionally emphasizes that the action or process in the verb phrase was experienced by the subject or experiencer of the

utterance. For example, the *pât* below highlights the fishermen’s experience of murdering fishermen from Taiwan on top of the perfect reading.

- (208) 'fi²²f.ɿ²²man⁵¹ pat⁵⁵ p^haʔ⁵⁵si⁵¹ tai²²wan³⁵e³³ 'fi²²f.ɿ²²man⁵¹ diⁿ⁵⁵
Fishermàn ***pât*** *phâhsì* *Taiwán=ē* *fishermàn* *dîn*
 fisherman PF kill Taiwan=GEN fisherman also
 ‘(The local) fishermen have experienced killing Taiwan’s fishermen too.’
 (CLIN-19-126:26804)

- (209) gun⁵⁵ pat⁵⁵ 'get²² tu²²'ge²²dɿ⁵¹ kio⁵¹ ʔi¹¹
Gûn ***pât*** *get* *togethèr* *kiô* *i...*
 1.PL.EXC PF get together call 3.SG
 ‘We have experienced, as one, calling them.’
 (CLIN-19-126:26842)

By default, an utterance with *pât* and no other temporal markers indicates that the completed prior action in the verb is relevant to the present.

- (210) gwa⁵⁵ pat⁵⁵ dim³³ tsi²²ge²² 'kan⁵¹ naŋ²² 'bi^r⁵¹
Guâ ***pât*** *dīm* *tsige* *càn* *ng* *beèr.*
 1.SG PF drink ART can of beer
 ‘I have experienced drinking a can of beer.’
 (elicitation, PC0068)

Compare the unmarked utterance above with a similar utterance below, where the time adverb *tsadit* ‘yesterday’ is added. Here, the entity’s experience of drinking of one can of beer is relevant to the past.

- (211) gwa⁵⁵ tsa²²dít³⁵ pat⁵⁵ dim³³ tsi²²ge²² 'kan⁵¹ naŋ²² 'bi^r⁵¹
Guâ *tsadít* ***pât*** *dīm* *tsige* *càn* *ng* *beèr.*
 1.SG yesterday PF drink ART can of beer
 ‘I had experienced drinking a can of beer yesterday.’
 (elicitation, PC0068)

If a speaker wants to indicate that the action is relevant to the future, they can use the possibility modal *ē* /eʔ³³/ and the *pât* particle in that order, before the verb.

- (212) gwa⁵⁵ bi²²na³⁵ ʔe³³ pat⁵⁵ dim³³ tsi²²ge²² 'kan⁵¹ naŋ²² 'bi^r⁵¹
Guâ *biná* ***ē*** ***pât*** *dīm* *tsige* *càn* *ng* *beèr.*
 1.SG tomorrow POS PF drink ART can of beer
 ‘I will have experienced drinking a can of beer tomorrow.’
 (elicitation, PC0068)

Using both perfect particles in a single verb phrase is possible in the variety. The utterance simply gets a perfect reading, with an emphasis on the experienced action or process. Speakers who wish to do so must place the *ū* particle before *pât*:

- (213) di⁵⁵ ʔu³³ pat⁵⁵ ʔiŋ⁵¹ bo¹¹
Dî *ū* *pât* *ìng* *bo?*
 2.SG PF PF use NEG
 ‘Have you ever experienced using it?’
 (PC0005-CLIN18)

Hokkien-derived *lè* /le⁵¹/ is multifunctional – before a verb phrase, it functions as a progressive aspect marker. After the verb, however, *lè* functions as a perfect marker. Post-verbal *lè* is used only if the entity subjected to the particular state expressed in the verb is the subject, unlike other perfect aspect markers (e.g., *Hîge thāngâ khuī lè* ‘The window [subject] is in an opened state’ but not **Guâ khuī lè hîge thāngâ* ‘I have opened the window [object] / I put the window in an opened state’).

- (214) gwa⁵⁵ tse²² le⁵⁵ t^haw²²tsuj³⁵
Guâ *tse* *lè* *thautsuí.*
 1.SG sit PF front
 ‘I am in a seated state in front (due to being seated prior)’
 (native speaker elicitation, PC0068)

- (215) ʔi⁵⁵ lak⁵⁵ le⁵⁵ hja³⁵
Í *lâk* *lè* *hiá.*
 3.SG fall PF there
 ‘They are in a fallen state there (due to being dropped prior)’
 (native speaker elicitation, PC0068)

- (216) so³³ji⁵⁵ ʔin⁵⁵e³³ t^ha³³ŋa⁵⁵ k^huj³³ le⁵¹
Sōyí *în=ē* *thāngâ* *khuī lè.*
 so 3.PL=GEN window open PF
 ‘So their window is an opened state (due to being opened prior).’
 (PC0004-CLIN18)

- (217) tsju³³ si³³ bo²² saŋ³⁵e³³ 'lan²²gwejd³⁵hap²² le⁵¹ lo¹¹
Tsiū *sī* *bō* *sáng=ē* *languâge* *hap* *lè* *lo.*
 at.onceCOP not same=MOD language combine PF PFV
 ‘It is the case that different languages completed combining and are in a state of combination (due to them being combined prior)’
 (PC0002-CLIN18)

To indicate inchoative aspect (the initiation of some process or action), two markers are used: the Hokkien-derived *lò* /lo^{X1}/ and the Tagalog-derived prefix *pà-* /pa⁵¹/. The *lò* particle, as shown earlier, is used to mark perfective aspect. But when it is not, it specifically indicates that an action or event – usually one that occurs or unfolds naturally (e.g., the sky getting dark) – is in the process of beginning. The particle is placed after the verb phrase.

- (218) gwa²²bin³⁵ ʔam²²'bun⁵⁵ lo⁵¹
Guabín *ambûn* ***lò.***
 outside drizzle INCH
 ‘It is starting to drizzle now (and it did not before).’
 (PC0068, elicitation)

- (219) ma²²ma³⁵ khja³³ hja³⁵ lo⁵¹
Mamá *khiā* *hiá* ***lò.***
 mother live there INCH
 ‘Mother (is beginning to) live there now (and she did not before).’
 (PC0068, elicitation)

The prefix *pà-* /pa⁵¹/ functions as an inchoative marker, similar to *lò*, when not attached to an adverb as an emphatic marker. Unlike *lò*, it is attached before the verb and not after the verb phrase.

- (220) ma²²ma³⁵ kaj²²pa²²ku²²loʔ⁵⁵ hi⁵⁵ge²² lu²²gaw⁵⁵
Mamá *kay-pa-kulô* *hîge* *lugâw.*
 mother CAUS-INCH-boil ART congee
 ‘Mother caused the congee to start boiling.’
 (PC0068, elicitation)

Before an adjective, the morpheme is used to derive a verb that denotes ‘getting Adj’; it can be attached to English-, Tagalog-, and Hokkien-derived adjectives. For instance, in the following examples, inchoative *pà-* is attached to the Hokkien-derived adjective *kui* ‘expensive’ and English-derived *shiny* ‘shiny’ to mean ‘getting expensive’ and ‘getting shiny’, respectively.

- (221) bi³³ko⁵⁵ke³³ miŋ³³kja³³ pa²²'kuj⁵¹ 'naŋ²²
Bīkōk=ē *mīngkiā* *pa-kui* *ng*
 America=GEN thing INCH-expensive LNK.ITER
- pa²²'kuj⁵¹ 'no³⁵
pa-kui *nóh?*
 INCH-expensive PRT

‘American things are getting more and more expensive, right?’
 (PC0072)

- (222) hi⁵⁵ge²²'kojn⁵¹ pa²²'fajni⁵¹ lo¹¹
Híge coin pa-shiny lo.
 ART coin INCH-shiny PFV
 ‘The coin is getting shiny.’

(PC0068, elicitation)

To mark durative aspect, three strategies are used. The first, total reduplication, is the general strategy. An identical copy of the verb is placed adjacent to the verb to indicate that the action or process lasts a certain, unspecified amount of time. The verb can only be copied once and there are no restrictions regarding the source language of the verb.

- (223) ts^hu⁵⁵ k^hi⁵⁵ kja²² kja³⁵ le⁵¹
Tshû khî kia kiá lè...
 exit DIR walk walk PF
 ‘(They) went out walking for some unspecified time.’
 (CLIN-19-127:27092)

- (224) lan⁵⁵ tsaw²² tsaw³⁵ go²² lak²² tsap²² ni³⁵
Lân tsau tsáú ... go, lak tsap ní.
 1.PL.INC run run five six tens year
 ‘We run for the foreseeable future, fifty to sixty years.’
 (FRST-19-115:19666)

- (225) 'bas²²ta⁵¹ 'ʔin²²tɿ²²vju²² 'ʔin²²tɿ²²vju⁵¹ djaw⁵⁵ kjo⁵¹ ʔi¹¹
Bastà interview interview diaû kiò i
 enough interview interview PFV call 3.SG
 ‘Enough/stop and believe me, (they) called him as soon as they finished interviewing them for a certain amount of time.’
 (CE-001)

The second strategy – the use of Hokkien-derived *tsè* /tse^{X1}/ after the verb phrase – has a similar function. Unlike total reduplication, however, the use of the *tsè* particle indicates that the duration of the action or process is short.

- (226) na³³ si³³ tsja³⁵e³³ laŋ³⁵ tjo^{ʔ22} ko⁵⁵ sjũ³³ tse³¹
Nā sī tsiá=ē lánŋ ... tioh kô siūnn tsè...
 if COP here=MOD person NEC again think DUR
 ‘If it is a local, you will need to think for a little bit.’
 (PC0092-CLIN19)

The final strategy is the use of Tagalog-derived adverb *munâ* /mu⁵¹na⁵⁵/ ‘first’ after the verb phrase. *Munâ* and *tsè* have identical functions and are used interchangeably. They can also co-occur in the same utterance without additional changes in meaning. If they do, *tsè* must be placed before *munâ*.

- (227) 'pe²²ro²² di⁵⁵ ʔu³³ khⁱ⁵⁵ sja²²mi^{ʔ35} ho²²'tel⁵¹
Pero dí ū khî siammih hotèl
 but 2.SG PF go what hotel
- tswe⁵⁵kaŋ³³ 'mu²²na⁵⁵ di³³'ba³⁵
tsuékāng munâ dibá?
 work first PRT

‘But you have gone to some hotel to work for a little bit, right?’
 (CE-001)

- (228) di⁵⁵ tsja^{ʔ22} tse⁵¹ 'mu²²na⁵⁵
Dî tsiah tsè munâ.
 2.SG eat DUR first
 ‘You eat for a little while.’
 (elicitation, PC0068)

The progressive aspect is marked using the multifunctional Hokkien-derived particle *lè* /le⁵¹/. When not used as a perfect marker, *lè* indicates that the action or process in the verb phrase is in progress without necessary time limits. It is placed before the verb phrase, as in the following:

- (229) 'juŋ³³ ma²²'ŋa³³ le⁵⁵ laj²² tsja³⁵ o³⁵ si³³
Yūng mgā lē lai tsiá ó si
 DEM PL PROG come here PRT COP
- ʔan⁵⁵ 'pja²²vins⁵⁵ laj²²e⁵¹
ân provincê lai=e
 PREP province come=MOD

‘Those who are coming here are from the province.’
 (C-001)

- (230) hi³³ge⁵⁵ gi³³na⁵⁵ le⁵⁵ khun⁵¹
Hīgê gīnâ lē khùn.
 DET kid PROG sleep
 ‘The kid is sleeping.’
 (FRST-19-98:14846)

To express the iterative aspect, two strategies are used. The first one is total reduplication of the verb phrase with the multifunctional Tagalog-derived *ng* /naŋ⁵⁵/ linker morpheme in between the verbs to indicate that the action or process is repeatedly done at a certain point in time or in a single event. When placed before a noun phrase, it functions as a preposition (Section 3.8); when placed between reduplicated verbs, it functions as an iterative aspect marker.

- (231) ʔi⁵⁵ tsjaʔ³⁵ naŋ²² tsjaʔ²² la⁵¹
Í tsiáh ng tsiáh là.
 3.SG eat LNK.ITER eat PRT
 ‘They ate (multiple times in a single event).’
 (elicitation, PC0068)

It is possible to reduplicate the verb more than once to highlight the degree of repetition, such as in (232).

- (232) t^hjā³³ naŋ²² t^hjā³³ naŋ²² t^hjā³³ 'daw⁵⁵ la⁵¹
Thiānn ng thiānn ng thiānn dâw là.
 listen LNK.ITER listen LNK.ITER listen PRT PRT
 ‘I was told that they listened (so many times in a single event).’
 (PC0005-CLIN18)

The second strategy is to use the Hokkien-derived particle *tīī* /ti³³ti³³/ before the verb phrase, such as in (233).

- (233) ʔi⁵⁵ ti³³ti³³ ts^{he51} pa²² din⁵⁵
Í tītī tshè padín.
 3.SG ITER search still.
 ‘They still searched (multiple times in a single event).’
 (PC0071-FRST18)

The first and second strategies are used interchangeably. To my knowledge, there are no factors that condition the use of the first strategy over the second.

Combining both strategies is possible. In such a case, the *tītī* particle must first be placed before the main verb before the reduplication of the verb phrase (234). Doing so does not encode additional meaning or function (e.g., greater emphasis).

- (234) ti³³ti³³ ts^{he51} naŋ²² ti³³ti³³ ts^{he51} ʔi¹¹
Tītī tshè ng tītī tshè i.
 ITER search LNK.ITER ITER search 3.SG
 ‘They searched and searched for them (in a single event).’
 (PC0071-FRST18)

To indicate habitual aspect (or indicate that an action or process is “characteristic of an extended period of time” (Comrie 1976:27)), Lánnang-uè does not use a distinct particle or a morphological process. Instead, it combines the perfect aspect marker *ū* /ʔu³³/ and the progressive marker *lê* /le⁵⁵/. The particles are placed in that order before the verb phrase. In (235), for example, the speaker uses both *ū* and *lê* to ask whether the listener has been thinking of resigning (habitually).

- (235) 'pe²²ro²² di⁵⁵ ʔu³³ le⁵⁵ sju³³ beʔ⁵⁵ .i²² sajn⁵¹
Pero dí ū lê siū bēh resìgn
 but 2.SG PF PROG think DES resign
- lo¹¹ ba³⁵
lo bá?
 INCH Q

‘But are you starting to think (habitually) about wanting to resign?’
 (CE-001)

Other examples from natural speech are as follows:

- (236) k^{ha55}nan⁵⁵ tʃ^hju³³ lan⁵⁵ k^{ha55} ʔu³³ le⁵⁵ tʃjaw⁵⁵
Khânân tshiū lân khâ ũ lê chiaû
 as like 1.PL.INC CMPV PF PROG follow
- hwaj²² 'ruls⁵⁵
huai rûle-s.
 ART.DEF.PL rule-PL

‘It is like we have been following (habitually) the rules better.’
 (PC0009-FRST19)

- (237) gwa⁵⁵ ʔu³³ le⁵⁵ koŋ⁵⁵ ʔjeŋ³³bun³⁵
Guâ ũ lê kông iēngbún
 1.SG PF PROG speak English
- ʔik²²'sept⁵⁵ 'pag³³ ʔu³³ 'si²²kiet⁵⁵
excêpt pāg ũ secrét.
 except CND have secret

‘I have been speaking (habitually) in English, except when there is a secret.’
 (C-001)

The aspect markers I discussed, along with their function(s), source, and distribution, are summarized in Table 10.

Table 10. Aspect markers in Lánngang-uè

Aspect marker	Function	Aspect	Source	Distribution
<i>bè</i>	negative counterpart of <i>ũ</i> ; indicates that the action or state is being invoked as a prior to a specific or implicit moment in time	perfect	Hokkien	__ VP
<i>diaû</i>	indicates a complete event, process, or action located at an undivided moment of time; emphasizes completion	perfective	Hokkien	Vb __ VP __
<i>kakà-</i>	indicates an event, process, or action that has been completed recently	perfective	Tagalog	__ Vb
<i>lè</i>	indicates a state resulting from a prior action	perfect	Hokkien	Vb __

	indicates that the action or process is in progress without necessary time limits	progressive	Hokkien	__ VP
<i>lò</i>	indicates a complete event, process, or action located at an undivided moment of time, regardless of the time contrasts which may be a part of it; the general perfective marker	perfective	Hokkien	VP __
	indicates the initiation of some process or action, usually one that occurs or unfolds naturally	inchoative	Hokkien	VP __
<i>munâ</i>	indicates continuity for a short unfixed period of time	durative	Tagalog	VP __
<i>ng</i>	indicates repetition	iterative	Tagalog	VP __ VP ... __ VP
<i>pà-</i>	indicates the initiation of some process or action	inchoative	Tagalog	__ Vb/Adj
<i>pât</i>	like <i>ū</i> , but also used to emphasize that the action or process was experienced by the subject/agent	perfect	Hokkien	__ VP
<i>-tiòh</i>	indicates successful completion	perfective	Hokkien	Vb __
<i>tīī</i>	indicates repetition	iterative	Hokkien	__ VP
total reduplication	indicates continuity for an unfixed period of time	durative	Hokkien? Tagalog?	VbVb
<i>tsè</i>	indicates continuity for a short unfixed period of time	durative	Hokkien	VP __
<i>ū</i>	indicates that the action or state is being invoked as a prior to a specific or implicit moment in time	perfect	Hokkien	__ VP
<i>u + lê</i>	indicates something done consistently or habitually	habitual	Hokkien	__ VP

3.5.3 The ability suffix *-dit*

The suffix *-dit* /dit^{X1}/, derived from Hokkien, creates adjectives that describe entities as being possible or capable of/suitable for the action or process indicated by the verb, analogous to the

English *-able* suffix. The ability suffix attaches to any Tagalog-, English-, or Hokkien-sourced verb. For instance, in (238), the suffix *-dit* was attached to Hokkien-sourced *thîthó* ‘play’ to form the adjective *thîthó-dit* [t^hi⁵⁵t^ho²²dit⁵¹] ‘playable’.

- (238) dî⁵⁵ ho³³ ʔi⁵⁵ hi⁵⁵ge²² t^hi⁵⁵t^ho²²dit⁵¹e¹¹ toj⁵¹
Dî hō î hîge thîthó-dit=ē tòy.
 2.SG give 3.SG ART play-ABI=MOD toy
 ‘Give them the playable toy.’
 (elicitation, PC0068)

Other examples can be found below.

- (239) hi⁵⁵ge²² sjen³³si³³ .i²²ko.ɪd⁵⁵dit⁵¹ bo¹¹
Hîge siēnsî recôrd-dit bo?
 DEM.SG teacher record-ABI NEG
 ‘Is the teacher recordable?’
 (elicitation, PC0068)

- (240) bwe³³ koŋ⁵⁵dit⁵¹ lo¹¹ a¹¹
Bo=ē kông-dit lo a.
 NEG=ABI say-ABI PFV PRT
 ‘(It is) not able to be commentable (naturally).’
 (CLIN-19-120:22123)

- (241) ʔin⁵⁵ bwe³³ pa²²sen²²ʃa⁵¹dit¹¹ a¹¹
În bo=ē pasensyà-dit a.
 3.PL NEG=ABI pardon-ABI PRT
 ‘They are not able to be pardonable (naturally).’
 (CLIN-19-122:23823)

3.5.4 Causative markers

Lánnang-uè has three causative markers: the *kāy-* prefix, the *tsiōng* particle, and the *hō* particle. The first – the *kāy-* /kaj³³/ prefix – is attached to any verb to derive another verb with the meaning ‘cause an entity to be Verb-ed (by a causer)’. The prefix, which cannot be definitively traced back to a source language, increases the valency – or the “range of syntactic elements either required or specifically permitted by a verb” (Matthews 2007:982) – of the verb it attaches to.

In constructions with an intransitive verb (e.g., S + V), the addition of the *kay-*, as in S + *kay-* + V, increases the number of arguments required by the verb. In (242), for example, the intransitive verb *hulôg* ‘drop’ without the prefix only requires one argument (i.e., the subject

mama ‘mother’). With the prefix *kay-*, the verb requires two arguments (i.e., the subject/causer and the object/causee) (243). The sentence is interpreted as *mamá* ‘mother’ causing *hîge rôck* ‘the rock’ (the causee) to be dropped.

- (242) ma³³ma³⁵ ʔan⁵⁵ hja³⁵ hu²²log⁵⁵
Mamá *ân* *hiá* *hulôg*.
 mother at there drop
 ‘**Mother** dropped/fell there.’
 (elicitation, PC0068)
- (243) ma³³ma³⁵ ʔan⁵⁵ hja³⁵ kaj²²hu²²log⁵⁵ hi⁵⁵ge²² .iak⁵⁵
Mamá *ân* *hiá* *kay-hulôg* *hîge* *rock*.
 mother at there CAUS-drop ART rock
 ‘**Mother** caused **the rock** to be dropped there.’
 (elicitation, PC0068)

In constructions with a transitive verb (e.g., S + V + O), the addition of the prefix, as in S + *kay-* + V + O, also increases the valency of the verb. In (244), for instance, the utterance is only interpreted as having two arguments, the subject ‘I’ and the object ‘the money’ (refer to the example gloss). The utterance with the prefixed verb (245), on the other hand, is interpreted as having three: the subject ‘I’, the patient/causee ‘the money’, and the agent/causer ‘myself/I’. In the absence of an explicitly stated agent or causer, the subject is interpreted as the causer (245).

- (244) gwa⁵⁵ le⁵⁵ ʔiks²²ʔfendʒ⁵⁵ hi⁵⁵ge²² tsi³⁵
Guá *lê* *exchânge* *hîge* *tsí*.
 1.SG PROG exchange ART.DEF money
 ‘**I** am exchanging **the money**.’
 (elicitation, PC0068)
- (245) gwa⁵⁵ le⁵⁵ ka²²ʔiks²²ʔfendʒ⁵⁵ hi⁵⁵ge²² tsi³⁵
Guá *lê* *kay-exchânge* *hîge* *tsí*.
 1.SG PROG CAUS-exchange ART.DEF money
 ‘**I** am causing **the money** to be exchanged (by **myself**).’
 (elicitation, PC0068)

I provide examples of *kay-* use from natural speech below:

- (246) 'juŋ²² dʒa³³ pan⁵¹ kaj²² bam⁵¹ 'tʃhaj³³ na⁵¹
Yung *Jāpàn* *kay-bòmb* *Chīnà*...
 when Japan CAUS-bomb China
 ‘When Japan caused China to be bombed (by Japan)...’
 (C-002)

- (247) ʔi⁵⁵ ma³³si³³bo²² kaj²²k^hwa⁵⁵ ʔi⁵⁵e³³ 'den²²tal²² pi²²'fe²²ʃon⁵¹
 Î māsī bo **kay-khuâ** î=ē dental profession
 3.SG still NEG CAUS-see 3.SG=GEN dental profession
 e³³ 'laj²²sens⁵⁵ pa²² 'ŋa⁵⁵
 =ē licensê pa ngâ.
 MOD license still PRT

‘They still haven’t caused their dental professional license to be seen/inspected (by themselves).’
 (CE-001)

The second marker – Hokkien-sourced *tsiōng* /tsjɔŋ³³/ – also has the same function as *kay-*. It indicates that the subject caused an entity (the object) to experience the verb in the sentence. However, it changes the order of the constituents into S + *tsiōng* + O + V (see section 3.6). It also does not increase the valency of the verb in the sentence.

- (248) ʔend²² sjɔŋ²²te⁵¹ ʔe³³ tsjɔŋ³³ di⁵⁵e³³ 'lajf⁵⁵ o³⁵ 'blɛs⁵⁵ lo⁵¹
 And Siongtè ē **tsiōng** dī=ē lifê ó blêss lò.
 and God POS CAUS 2.SG=GEN life PRT bless PFV
 ‘And God will cause your life to be blessed by him.’
 (PROT-16-NA:37972)

The third causative marker is the multifunctional morpheme *hō* /ho³³/, derived from Hokkien. In ditransitive sentences that have verbs of giving, it functions as a dative marker in an S + V_{giving} + O_{direct} + dative *hō* + O_{indirect} construction (section 3.6). The morpheme can also function as the verb itself – specifically, a verb meaning ‘to give’ (e.g., *Guâ ho î tsige gift*. ‘I gave them a gift.’). Finally, the morpheme can also be a causative marker that can be interpreted in two different manners – the intentional causative and the unintentional causative (the ‘passive’).

In the first interpretation, the causation involves voluntary action on the subject/agent’s part, as in the examples below.

- (249) gwa⁵⁵ ho³³ gun⁵⁵ k^han²²tʃju⁵⁵ ʔa²²plaj⁵¹ fi²²li²²pi²²no²² si²²ti²²zen⁵¹
 Guâ **ho** gûn khantshiû apply Filipino citizèn.
 1.SG CAUS 1.PL spouses apply Filipino citizen
 ‘I caused us spouses to apply for Filipino citizenship.’
 (CLIN-19-68:40227)

- (250) gwa⁵⁵ ho²² di⁵⁵ k^hwa⁵¹
Guá ho dí khuà ...
 1.SG CAUS 2.SG look
 ‘I caused you to look.’
 (CLIN-19-68:23872)

In the second interpretation, the causation is by virtue of the mere existence of an object, event, or state, as expressed by the subject. This ‘unintentional causative’ interpretation can also be regarded as the ‘passive’ interpretation (section 3.6).

- (251) lan⁵⁵ ho³³ tsin³³ tswe³³ bo²² sa³³saŋ^{35e 11} miŋ³³kjã³³ k^hi⁵⁵ ʔjeŋ³³hjoŋ⁵⁵
Lân hō tsīn tsuē bo sāsáng=e mīngkiānn k^hi iēng^hiōng.
 1.PL CAUS very many NEG same=MOD thing go influence
 ‘We were influenced by very many different things. (literally, Our state of existence unintentionally caused very many different things to influence us.)’
 (CLIN-18-4:1268)

The exact interpretation of the causative *hō*, to my knowledge, is determined via context.

Unlike *kay-*, which is placed before a verb, *hō* is placed before a noun (section 3.6); it also does not increase the valency of the verb in the sentence like *tsiōng*. Unlike both *kay-* and *tsiōng*, *hō* is used in causative constructions where the subject causes an entity to *do* an action/process, rather than *experience* it.

3.5.5 Benefactive markers

Lánng-uè has two markers that indicates that the entities indicated in the object benefit from the action of the verb – Hokkien-derived *kā* /ka³³/ and *kāng* /kaŋ³³/. The first is used more frequently than *kāng*. To my knowledge, there are no factors that condition the use of *kā* over *kāng*; they are used interchangeably. These markers change the default transitive word order (section 3.6).

- (252) ʔu²² laŋ³⁵ bεʔ⁵⁵ ka³³ di⁵⁵ kaj³³sjak⁵⁵ la⁵¹
Ū lánng bēh kā dí kaīsiák là.
 have person will BEN 2.SG explain PRT
 ‘There will be someone who will explain it to you for your benefit.’
 (E-004)

(253) bas²²ta⁵¹ gwa⁵⁵ ʔe³³ kaŋ³³ ʔin⁵⁵ ka⁵¹ la¹¹
Bastà *guâ* *ē* *kāng* *în* *kà* *la*.
 PRT 2.SG will BEN 3.PL teach PRT
 ‘No matter what, I will teach them for their benefit.’
 (PC0019-CLIN18)

(254) di⁵⁵ kaŋ³³ gwa⁵⁵ 'tʃhek⁵⁵ʃa²²mɪʔ⁵⁵ a⁵¹
Dî *kāng* *guâ* *chêck* *siammîh* *à?*
 2.SG BEN 1.SG check what PRT
 ‘What did you check for my benefit?’
 (PC0019-CLIN18.eaf)

3.5.6 Directional markers

Lánngang-uè has two multifunctional morphemes that mark verb direction: *khì* [khi⁵¹] and *lai* [laj³⁵], both Hokkien-derived. After a noun, these morphemes function as lexical verbs that mean ‘go’ and ‘come’, respectively. After a verb, they function as particles that indicate its direction (e.g., the direction the agent is enacting an action, the direction the patient is undergoing an action or process). *Khì* roughly means ‘away’ whereas *lai* denotes ‘towards’.

(255) hi⁵⁵ge²² ʔawl⁵¹ tsa⁵⁵ pɛʔ⁵⁵ pɛʔ⁵⁵ k^hi⁵⁵ tsi²²ge²² ja³⁵
Hîge *òwl* *tsâ* *pêh* *pêh* *khî* *tsige* *yá...*
 ART.DEF.SG owl then climb climb DIR ART.INDEF.SG very

 twa²² twa²² te⁵⁵ tsjoʔ²² t^haw³⁵
tua *tua* *tê* *tsiohtháú*.
 big big CLS rock

‘The owl then climbed away (for some time) to a rock that is very... sort of big.’
 (FRST-19-115:19677)

(256) ʔin⁵⁵ tjoʔ²² tsaw³⁵ tsaw³⁵ tsaw²² k^hi⁵⁵ ts^he⁵¹
În *tioh* *tsáú*, *tsáú*, *tsáú* *khî* *tshè*.
 3.PL NEC run run run DIR find
 ‘They should run (away) for some time to find.’
 (FRST-19-116:19841)

(257) hi⁵⁵ tsjaʔ⁵⁵ ts^haŋ²²ka²²ba⁵⁵ pɛʔ⁵⁵ ts^hu⁵⁵ laj⁵¹ la¹¹
Hî *tsiâh* *tshangkabâ...* *pêh* *tshu* *lai* *la*.
 DEM CLS frog climb out DIR PRT
 ‘That frog ... climbed out (towards something/someone).’
 (FRST-19-119:21745)

- (258) gwa⁵⁵ ʔwat⁵⁵ laj²² tsi⁵⁵ ta ʔ⁵⁵e³³ si²²tsun⁵¹
 guâ ... uât **lai** tsî tâh=ē sitsùn
 1.SG return DIR DEM CLS=MOD time
 ‘a time where I returned towards this place’
 (CLIN-19-120:21815)

3.5.7 Resultative marker

Lánnang-uè uses the multifunctional Hokkien-derived *tsuè* /tswe⁵¹/ morpheme for resultatives.

After a noun, *tsuè* is as a verb that means ‘work’ or ‘produce’ (e.g., *Guâ lê tsuê phai* ‘I produce bad [things]’); after a verb or verb phrase, it is a resultative particle indicating that the action or process expressed in the verb or verb phrase results in a state represented by the constituent/complement after *tsuè*.

If the particle is placed after a verb (i.e., S + V + *tsuè* + complement), the particle indicates that the action or process (V) resulted in the patient’s (S) state of being something (complement).

- (259) hi⁵⁵ge²² laŋ³⁵ ‘ba²²lot⁵⁵ tswe⁵⁵ bu²²’ .ii²²to⁵¹
Hige láng balôt tsuê burritò.
 ART.DEF.SG person wrap RSLT burrito
 ‘The person is wrapped in a burrito-like state.’
 (elicitation, PC0068)

If the particle is placed after a verb phrase (i.e., S + VP + *tsuè* + complement), the particle indicates that the subject is responsible for the action or process (the V in the VP) that resulted in the patient’s (the NP complement in VP) state of being something (complement).

- (260) hi⁵⁵ge²² laŋ³⁵ ‘ba²²lot⁵⁵ hi⁵⁵ge²² du⁵¹we⁵¹tswe⁵⁵bu²²’ .ii²²to⁵¹
Hige láng balôt hige dúwè tsuê burritò.
 ART.DEF.SG person wrap ART.DEF.SG girl RSLTburrito
 ‘The person wrapped the woman in a burrito-like state.’
 (elicitation, PC0068)

Some examples exemplifying the use of the resultative particle in spontaneous conversations are found below.

- (261) so³⁵ⁱ²² di⁵⁵ pjen⁵⁵ tswe⁵⁵ k^{hi55} ʔoʔ²²tjoʔ²² ʔi⁵⁵e³⁵
Sói dí piên tsuê khî oh-tioh î=é.
 So 2.SG change RSLT go learn-PFV 3.SG=GEN
 ‘So you are in a state of learning their (language) successfully (as a result of being changed into that state)’
 (CLIN-19-116:19822)

- (262) tɕəŋ²²paj⁵⁵ k^{ha55} bwe³³hjaw⁵⁵ ʃu²² tswe⁵⁵
Tiungpaí khâ bō=ēhiau shiu tsuê
 long.ago CMPV NEG=ABI think RSLT
- p^{ha55}e³³ taj²²tsi⁵¹
phaí =ē taitsi
 bad=MOD thing

‘Before, (they) were less able to think of (or treat) things as malicious things (as a result of thinking).’
 (CLIN-19-124:25443)

- (263) law³³e³³ lan³⁵ t^{hak22} t^{hak22} tswe⁵⁵ san²²san²² ʔoʔ²²tɕ³⁵ a⁵¹
Laū=ē lán...thak thak tsuê sangsang ohtúng a
 old=MOD person read read RSLT same.same school PRT
 ‘The old person/people are responsible for reading (for some time) that resulted in the name being read as (the name of) sort of the same school’
 (CLIN-19-124:25738)

3.5.8 Negation

Lánnang-uè has a set of five semantically distinct negative markers at the verb phrase level that are nuanced as to the type of negation. They are all derived from Hokkien. The first marker is *bè* [be⁵¹], which is used to negate an expectation or presupposition expressed in the verb phrase. *Bè* also marks perfect aspect and is placed before verb phrases. For instance, it is placed before the verb phrase *thaktshêh* ‘study’ in (264) to mark perfect aspect and negate the presupposition that *dí* ‘2.SG’ has already studied.

- (264) di⁵⁵ be³³ t^{hak22}ts^{he?}
Dí bē thaktshêh.
 2.SG NEG.PF study
 ‘You have not studied (contrary to my expectation of you having studied).’
 (CLIN-19-116:19717)

- (265) tan³³si³³ hwaj³⁵ sin³³ laj²²e⁵¹ ma³³si³³
Tānsī *huai* *sīn* *lai=e* *māsī*
 but DEM.PL new come=MOD also
- be³³ ʔoʔ²² pa⁵¹
bē *oh* *pà.*
 NEG.PF learn yet

‘But those newcomers also have not learned yet (contrary to the expectation that the newcomers have learned).’
 (CLIN-19-68:21185)

The second marker is *m* [m³³] or *ūm* [ʔəm³³], whose two forms are used interchangeably to negate copular verbs, verbs with the perfect marker *pât*, and the verb *tsa-iâ* ‘know’, as exemplified below. The syllabic variant is used more frequently.

- (266) m³³ tsa³³ʔja⁵⁵ kuŋ²² si²² tsi²² ge²² laŋ³⁵ ʔoɿ²²
m *tsa-iâ* *kung* *si* *tsi* *ge* *lang* *or...*
 NEG know if COP one CLS person or
 ‘do not know if it is just one person or...’
 (CFH-001)

- (267) si³³ 'ma²²ŋa²² 'fil²²tʃaj⁵¹ ʔəm³³ si³³ 'ma²²ŋa²²
Sī *mga* *Fil-Chi* *ūm* *sī* *mga*
 COP PL Filipino-Chinese NEG COP PL
- la³⁵naŋ³⁵ na²² tjam⁵⁵ 'mejn²²land⁵⁵
lānnāng *na* *tiām* *Mainland*
 Lannang MOD from Mainland
 ‘It is the Filipino-Chinese, not the Lannangs from the Mainland.’
 (PC0095-CLIN19)

- (268) gwa⁵⁵ ʔəm³³ pat⁵⁵ k^hi⁵¹tjo^ʔ¹¹
Guā *ūm* *pât* *khì-tioh.*
 1.SG NEG PF go-PFV
 ‘I have not successfully experienced going (there).’
 (CLIN-19-55:9417)

The third marker is the negation particle *māng* /maŋ³³/, used interchangeably with the particles *methāng* [m³³ t^haŋ³³] / *ūmethāng* [ʔəm³³ t^haŋ³³], from which it arguably derived. Apart from encoding negation, the particle also encodes prohibition (see section 3.5.1). The negated verb phrase is interpreted as a warning.

(269) di⁵⁵ maŋ³³ kjo⁵⁵ ʔi⁵⁵ lo²² kʰi⁵¹ lo¹¹ a³⁵
Dì māng kiô î loh khi lo á...
 2.SG NEG.PROH call 3.SG fall DIR PFV PRT
 ‘Don’t tell him to come down, okay?’
 (PROT-16-NA:37888)

(270) di⁵⁵ m³³tʰaŋ³³ koŋ⁵⁵
Dì mthāng kōng.
 2.SG NEG.PROH say
 ‘Don’t talk.’
 (PROT-16-NA:38180)

The fourth marker is *miên* [mjɛn⁵⁵], a negation particle that encodes necessity. The negated verb phrase *ka î kōng* ‘tell them (for their benefit)’ in the example below for example means ‘do not need to tell them (for their benefit).’

(271) hi²² ge⁵⁵ ʔa²²si²² mjɛn³⁵ ka²² ʔi⁵⁵ koŋ⁵⁵
Hi gé, āsī mién ka î kōng.
 DEM CLS also NEG.NEC BEN 3.SG say
 ‘That, you also don’t need to tell them (for their benefit).’
 (CLIN-19-44:7860)

In all other contexts of negation (e.g., before most verbs, modals of want/desire), the negative marker *bó* [bo³⁵] is used.³⁶

(272) ‘pe²²ro²² hi⁵⁵ ge³⁵ bo²² səŋ⁵⁵ hwaj³⁵
Pero hí gé bo sūng huái
 but that CLS NEG count ART.DEF.PL

 ‘ma²²ŋa²² sin³³kjaw³⁵ tse²²tsun⁵¹ dik³³ laj⁵¹ la¹¹
mga sīnkiaú tsetsùn dik laì la.
 PL new.immigrants now enter DIR PRT
 ‘But that doesn’t count the new immigrants who have entered.’
 (PC0005-CLIN18.eaf)

(273) di⁵⁵ e³³ bo²² bɛ^{ʔ55} koŋ³⁵ hwa²²na²²ʔwe⁵¹
Dì ē bo bēh kóng Huana-uè.
 2.SG POS NEG DES speak Filipino
 ‘You will no longer want to speak Filipino.’
 (PC0096-CLIN19)

³⁶ This marker is also used before adjectives and nouns (see Section 3.4).

I summarize the negative markers of the variety in Table 11.

Table 11. Negative markers in Lánnang-uè

Marker	Context of negation	Source	Distribution
<i>bè</i>	expectation or presupposition, perfect	Hokkien	__ VP
<i>bó</i>	general; any context that is not listed in this table	Hokkien	__ VP
<i>m</i> <i>ūm</i>	copula, experiential perfect <i>pât</i> , some verbs (e.g., <i>tsa-ia</i> ‘know’)	Hokkien	__ VP
<i>māng</i> <i>mthāng</i> <i>ūmthāng</i>	imperative	Hokkien	__ VP
<i>miên</i>	necessity	Hokkien	__ VP

Lánnang-uè does not allow negative concord.

3.5.9 Adverbial modification

Verbs and verb phrases in Lánnang-uè can be modified by adverbs. Adverbs that modify a verb are placed preverbally. These adverbs may be derived from adjectives, but not necessarily. If they are not derived from adjectives (e.g., *nevèr* ‘never’, *suīsī* ‘immediately’), the construction for modifying a verb is Adv + Vb. The adverb is followed by the verb:

- (274) swi³³si³³ tu⁵⁵tjo⁵¹
suīsī *tū-tiòh*
 immediate meet-PFV
 ‘immediately (successfully) met’
 (CLIN-19-114:19167)

If the adverb is derived from an adjective, the general strategy is to use the Adj~Adj + =*ē* [e³³] + Vb construction. The adjective is reduplicated, after which it is succeeded by the modifier clitic =*ē* and the verb that is being modified, as in the following examples:

- (275) ʔi⁵⁵ hi²²ja²² hi²²ja⁵⁵e³³ 'beg⁵⁵ fo²² tsi³⁵
Ī *hiya* *hiyâ=ē* *bég* *for* *tsí*
 3.SG shy shy=MOD beg for money
 ‘They shyly begged for money.’
 (elicitation, PC0068)

- (276) ban²² ban²²e³³ si⁵⁵
ban ban= \bar{e} s \bar{i}
 slow slow=MOD die
 ‘slowly die’
 (CLIN-19-68:24002)

If the adverb is preceded by another adverb that modifies it, the adjective is most often not reduplicated. For example, in (277), instead of using *yá kinkîn*= \bar{e} ‘very quickly’, the speaker uses *yá kîn*= \bar{e} ‘very quickly’ to modify the verb *khíóh-tiòh* ‘to learn’.

- (277) di⁵⁵ ja²² kin⁵⁵e³³ k^hi⁵⁵ ʔoʔ²²tjoʔ⁵¹
Dì ya kîn= \bar{e} *khì oh-tiòh*...
 2.SG very fast=MOD go learn-PFV
 ‘You quickly (successfully) learned ...’
 (CLIN-19-116:19822)

If the adverb is derived from an adjective sourced from English, the adjective is suffixed by *-ly* /li⁵¹/ instead of being reduplicated (i.e., Adj + *-ly* + \bar{e} + Vb):

- (278) ‘slow²²li⁵¹e³³ bo²²k^hi⁵¹
slow-ly= \bar{e} *bokhì*
 slow-ADV=MOD disappear
 ‘slowly disappear’
 (CLIN-18-70:10297)

Often, speakers do not use the modifier clitic = \bar{e} . There are no semantic or pragmatic differences, to my knowledge, between adverbial modification constructions with the clitic = \bar{e} and those without. Some examples of adverbial modification without = \bar{e} are as follows:

- (279) kin²² kin⁵⁵ k^hi⁵¹
kin kîn *khì*
 fast fast go
 ‘quickly go’
 (PROT-16-NA:38012)

- (280) ban²² ban²² ban²² ban²² k^hwa⁵¹
ban ban ban ban *khuà* ...
 slow slow slow slow see
 ‘sort of slowly see’
 (PROT-16-NA:37934)

- (281) ʔi⁵⁵ big²²laʔ⁵⁵ big²²laʔ⁵⁵ ts^hu⁵⁵lai⁵¹
*Í **bigla** **bigla** tshùlai.*
 3.SG sudden sudden appear
 ‘They suddenly appear.’
 (elicitation, PC0068)

To emphasize the manner of doing the action, the multifunctional prefix *pà-* /pa⁵¹/ is used before the adverb. Before verbs and adjectives, the prefix functions as an inchoative marker (see 3.5.2). However, before an adverb, it encodes emphasis of manner and/or intentionality. Examples illustrating the semantic differences between utterances with and without this prefix are shown below:

- (282) hi⁵⁵ge²² ja²²jaʔ³⁵ ban²²ban²² le⁵⁵ tsjaʔ³⁵
Hige yayá banban lê tsiáh.
 ART.DEF.SG domestic.helper slowly PROG eat
 ‘Our domestic helper is slowly eating.’
 (elicitation, PC0068)

- (283) hi⁵⁵ge²² ja²²jaʔ³⁵ pa ban²²ban⁵¹ le⁵⁵ tsjaʔ³⁵
*Hige yayá **pa-banbàn** lê tsiáh.*
 ART.DEF.SG domestic.helper EMPH-slowly PROG eat
 ‘Our domestic helper is eating in a manner that is (intentionally) slow.’
 (elicitation, PC0068)

Adverbs that modify verb phrases are placed in either the pre- or post-verbal position, depending on the adverb, as illustrated in (284) to (291). The most commonly used phrase-level adverbs, with their origins and distributions, are summarized in Table 12.

- (284) gun⁵⁵ tak²²maj⁵⁵ k^hi⁵⁵ pu²²ʔok²² bitʃ⁵⁵
*Gún **takmaí** khí Puok beách.*
 1.PL.EXC always go Puok beach
 ‘We always go to Puok beach.’
 (PC0019-CLIN18)

- (285) ʔin⁵⁵ ma³³si³³ le⁵⁵ k^haj²²si⁵⁵ le⁵⁵ sa²²taŋ³⁵
*Ín **māsī** lê khaisī lê satáng.*
 3.PL also PROG begin PROG assimilate
 ‘They are also beginning to do the action of assimilating.’
 (PC0071-CLIN18)

(286) di⁵⁵ ke⁵⁵tsaj⁵⁵ ba²²'lik⁵⁵ ko⁵⁵ ti³³ 'past⁵⁵
Dì kêtsai balik kô tī³³ past.
 2.SG again return PRT in past
 'You return to the past again.'
 (PC0083-CLIN19)

(287) ʔin⁵⁵ le⁵⁵ ʔo^{ʔ35} 'din⁵⁵ la⁵¹
În lê óh dîn là.
 3.PL PROG learn also PRT
 'They are also learning.'
 (PC0072-CLIN18)

(288) so³⁵ ʔin⁵⁵ nəŋ³³ e³⁵ k^haj²² si⁵¹ ts^he⁵¹ 'ʔu²²'lit⁵⁵
Só, îŋ nūŋ é khaisî tshè ulít.
 So 3.PL two CLS start find again
 'So the two of them started to search again.'
 (PC0012-FRST19)

(289) 'pe²²ro²² di⁵⁵ ʔu³³ k^hi⁵⁵ sja²²mi^{ʔ35} ho²²'te^l⁵¹
Pero dí ū khî siammih hotèl
 but 2.SG PF go what hotel

 tswe⁵⁵kaŋ³³ 'mu²²na⁵⁵ di²²'ba³⁵
tsuékāng munâ dibá?
 work first PRT

'But you have gone to some hotel to work temporarily first, right?
 (CE-001)

(290) le⁵⁵ ts^he²² hi²² ge⁵⁵ 'fɿog⁵⁵ pa²²din⁵⁵
Lê tshè hi gê frog padîn.
 PROG find DEM CLS frog still
 '(They) are still looking for the frog.'
 (PC0012-FRST19)

(291) dan⁵⁵e³³ kul²²tʃur⁵¹ ja²²ko⁵⁵ be²² 'blend⁵⁵ pa⁵¹
Dân=ē cultùre yakô be blênd pà.
 1.PL.INC=GEN culture still NEG blend yet
 'Our culture has still not yet blended (with other cultures).'
 (PC0091-CLIN19)

(292) dan⁵⁵ ts^hɿŋ²² dɿŋ²² laŋ³⁵ ʔi⁵⁵kjeŋ³³ bo²² le⁵⁵
Dân *tshingding* *láng* *ikiēng* *bo* *lê*
 1.PL.INC youth person already NEG PROG

ʔjeŋ³³ hok²²kjeŋ⁵⁵ʔwe⁵¹ lo¹¹
iēng *Hokkiēnuè* *lò.*
 use Hokkien PFV
 ‘We, the youth, are not using Hokkien already.’
 (CLIN-19-68:22263)

Table 12. List of the most common verb phrase adverbs, with origin and distribution

Adverb	IPA	Origin	Gloss	Distribution
<i>atâ</i>	[ʔa ²² taʔ ⁵⁵]	Tagalog	‘maybe’	VP __
<i>āsī</i>	[ʔa ³³ si ³³]	Hokkien	‘also’	__ VP
<i>baka</i>	[ba ²² ka ²²]	Tagalog	‘maybe’	__ VP
<i>dîn</i>	[din ⁵⁵]	Tagalog	‘also’	VP __
<i>ikiēng</i>	[ʔi ⁵⁵ kjeŋ ³³]	Hokkien	‘already’	__ VP
<i>kêtsai</i>	[ke ⁵⁵ tsaj ⁵¹]	Hokkien	‘again’	__ VP
<i>lâng</i>	[laŋ ⁵⁵]	Tagalog	‘only’	VP __
<i>(m)āsī</i>	[ʔa ³³ si ³³ [ma ³³ si ³³]	Hokkien	‘also’	__ VP
<i>mēdyò</i>	[me ³³ dʒo ⁵¹]	Tagalog	‘somewhat’	__ VP
<i>munâ</i>	[mu ²² na ⁵⁵]	Tagalog	‘first’	VP __
<i>nanamân</i>	[na ²² na ²² man ⁵⁵]	Tagalog	‘again’	VP __
<i>pà</i>	[pa ⁵¹]	Tagalog	‘yet/even/ still’	VP __
<i>padîn/parîn</i>	[pa ²² din ⁵⁵ [pa ²² rin ⁵⁵]	Tagalog	‘still’	VP __
<i>takmaî takpaî</i>	[tak ²² maj ⁵⁵ [tak ²² paj ⁵⁵]	Hokkien	‘always’	__ VP
<i>sigurò</i>	[si ²² gu ²² .io ⁵¹]	Tagalog	‘possibly’	__ VP VP __
<i>talagà</i>	[ta ²² la ²² ga ⁵¹]	Tagalog	‘really’	__ VP VP __
<i>tsiū</i>	[tsju ³³]	Hokkien	‘at once’	__ VP
<i>ulît</i>	[ʔu ²² lit ⁵]	Tagalog	‘again’	VP __
<i>yakô</i>	[ja ²² ko ⁵⁵]	Hokkien	‘still’	__ VP

Hokkien-derived VP-level adverbs are always pre-verbal, while those derived from Tagalog are pre-verbal, post-verbal, or both. For Tagalog-derived adverbs that can be placed before or after the verb phrase (e.g., *talagà* ‘really’), there do not seem to be linguistic factors that condition the position of the adverb, to the extent of my knowledge and analysis.

There is variation in the adverbs for ‘still’, ‘only’, ‘really’, and ‘again’ with regard to source language. The Tagalog-origin variant is more commonly used for ‘still’, ‘only’, and ‘really’, whereas the Hokkien-sourced variant is more popular for ‘again’ (Table 13).

Table 13. Frequency distribution of variants for selected adverbs

	Hokkien-derived	Tagalog-derived
still	1 (3.45%)	28 (96.55%)
also	66 (44%)	84 (56%)
really	46 (14.4%)	273 (85.6%)
again	45 (63.38%)	26 (36.62%)

3.5.10 The copula

The copula in Lánnang-uè is the Hokkien-origin *sī* /si³³/. To link the subject noun phrase to a noun phrase complement, *sī* is placed between the two constituents.

- (293) tsi⁵⁵ge³⁵ si²² tsi²²ge²² va²²ˈja²²je²²ti⁵¹ ba³⁵
Tsígé *sī* *tsige* *variety* *bá?*
 this COP ART.INDEF.SG variety Q
 ‘Is this a variety?’
 (PC0019-CLIN18)

It is also used to link the subject noun phrase and adjective phrase complements.

- (294) ʔjəŋ²²bun³⁵ si³³ ʃuɹ⁵¹
lungbún *sī* *sùre...*
 English COP sure
 ‘English is certain...’
 (CLIN-18-4:1521)

Although copula constructions are used in Lánnang-uè, copula-less constructions are preferred over them.

(295) ʔi⁵⁵ ja³⁵ hwa²²hi⁵⁵ na⁵⁵
 Í yá huahî ná.
 3.SG very happy PRT
 ‘They (singular) are very happy, you know.’
 (PC0002-FRST18)

(296) hi⁵⁵ge²² kaw⁵⁵ ja³⁵ hwa²²hi⁵⁵
 Hîge kaû yá huahî.
 ART dog very happy
 ‘The dog is very happy.’
 (PC0071-FRST18)

3.6 Clause-level constituent order

The canonical order in Lánnang-uè is SVO. The subject is placed before the verb:

Intransitive sentences

(297) hige⁵⁵ pa²²la²²ka⁵⁵ t^hjaw⁵¹ ts^hu¹¹laj¹¹
 Hîge palakâ thiaù tshulai.
 ART.DEF.SG frog jump out
 ‘The frog jumped out.’
 (FRST-20-175:37042)

(298) hige⁵⁵ pa²²la²²ka⁵⁵ kap⁵⁵ hi⁵⁵ge²² gi²²na⁵⁵ to⁵¹k^hi²² lo²²a²²
 Hîge palakâ kâp hîge ginâ tòkhi lo a.
 ART frog and ART kid return.DIR PFVPRT
 ‘The frog and the kid returned.’
 (FRST-20-175:37070)

(299) hi⁵⁵ge²² bit²²p^haŋsju⁵¹ lak⁵⁵ lo⁵¹laj¹¹
 Hîge bitphangsiù lâk loh lai.
 ART.DEF.SG beehive fall down
 ‘The beehive fell down.’
 (FRST-20-164:36890)

Transitive sentences

(300) ‘juŋ²² dʒa³³pan⁵¹ kaj²²bam⁵¹ ‘tʃhaj³³na⁵¹
 Yung Jāpàn kay-bòm̃b Chīnà...
 when Japan CAUS-bomb China
 ‘When Japan bombed China... (literally, when Japan caused China to be bombed by Japan)...’
 (C-002)

- (301) ʔi⁵⁵ ʔu³³ dja²²tjo²² tsi²²ge²² 'pa²²la⁵⁵ka²²⁵⁵
 Í ū diah-tioh tsige palakâ.
 3.SG PF catch-PFV ART frog
 'They have successfully caught a frog.'
 (FRST-18-2:855)
- (302) ʔi⁵⁵ k^hwa⁵¹tjo²² hwaj³⁵ pa²²de²² swe⁵⁵ swe⁵⁵e³³
 Í khuà-tioh huaí pade suê suê=ē .
 3.SG see-PFV DEM.PL other small small=MOD
 'pa²²la²²ka²² ʔ⁵⁵ pa⁵¹
 palakâ pà
 frog even
 'They even (successfully) saw those other sort of small frogs.'
 (FRST-20-175:37069)

If the verb is ditransitive, the subject begins the sentence, followed by the verb, the indirect object, and then the direct object.

- (303) dan⁵⁵e³³ ʃjoŋ²²tsoŋ⁵⁵ ʔu³³ ho²² ʔin⁵⁵ tsĩ²² la⁵¹
 Dân=ē Shiongtsóng ū ho ín tsínn là.
 1.PL.INC=GEN Shiongtsong PFgive 3.PL money PRT
 'Our Shiongtsong has given them money.'
 (CLIN-19-142:41322)

The default clause-level constituent order changes in constructions with topicalized constituents, passive voice (unintentional causatives), and benefactive marking. It also changes in intentional causative and dative constructions, and occasionally, in a certain type of *wh*-question.

To topicalize a constituent, the constituent that “functions as [the] topic” (Crystal 2008:488) is placed in the sentence-initial position. In (304), the object noun phrase of the utterance *hí ge mīngkiā* ‘that thing’ is topicalized, whereas in (305), the verb phrase (i.e., *suê hí ge mīngkiā* ‘washed that thing’) is. In these cases of topicalization, the default word order is changed from SVO to OSV and VOS, respectively.

- (304) hi⁵⁵ ge²² miŋ³³kja³³ gwa⁵⁵ swe²² djaw⁵⁵ lo⁵¹
 Hí ge mīngkiā guâ sue diaû lò.
 DEM CLS thing 1.SG wash PFV PFV
 'That thing, I washed.'
 (elicitation, PC0068)

- (305) swe²² djaw⁵⁵ hi⁵⁵ ge²² mŋ³³kja³³ lo³¹ gwa⁵⁵
Sue diaû hî ge mŋkiā lò guâ.
 wash PFV DEM CLS thing PFV 1.SG
 ‘Washed that thing, I did.’
 (elicitation, PC0068)

Passivizing the sentence also changes the default constituent order. If the user wants to shift the focus from the agent to the patient, such that the grammatical subject is the patient and the object is the agent, the S + unintentional causative or ‘passive’ (henceforth, passive) *hō* + O + V construction is used (section 3.5.4).

- (306) ʔi⁵⁵ ho³³ gwa⁵⁵ ts^hjo⁵¹ lo¹¹
Í hō guâ tshìò lò.
 3.SG CAUS 1.SG laugh PFV
 ‘They (singular) were laughed at by me.’
 (elicitation, PC0068)

Below are some examples of passive constructions found in natural speech.

- (307) gwa⁵⁵ tjeŋ²²maj⁵⁵ ko⁵⁵ ho³³ laŋ³³ ts^hjo⁵¹
Guâ tiungmaî kô hō lāng tshìò.
 1.SG before PRT CAUS person laugh
 ‘I was laughed at by people before.’
 (PC0004-CLIN18.eaf)

- (308) hi⁵⁵ge²² kaw⁵⁵ ho³³ hwaj³⁵e³³ ‘bis⁵⁵ ko⁵⁵ le⁵⁵ tsip⁵⁵
Hîge kaû hō huai=ē beês kô lê tsíp.
 ART.DEF.SG dog CAUS DEM.PL=MOD bees PRT PROG chase
 ‘The dog is being chased by those bees.’
 (PC0071-FRST18)

- (309) ke⁵⁵ ho³³ laŋ²² phi³³ la³¹
Kê hō lāng phī là.
 all CAUS person bully PRT
 ‘Everyone was bullied (by people).’
 (PC0097-CLIN19)

Using ‘intentional causative’ (section 3.5.4 henceforth, causative) constructions also changes the default constituent order. The default order (i.e., SVO) is maintained for the causative construction that only involves the causative *kāy-* prefix (S + *kāy-* + V + O, see Section 3.6) – one strategy used to indicate that the subject caused the object to *experience* the action or process (in the verb) caused by the subject themselves.

- (310) ʔi⁵⁵ kaj³³ ts^hjo⁵¹ gwa¹¹ lo¹¹
 Í *kāy-* tshìò gua lo.
 3.SG CAUS- laugh 1.SG PFV
 ‘They caused me to be laughed at (by them).’
 (elicitation, PC0068)

The other strategy – the use of Hokkien-sourced causative marker *tsiōng* – also has the same function,³⁷ but changes the order of the constituents to S + *tsiōng* + O + V:

- (311) ʔend²² sjoŋ²²te⁵¹ ʔe³³ tsjoŋ³³ di⁵⁵e³³ 'lajf⁵⁵ o³⁵ 'bles⁵⁵ lo⁵¹
 And Siongtè ē *tsiōng* dī=ē lifê ó blêss lò.
 and God POS CAUS 2.SG=GEN life PRT bless PFV
 ‘And God will cause your life to be blessed by him.’
 (PROT-16-NA:37972)

Both causative strategies³⁸ are used interchangeably, although the *kāy-* causative construction is used more frequently than the other construction. It is possible to use the *kay-* prefix before the verb in the *tsiōng* construction without any changes in meaning.

- (312) ʔi⁵⁵ bo³³ tsjoŋ³³ hi⁵⁵ge²²bu²²ti⁵⁵kan³³ kaj³³ k^huj³³ le³¹
 Í bō *tsiōng* hige butîkân *kāy-* khuī le.
 3.SG NEG CAUS ART.DEF.SG bottle CAUS-open PF
 ‘They did not cause the bottle to be in a state of being (caused to be) opened.’
 (FRST-19-130:29262)

The default SVO order also changes for another type of causative construction where the subject caused the object to *do* (not experience) an action expressed by the verb. Here, the SVO order changes to SOV in an S + causative *hō* + O + V construction.

- (313) ʔi⁵⁵ ho³³ gwa⁵⁵ ts^hjo⁵¹
 Í *hō* guâ tshìò.
 3.SG CAUS 1.SG laugh
 ‘They (singular) caused me to laugh.’

³⁷ The speakers I sampled never used the S + *kay-* + V + O and S + *tsiōng* + O + V causative strategies to indicate that the agent subject caused the patient object to do (instead of experience) the verb (e.g., ‘They caused me to laugh.’). They instead use the S + causative *hō* + O + V construction (described in this section).

³⁸ Unlike the default constructions (e.g., S + V + O), the two causative strategies emphasize that the subject is the cause of the action/process expressed in the sentence. For example, the causative sentence *Guâ tsiōng erasèr tansák* ‘I caused the eraser to be discarded’ is different from the default-construction sentence *Guâ tansák erasèr* ‘I discarded the eraser’ in that the first sentence emphasizes that the subject *guâ* is the cause of the object *erasèr* being discarded. There does not seem to be this reading in the second sentence.

(elicitation, PC0068)

Some examples involving constructions with causative *hō* are as follows:

- (314) ʔin³³wi³³tsi³³ ge⁵⁵ tsa⁵⁵ ho²² lan²² tsaj³³ʔja⁵⁵
Īnwī tsī gē tsā ho lang tsai-iā
because DEM CLS then CAUS person know

koŋ³⁵ di⁵⁵ si²² lan³⁵naŋ³⁵
kóng dí si Lánnáng.
speak 2.SG COP Lannang
'Because this is what will cause people to know that you are Lannang.'
(PC0072-CLIN18)
- (315) di⁵⁵ be^{ʔ55} ho³³ ʔi⁵⁵ tswan²²tswan²² koŋ³⁵ hwi³³di³³pin³³ʔwe⁵¹
Dī bēh hō î tsuantsuan kóng Huīdīpīn-uè?
2.SG will CAUS 3.SG whole speak Filipino
'Will you make them (or cause them to) speak in Filipino entirely?'
(PC0006-CLIN18.eaf)
- (316) 'ka²²si⁵⁵ gwa⁵⁵ ʔu²² ho³³ ʔin⁵⁵ ts^hjam³³ 'wej²²v.ɿ⁵¹ e⁵⁵
Kasī guā u hō ín tshiam waivèr eh.
because 1.SG PF CAUS 3.PL sign waiver PRT
'Because I have made them (or have caused them to) sign the waiter.'
(PC0005-CLIN18)

To express that a patient-role object is a benefactor of the agent-role subject's action (Section 3.5.5), the default order changes from SVO to the benefactive construction S + *kā/kāng* + O + V. Hokkien-sourced *kā* is used more frequently than *kāng*, which is also sourced from Hokkien. To my knowledge, there are no factors that condition the use of *kā* over *kāng*. They are used interchangeably.

- (317) ʔu²² lan³⁵ be^{ʔ55} ka³³ di⁵⁵ kaj³³sjak⁵⁵ la⁵¹
Ū láng bēh kā dí kaīsiák là.
have person will BEN 2.SG explain PRT
'There will be someone who will explain it to you for your benefit.'
(E-004)
- (318) di⁵⁵ ka³³ gwa⁵⁵ mŋ²² k^hwa⁵⁵ maj⁵¹ tse¹¹
Dī kā guā mng khuā mai tse.
2.SG BEN 1.SG ask look try first
'Try asking them first for my benefit.'
(F-002)

(319) hi⁵⁵ge²² ma²²ne²²dʒɿ⁵¹ ka³³ di⁵⁵ khuj³³ tsi²²ge²² ʔa²²kawnt⁵⁵
Hige manager kã dí khuī tsige account.
 ART.DEF manager BEN 2.SG open ART.INDEF account
 ‘The manager opened an account for your benefit.’
 (CE-001)

(320) bas²²ta⁵¹ gwa⁵⁵ ʔe³³ kaŋ³³ ʔin⁵⁵ ka⁵¹ la¹¹
Bastà guâ ē kãng ín kà la.
 PRT 2.SG will BEN 3.PL teach PRT
 ‘No matter what, I will teach them for their benefit.’
 (PC0019-CLIN18)

(321) di⁵⁵ kaŋ³³ gwa⁵⁵ ʔʃhek⁵⁵ʃa²²mi^{ʔ55} a⁵¹
Dí kãng guâ chêck siammîh à?
 2.SG BEN 1.SG check what PRT
 ‘What did you check for my benefit?’
 (PC0019-CLIN18.eaf)

The default order sometimes changes from SVO to OSV in *wh*-questions where the *wh*-phrase has an object role (i.e., *Shangá/Siangá* ‘who’, *Shammîh/Siammîh* ‘what’) (see Section 3.7.2 and Chapter 6).

(322) sjam²²mi^{ʔ55} toŋ²²but³⁵ a⁵¹ ʔi⁵⁵ kaj³³ p^ha^{ʔ55}si⁵¹
Siammîh tongbút à í kãy- phâhsì?
 what animal PRT 3.SG CAUS kill
 ‘What animal did they cause to be killed by them?’
 (elicitation, PC0005)

The default ditransitive order (i.e., SVO_{indirect}O_{direct}) changes for particular dative constructions. In dative utterances involving verbs of communication (e.g., *kông* ‘speak’, *kaisiâk* ‘explain’), the default order changes to S + *kâp* + O_{indirect} + V_{communication} + O_{direct}. The multifunctional Hokkien-derived morpheme *kâp* /kap⁵⁵/ functions as a dative marker when not used as a conjunction denoting ‘and’ (Section 3.9).

(323) gwa⁵⁵ hwaj³⁵ gi²²na⁵⁵ bε^{ʔ55} kap⁵⁵ gwa⁵⁵
Guâ huai ginâ... bêh kâp guâ
 1.SG DEM.PL kid DES DAT 1.SG

koŋ⁵⁵ hwa²²na²²ʔwe⁵¹ la¹¹
kông Huana-uè la.
 speak Filipino PRT

‘My kids want to speak Filipino to me.’
 (CLIN-19-117:19979)

In utterances involving verbs of giving (e.g., *sâng* ‘deliver’, *donâte* ‘donate’), the default ditransitive order changes to S + V_{giving} + O_{direct} + dative *hō* + O_{indirect}. The multifunctional morpheme *hō* functions as a lexical verb ‘give’ or a causative marker when not used as a dative marker.

(324) ʔi⁵⁵ t^hjaʔ⁵⁵ tsi²² paʔ⁵⁵ k^ho³³ ho²² ʔin⁵¹
 Î ... thiâh tsi pâh khō **ho** ìn.
 3.SG tear.and.give one hundreds CLS DAT 3.PL
 ‘They tore (and gave a check of) one hundred units (of money) to them.’
 (CLIN-19-117:20102)

I summarize the clause-level constituent orders below.

Table 14. Clause-level constituent orders

Construction	Order	Reading
Default	S + V	The subject verbs.
	S + V + O _{direct}	The subject verbs the direct object.
	S + V + O _{indirect} + O _{direct}	The subject verbs the indirect object the direct object.
Benefactive	S + <i>kā/kāng</i> + O + V	The subject verbs for the benefit of the object.
Intentional causative	S + <i>kāy-</i> + V + O	The subject intentionally caused the object to be verbed.
	S + <i>tsiōng</i> + O + (<i>kāy -</i>) + V	The subject intentionally caused the object to be verbed.
	S + <i>hō</i> + O + V	The subject intentionally caused the object to do the verb.
Unintentional causative or 'Passive'	S + <i>hō</i> + O + V	The existence of the subject unintentionally caused the object to verb the subject. OR The subject was verbed by the object.
Dative	S + dative <i>kâp</i> + O _{indirect} + V _{communication} + O _{direct}	The subject verbs (of communication) the direct object to the indirect object.
	S + V _{giving} + O _{direct} + dative <i>hō</i> + O _{indirect}	The subject verbs (of giving) the direct object to the indirect object.
Some <i>wh</i> -questions with an object-role <i>wh</i> -phrase	O + S + V	What/who did the subject verb?
Topicalization	O + S + V	The object is what the subject verbed.
	V + O + S	Verbed the object is what the subject did.

3.7 Questions

3.7.1 Yes-no questions

There are two primary ways of constructing *yes-no* questions: (1) using the clause + negative particle strategy and (2) attaching the Tagalog-derived morpheme *bá* clause-finally.

In the first strategy, a Hokkien-derived negative particle (i.e., *bò*, *bè*) is placed after the clause. This strategy converts the declarative utterance into an interrogative. It is only used (1) if the clause is an affirmative (positive) declarative statement and (2) if the main verb in the clause is not a copula:

- (325) pa²²pa³⁵ ʔu²² tsja²² bo⁵¹
Papá *u* *tsiah* *bò?*
 father PF eat NEG
 ‘Has father eaten?’
 (elicitation, PC0068)

The particle *bè* /be⁵¹/ is used if the action in the question is implied to have not been completed yet but will be at a certain point of time. If no such implication exists, the particle *bò* is used instead.

- (326) di⁵⁵ tsja²² lo⁵¹ be¹¹
Dì *tsiah* *lò* *bè?*
 2.SG come PFV NEG
 ‘Have you eaten?’
 (CLIN-19-134:31600)

- (327) ʔu³³ pat⁵⁵ tʰja³³tjo²² bo²¹
Ū *pât* *thiã-tioh* *bò?*
 PF PF listen-PFV NEG
 ‘Have you experienced successfully listening to this?’
 (CLIN-19-68:32591)

Only the two negative particles, *bó* and *bè*, can be used in the first strategy. The rest (see Section 3.5.8) cannot be. For instance, the *ũm/m* negative particles are never placed in sentence-final position to create a *yes-no* interrogative, unlike Hokkien (Chappell 2019:218).

- (328) pa²²pa³⁵ u²² tsja²² m²¹
 **Papá* *u* *tsiáh* *m?*
 father PF eat NEG
 ‘Has father eaten?’
 (judgment, PC0068)

Unlike the first strategy, the second strategy – the use of *bá* /ba³⁵/ sentence-finally – is a general strategy. It is used both in negative and affirmative contexts.

- (329) di⁵⁵ ʔu³³ ʃu³³kʰi⁵¹ ba³⁵
Dì *ũ* *shūkhì* *bá?*
 2.SG PF anger Q?
 ‘Have you gotten mad?’
 (PC0005-CLIN18)

- (330) si³³ 'tem²²po²²' ia²²ii⁵¹ la η⁵⁵ ba³⁵
Sī temporary lāng bá?
 COP temporary only Q
 'Is it just temporary?'
 (PC0005-CLIN18)
- (331) la³⁵naŋ²² bo²² ʔaj⁵⁵ kjaw³³ hwa²²na⁵⁵ sa²²taŋ³⁵ ba³⁵
Lánnáng... bo aī kiaū huanâ ... satáng bá?
 Lannang NEG love with locals same Q
 'Do Lannangs not love being similar with the locals?'
 (CLIN-19-141:351190)

Unlike the first strategy, the second is also used when converting utterances with the copula into *yes-no* questions.

- (332) hi⁵⁵ ge³⁵ si²² tsi²²ge²² 'tɿend⁵⁵ ba³⁵
Hī gé si tsiġe trênd bá?
 DEM CLS COP ART.INDEF.SG trend Q
 'Is that a trend?'
 (CLIN-19-141:35581)

The *bá* interrogative construction is used more commonly than to the *bó/bè* constructions.

3.7.2 *Wh-questions*

Forming *wh*-questions in Lánnang-uè involves careful attention to the position of the *wh*-phrase,³⁹ which is conditioned by the type of the *wh*-phrase (see Chapter 6). The *wh*-phrase of a *wh*-question is placed sentence-initially if the phrase is either a *why*-phrase (i.e., *kàna* 'why', *uisiammih* 'why')⁴⁰ or if it is the subject of the sentence. Otherwise, the *wh*-phrase is either placed in the adverbial position (before the verb phrase) or in the verb phrase complement position (after the verb). The position depends on the type of *wh*-phrase: the *wh*-phrase is placed in the verb phrase complement position if the *wh*-phrase is the object of the utterance (e.g.,

³⁹ I refer to *wh*-words and phrases headed by *wh*-words as *wh*-phrases in line with contemporary work (Atkinson et al. 2016).

⁴⁰ Hokkien-derived *kàna* [ka⁵¹na¹¹] 'why' and *uisiammih* [ʔuj²²sja²²mi^{ʔ55}] 'why' are used interchangeably. To my knowledge, there are no linguistic factors that condition the use of one variant over the other. The first variant is more ubiquitously used.

siammîh ‘what’, *shammîh* ‘what’,⁴¹ *siangá*, ‘who’, *shangá* ‘who’,⁴² *siammîh mîngkiā* ‘what thing’). All other *wh*-phrases – adjunct *wh*-phrases including *tisí* ‘when’, *tolóh* ‘where’, *tsai-iùnn* ‘how’, *tsiúwá* ‘how’⁴³ – are placed in the adverbial position.

For example, in (333), the Hokkien-derived *wh*-phrase *kàna* /ka⁵¹na¹¹/ ‘why’ is placed sentence-initially because it is a *why*-phrase.

- (333) ka⁵⁵na⁵⁵ dîn⁵⁵ beʔ⁵⁵ tsaw⁵⁵ a⁵¹
Kânâ *dîn* *bêh* *tsaú* *a?*
 why 2.PL want run PRT
 ‘Why do you want to run?’
 (elicitation, PC0071)

The Hokkien-derived phrases *siangá* /sja⁵¹ŋa³⁵/ ‘who’ and *siammîh insect* /sja⁵¹mɪʔ⁵⁵ʔin⁵¹sɛkt⁵⁵/ ‘what insect’ in the examples below function as the subject in their respective sentences so they are placed in the sentence-initial position as well.

- (334) sja²²ŋa³⁵ beʔ⁵⁵ tsaw⁵⁵
Siangá *bêh* *tsaú?*
 who will run
 ‘Who will run?’
 (elicitation, PC0007)

- (335) sja²²mɪʔ³⁵ ʔin²²sɛkt⁵⁵ le⁵⁵ ka³³
Siammîh *insêct* *lê* *kā?*
 what insect PROG bite
 ‘What insect is biting?’
 (elicitation, PC0068)

Shammîh [ʃam²²mɪʔ³⁵] ‘what’, *shangá* [ʃa²²ŋa³⁵] ‘who’, and *to tsigé* [to³⁵tsi²²ge³⁵] ‘which one’ – all Hokkien-derived – are placed in the verb complement position because they function as the objects of their respective sentences.

⁴¹ Hokkien-derived *siammîh* [sja²²mɪʔ⁵⁵] ‘what’ and *shammîh* [ʃa²²mɪʔ⁵⁵] are used interchangeably. To my knowledge, there are no linguistic factors that condition the use of one variant over the other. The first variant is more frequently used. The second variant is used more frequently by younger speakers.

⁴² Hokkien-derived *siangá* [sja²²ŋa³⁵] ‘what’ and *shangá* [ʃa²²ŋa³⁵] are used interchangeably. There are no linguistic factor, to my knowledge that condition the use of one variant over the other. The first variant is more frequently used. The second variant is used more frequently by younger speakers.

⁴³ Hokkien-derived *tsiúwá* [tsju⁵⁵wa⁵⁵] ‘how’ and *tsai-iùnn* [tsaj⁵⁵ʔjũ⁵¹] ‘how’ are used interchangeably. To my knowledge, there are no linguistic factors that condition the use of one variant over the other. The first variant is more frequently used.

- (336) ʔin⁵⁵ le⁵⁵ sɿ⁵⁵ tswe⁵⁵ ʃam²²mi^{ʔ35}
*În lê sũng tsuê **shammih?***
 3.PL PROG count RSLT what
 ‘What did they count that as?’
 (elicitation, PC0068)
- (337) le⁵⁵ mŋ²² ʃa²²ŋa³⁵
*Lê mung **shangá?***
 PROG ask who
 ‘Who are you asking?’
 (elicitation, PC0068)
- (338) di⁵⁵ bɛ^{ʔ55} ʔo^{ʔ22} to³⁵ tsi²² ge³⁵ a⁵¹
*Dî bêh oh **tó tsi gé a?***
 2.SG DES learn which one CLS PRT
 ‘Which one do you want to learn?’
 (CLIN-18-90:12778)

The *wh*-phrases in the examples below are found in adverbial position (i.e., before the verb phrase) because they do not belong to any of the classes mentioned earlier (i.e., *why*-phrases, subject or object *wh*-phrases).

- (339) ʔin⁵⁵ tsju⁵⁵wa⁵⁵ tsju⁵⁵wa⁵⁵ k^haŋ³³k^ho⁵⁵ a⁵¹
*În **tsiûwâ** ... **tsiûwâ** khāngkhô a?*
 3.PL how how labor PRT
 ‘How did they labor?’
 (elicitation, PC0068)
- (340) di⁵⁵ tsaj⁵⁵ʔjũ⁵¹ ka⁵⁵ di⁵⁵je³³ gi²²na⁵⁵?
*Dî **tsai-iunn** kâ dî=e ginâ?*
 2.SG how teach 2.SG=GEN kid
 ‘How did you teach your kid?’
 (CLIN-18-19:5561)
- (341) tsi⁵¹ si³³ ti³³si³⁵ 'pub²²lɿ⁵⁵ a⁵¹
*Tsí sī **tīsí** 'publîsh a?*
 DEM COP when publish PRT
 ‘When was this published?’
 (elicitation, PC0068)
- (342) ʔi⁵⁵ to³³lo^{ʔ35} be^{ʔ55} p^hjaŋ⁵¹ a¹¹
*Î **tōlôh** bêh phiàng a?*
 3.SG where POS explore PRT
 ‘Where will they (singular) be about to explode?’
 (elicitation, PC0071)

- (343) hwaj³⁵ laŋ³⁵ bɛŋ⁵⁵ tjam⁵⁵ to³³loŋ³⁵ 'ɿes²²kju⁵¹ a¹¹
Huai *láng* *bêh* *tiâm* *tōlôh* *rescuè* *a?*
 DEM.PL people POS PREP where rescue PRT
 ‘Where will the people be about to rescue?’
 (elicitation, PC0001)

Hokkien-derived *tsiuwâ* [tsju⁵⁵wa⁵⁵] ‘how’ is found in the adverbial position but is in the sentence-initial position because the speaker did not produce the subject.

- (344) tsju⁵⁵wa⁵⁵ k^haŋ³³k^ho⁵⁵ a⁵¹
Tsiûwâ *khāngkhô* *a?*
 how labor PRT
 ‘How did (they) labor?’
 (elicitation, PC0068)

In general, the position of the *wh*-phrase in Lánang-uè *wh*-questions is conditioned by its type and grammatical role. However, some speakers do not follow the distributional pattern described. These speakers sometimes do not front the *why*-phrases in *why*-questions.

- (345) ʔi⁵⁵ ka⁵⁵na⁵⁵ bɛŋ⁵⁵ ho²² gwa⁵⁵ 'ken²²di⁵¹ a¹¹
Í *kânâ* *bêh* *ho* *guâ* *candy* *a?*
 3.SG why want give 1.SG candy PRT
 ‘Why will they give me candy?’
 (elicitation, PC0071)

Others sometimes front *wh*-phrases even if they are not *why*-phrases or subject *wh*-phrases:

- (346) ti²²si³⁵ a⁵¹ tsi⁵⁵ge²² snejk⁵⁵ be ʔ⁵⁵ tsja ʔ²² 'ɿa²²bit⁵⁵
Tisi *à* *tsíge* *snâke* *bêh* *tsiah* *rabbít?*
 when PRT DEM snake want eat rabbit
 ‘When will this snake eat the rabbit?’
 (elicitation, PC0071)

- (347) sja²²miŋ³⁵ toŋ²²but³⁵ a⁵¹ ʔi⁵⁵ kaj³³ p^haŋ⁵⁵si⁵¹
Siammîh *tongbút* *à* *í* *kāy-* *phâhsì?*
 what animal PRT 3.SG CAUS-kill
 ‘What animal did they cause to be killed by them?’
 (elicitation, PC0005)

- (348) to³⁵ tsi²² ge²² k^hju³³ di⁵⁵ le⁵⁵ koŋ⁵⁵
Tó tsi ge khiū dí lê kông?
 which one CLS accent 2.SG PROG say
 ‘Which accent are you referring to?’
 (CLIN-19-68:24062)

Chapter 6 provides a deeper analysis of variation in *wh*-question formation.

3.8 Prepositions

In this section, I describe all the prepositions in Lánang-uè, based on corpus and elicitation data. These prepositions, in general, express temporal, spatial, and conceptual relations between constituents.

To express that an entity is accompanied by another entity, the English-derived preposition of accompaniment *with* ‘with’ [wid⁵⁵] is used.

- (349) bo²² beʔ⁵⁵ miks⁵⁵ wid²² hjo²²ŋe⁵⁵ la⁵¹
bo bêh mîx... with hiongê là
 NEG DES mix PREP DEM PRT
 ‘don’t want to mix [something] with that.’
 (CLIN-19-141:35282)

- (350) di⁵⁵ beʔ⁵⁵ sajd⁵⁵ wid²² to³⁵ tsi²² ge³⁵ a⁵¹
Dí bêh síde with tó tsi gé a?
 2.SG DES side PREP which one CLS PRT
 ‘Which one do you want to side with?’
 (CLIN-19-68:35729)

To express the meaning ‘of’, particularly a correlative, associative, meronymic (part of whole), or possessive relationship between an entity and another one, English-derived *ôf* ‘of’ /ʔof⁵⁵/ is generally used.

(351) 'miks²²tʃu.ɿ⁵¹ ʔof²² fi²²li²²'pi²²no⁵¹' ʔend²² 'tʃaj²²nis⁵¹e³³
mixture of Filipino and Chinese=ē
 mixture PREP Filipino and Chinese=MOD

ʔu²²'ga²²liʔ⁵⁵
ugali
 behavior

‘behavior that is a mixture of Filipino and Chinese’
 (CLIN-18-20:6697)

Occasionally, some speakers use the multifunctional Tagalog-derived linker morpheme *ng* /naŋ⁵⁵/. When between reduplicated verbs, it functions as an iterative aspect marker (Section 3.5.2). However, before a noun phrase, it functions as a preposition meaning ‘of’. It is interchangeable with *ōf*, to the best of my knowledge.

(352) hwaj³⁵ pa²²de³⁵ kap⁵⁵ ʔi⁵⁵e³³ bu²²'ʔoŋ²² 'fa²²mi²²li⁵¹
Huai padé kâp î=ē buong family
 DEM.PL other and 1.SG=GEN whole family

naŋ²² pa²²la²²'kaʔ⁵⁵ le⁵⁵ k^hwa⁵¹ laŋ⁵⁵
ng palakâ lê khuà lāng.
 LNK frog PROG see only

‘Those other [frogs] and their whole family of frogs are only watching.’
 (FRST-20-175:37071)

(353) hi⁵⁵ge²² 'wej⁵¹ naŋ²² ko²²'mju²²ni²²'kej²²ʃon⁵¹
hîge wày ng communication.
 ART way LNK communication
 ‘the way of communication’
 (CLIN-19-9:3889)

(354) 'so²² k^ha⁵⁵nan⁵⁵ 'paɪ⁵⁵ naŋ²² gwa⁵⁵e³³ ʔaj²²'den²²ti²²ti⁵¹
 ... *so khânân pârt ng guâ=ē identity*
 so like part LNK 1.SG=GEN identity
 ‘...so it’s like part of my identity.’
 (CLIN-19-9:3810)

(355) di⁵⁵ ʔu³³ hi⁵⁵ge²² 'best⁵⁵ naŋ²² hi⁵⁵ge²² 'bowt²²'woɪld⁵⁵
Dî ū hîge best ng hîge both wôrld-s.
 2.SG have ART best LNK ART both world-PL
 ‘You have the best of the two worlds.’
 (PC0009-CLIN19)

- (356) hi⁵⁵ge²² t^haw³⁵ naŋ²² ʔa ʔ⁵⁵ bo²² lo⁵¹ ba³⁵
Híge *tháú ng áh bó lò bá?*
 ART head LNK duck NEG PFV Q
 ‘Is the head of the duck gone?’
 (PROT-16-NA:37769)

Based on corpus data, there are no linguistic factors (e.g., source language of the head noun and the phrase complement) that condition the use of one preposition of conceptual relationship over the other.

Lánnang-uè has prepositions that indicate a general area or region in a physical, conceptual, or temporal space where an entity can be found, all sourced from Hokkien. Prepositions of specific location are discussed later in this section. These general locative prepositions are *tī* ‘at’ /ti³³/, *ân* ‘at’ /ʔan⁵⁵/,⁴⁴ *tiâm* ‘at’ /tjam⁵¹/, *tuâ* ‘at’ /twa⁵⁵/, and *kūn* ‘near’ /kun³³/. The first four are used to express a general area anchored to a point in space. By default, *tī* is used.

- (357) gwa⁵⁵e³³ t^hin³³ ti²² taj²²djok³⁵
guâ=ē *chīn* *tī* *Taidiók.*
 1.SG=GEN relative PREP China
 ‘my relative(s) in China.’
 (CLIN-19-126:26678)

- (358) tsi⁵⁵ ge³⁵ si³³ ti²² to²²lo^{ʔ35}
Tsí gé sī tī tolóh?
 DEM CLS COP PREP where
 ‘(At) where is this?’
 (PROT-16-NA:37396)

- (359) gi²²na⁵⁵ ti³³ bin²²tjeŋ⁵⁵ ʔe²² ts^he⁵¹
Ginâ *tī* *bintiêng* *e* *tshè.*
 kid PREP face.up POS find
 ‘Kid(s) will find [something] on top.’
 (FRST-20-159:36699)

The other three general locative prepositions are only used if the main verb of the utterance is a non-copula verb. It is, for example, ungrammatical to say *Guâ sī ân/tuâ/tiâm tshûlāi* ‘I am at home’ because *ân/tuâ/tiâm* requires a non-copula verb. Changing the main verb to a non-copula

⁴⁴ This morpheme functions as a locative preposition as well as a preposition of range/path. See discussion below.

verb, as in *Guâ khùn ân/tuâ/tiâm tshûlâi* ‘I am sleeping at home’, makes the use of these prepositions grammatical. These three prepositions are used interchangeably. I have not found these usage patterns in my corpus data.

- (360) $hi^{55}ge^{22}gi^{22}na^{55}$ $k^{hwa^{51}}tjo^{22}$ $tsi^{22}ge^{22}$ $khaŋ^{33}$
Hige ginâ *khuà-tioh* *tsige* *khāng*
 ART kid see-PF ART hole
- $ʔan^{55}$ $tsi^{55}ge$ $ʔui^{51}$ $laj^{22}bin^{51}$
ân *tsîge* *tree* *laibìn.*
 PREP DEM tree inside

‘The kid saw a hole at the insides of this tree/ inside this tree.’
 (FRST-20-160:36779)

- (361) $ʔi^{55}$ twa^{55} hja^{35} $t^{hak^{35}}$ ho^{35}
Í *tuâ* *hiá* *thák* *hó...*
 3.SG PREP DEM read PRT
 ‘They read there, right?’
 (CLIN-18-4:1591)

- (362) gua^{55} $tjam^{55}$ tsi^{22} ta^{255} $twa^{22}han^{51}$
Guâ *tiâm* *tsi* *tâh* *tuahàn.*
 1.SG PREP DEM CLS grow.up
 ‘I grew up here.’
 (CLIN-19-44:7617)

The locative preposition *kūn* /*kun*³³/ is used to indicate general proximity.

- (363) $ʔEs^{22}kol^{22}ta^{51}$ ti^{33} kun^{33} $so^{22}leɪ^{55}$ hi^{55} ta^{255} ko^{55}
Escoltà *tī...* *kūn* *Solêr* *hî* *tâh* *kó...*
 Escolta PREP PREP Soler DEM CLS PRT
 ‘So Escolta is at... near Soler.’
 (CLIN-19-55:9363)

To indicate that an entity is oriented towards another entity, the variety uses the prepositions of orientation *hiòng* /*hjoŋ*⁵¹/ ‘towards’ and *tui* /*tuj*⁵¹/ ‘towards’. The first is only used in contexts that involve physical orientation (i.e., facing the direction of another entity).

- (364) ʔi⁵⁵ k^haj²²si³⁵ hɔŋ⁵⁵ gwa²²khaw⁵⁵ tsaw⁵⁵ k^hi⁵¹ lo¹¹
Í khaisí hiông guakhaú tsaú khi lo.
 3.SG start PREP outside run DIR PFV
 ‘They started facing outside and ran away.’
 (FRST-19-135:32502)

The second preposition of orientation, which can also be used as a preposition of path or range (see description below), is used for both physical and non-physical orientation (e.g., ‘in relation to’, ‘as regards’).

- (365) le⁵⁵ tuj⁵⁵ di⁵⁵ koŋ²²ʔwe⁵¹
lê tuî dî konguè
 PROG PREP 2.SG speak
 ‘speaking facing you.’
 (PROT-16-NA:38101)

- (366) pa²²raŋ²² ʔi⁵⁵ tuj⁵⁵ hi⁵⁵ ge²² 'bi²²hajv⁵⁵
Parang í tuî hî ge beehivê
 like 3.SG PREP DEM CLS beehive

 'fa²²si²²'nej²²ted⁵⁵ ko⁵⁵
fascinatêd kô.
 fascinated PRT

‘Like, they feel fascinated towards that beehive.’
 (FRST-20-158:36545)

Lánng-uè also has three Hokkien-derived prepositions of range or path: *tuì* ‘from’ /tuj⁵¹/, *ân* ‘from’ /ʔan⁵⁵/, and *kau* ‘to’ /kaw⁵¹/ . These mark the point in a path-like space or time or a range at which an entity begins or ends. When *tuì* and *ân* are not used as prepositions of orientation and general location, respectively, they function as these prepositions. Specifically, the morphemes *tuì* and *ân* are used to mark the starting point (or a point of origin). They are used interchangeably.

- (367) hi⁵⁵ge²² fɿog⁵⁵ tuj⁵⁵ hi⁵⁵ge²² dʒaɿ⁵¹ le⁵⁵ t^haw²²tsaw⁵⁵ ts^hut⁵⁵khi⁵¹
Hîge frôg tuî hîge jàr lê thautsaú tshutkhi.
 ART frog PREP ART jar PROG steal.run out.DIR
 ‘The frog secretly ran away from the jar.’
 (FRST-20-158:36511)

- (368) gun^{55} tuj^{55} $gun^{22}han^{35}$ $tu^{55}tjo^{22}$ i^{21}
Gún *tuí* *gunháng* *tú-tioh* *ì.*
 1.PL.EXC PREP bank meet-PFV 3.SG
 ‘We successfully met them from the bank.’
 (PROT-16-NA:37152)
- (369) η^{55} tuj^{55} tsi^{33} pa^{255} $ki^{22}lo^{51}$ san^{35} kaw^{55} pwe^{255} $tsap^{35}$ na^{22} lan^{55}
Í *tuí* *tsi* *pâh* *kilò* *sán* *kaú* *puêh* *tsáp* *nalâng.*
 3.SG PREP one hundreds kilo thin PREP eight tens PRT
 ‘They weighed eighty kilograms from a hundred kilograms.’
 (elicitation, PC0068)
- (370) $hwaj^{35}$ $gin^{22}na^{55}$ η^{55} $tjo^{22}kok^{55}$ laj^{35}
Huai *ginnâ* *ân* *Tiongkôk* *lái.*
 DEM.PL kid PREP China come
 ‘Those kids came from China.’
 (CLIN-19-138:33656)

Kau / kaw^{51} / is used to mark the end point.

- (371) kaw^{55} $ma^{22}kos^{55}=e^{33}$ $si^{33}tsun^{51}$
 ... *kaú* *Marcôs=ē* *sītsùn.*
 ... until Marcos=GEN time
 ‘... until Marcos’ time.’
 (PC0096-CLIN19)

The prepositions *tui* / tuj^{51} / ‘from’ and *ân* / η^{55} / ‘from’ may not be used if the speaker expresses a temporal frame, as in the example below.

- (372) gun^{55} $ts^{h}ut^{55}si^{51}e^{33}$ si^{35} kaw^{55} gwa^{55} $t^{h}ak^{22}ts^{h}e^{255}e^{33}$ si^{35}
gún *tshùtsì=ē* *sí* *kaú* *guâ* *thaktshêh=ē* *sí*
 1.PL.EXC born=MOD time PREP 1.SG study=MOD time
 ‘from the time we were born to the time I studied’
 (CLIN-19-140:34326)

To express temporal relations, the multifunctional morphemes *af̀t̀er* ‘after’ / $\eta^{55}t_{\eta}^{51}$ / and *bef̀ore* ‘before’ / $bi^{51}fo:r^{51}$ / are used. When placed before a subordinate clause, these English-derived morphemes function as conjunctions (Section 3.9). Before a noun phrase that is not in a subordinate clause, they are prepositions of temporal relations. *Af̀t̀er* refers to a time later than the time or event mentioned while *bef̀ore* refers to a time earlier.

- (373) la³⁵naŋ²² ʔoʔ²²tuŋ³⁵ ʔaf²²tɿ²² lantɕ¹¹ ham⁵⁵bun³⁵
Lánnang *ohtúng* **after** *lúnch* *Hâmbún.*
 Lannang school after lunch Chinese
 ‘Lannang schools have Chinese classes after lunch.’
 (PROT-16-NA:37235)
- (374) hwaj³⁵ loŋ²²tsoŋ⁵¹ si³³ bi²² foɿ²²hi⁵⁵ge²² 'tejk²²ʔow²²vɿ⁵¹a¹¹
Huái *longtsòng* *sī* **before** *hîge* *takeovèr* *a.*
 DEM.PL all COP before ART.DEF.SG takeover PRT
 ‘All of those were [done] before the takeover.’
 (CLIN-19-141:35440)
- (375) bi²²foɿ²² 'najn²²tin²² 'siks²²ti²² 'tu⁵¹ ʔu³³
before *nineteen* *sixty* *twò* *ū*
 before nineteen sixty two have
 ‘had before 1962’
 (CLIN-19-141:35260)

Lánnang-uè has nineteen prepositions of specific location, all derived from English. These prepositions, unlike the prepositions of general location discussed earlier, mark a point or an area in space that is more specific than locations marked with general locative prepositions (e.g., *tī* ‘at’). Prepositions of specific location indicate the position of an entity relative to the position of another entity. I enumerate them below, after which I give examples for some of them. Brief descriptions for each one of them can be found in Table 15, which summarizes all the prepositions discussed in this section.

1.	<i>abové</i>	‘above’	/ʔa ⁵¹ ‘bav ⁵⁵ /
2.	<i>acrôss</i>	‘across’	/ʔa ⁵¹ ‘kios ⁵⁵ /
3.	<i>against</i>	‘against’	/ʔa ⁵¹ ‘genst ⁵⁵ /
4.	<i>alòng</i>	‘along’	/ʔa ⁵¹ ‘lon ⁵¹ /
5.	<i>amòng</i>	‘among’	/ʔa ⁵¹ ‘mon ⁵¹ /
6.	<i>aroùnd</i>	‘around’	/ʔa ⁵¹ ‘.iawnd ⁵⁵ /
7.	<i>befòre</i>	‘before’	/bi ⁵¹ ‘foi ⁵¹ /
8.	<i>behînd</i>	‘behind’	/bi ⁵¹ ‘hajnd ⁵⁵ /
9.	<i>belòw</i>	‘below’	/bi ⁵¹ ‘low ⁵¹ /
10.	<i>beside</i>	‘beside’	/bi ⁵¹ ‘sajd ⁵⁵ /
11.	<i>betweèn</i>	‘between’	/bi ⁵¹ ‘twin ⁵¹ /
12.	<i>ìn</i>	‘in’	/‘ʔin ⁵¹ /
13.	<i>inside</i>	‘inside’	/ʔin ⁵¹ ‘sajd ⁵⁵ /
14.	<i>òn</i>	‘on’	/‘ʔon ⁵¹ /
15.	<i>outsìde</i>	‘outside’	/ʔawt ⁵⁵ ‘sajd ⁵⁵ /
16.	<i>ovèr</i>	‘over’	/ʔow ⁵¹ ‘vɿ ⁵¹ /
17.	<i>througħ</i>	‘through’	/‘θru ⁵¹ /
18.	<i>undèr</i>	‘under’	/‘ʔan ⁵¹ dɿ ⁵⁵ /
19.	<i>withìn</i>	‘within’	/wid ⁵⁵ ‘ʔin ⁵¹ /

(376) gwa⁵⁵ tjoʔ²² ‘pas⁵⁵ ʔa²²’kios²² da²² ‘.iowd⁵⁵ ko⁵⁵
Guâ tìoh pâss across the road kô.
 1.SG NEC pass PREP ART.DEF.SG road PRT
 ‘I should pass across the road.’
 (CLIN-19-68:40519)

(377) di⁵⁵ k^hwa⁵¹ ʔu²² di²²fɛ²².iɛns⁵⁵ bo⁵¹ bi²²‘twin²² ‘fɿl²² tʃ^haj⁵¹
Dî khuà u difference bò between FilChì
 2.SG look have difference NEG between Filipino.Chinese
 kap⁵⁵ ‘tʃ^haj²²nis²² fɿ²²li²²‘pi²²no⁵¹
kâp Chinese Filipino
 and Chinese Filipino
 ‘Do you think these is a difference between Filipino-Chinese and Chinese Filipino?’
 (PC0096-CLIN19)

(378) gun⁵⁵ tswaj³⁵ kǰǎ⁵⁵ k^ha⁵⁵ tsjo²² he⁵¹ ‘ʔan²²dɿ⁵¹ ʔi¹¹
Gûn tsuai kiânn khâ tsio hè undèr i.
 1.PL DEM.PL son CMPV less age under 3.SG
 ‘We sons are younger compared to him.’
 (PC0019-CLIN18)

Prepositions are all placed before a noun phrase (specifically, the head position of preposition phrases). Prepositional phrases that are headed by Hokkien-derived prepositions can only be placed before or after the verb phrase (i.e., adverbial position), interchangeably. There are no

factors, to my knowledge, that condition the position of the prepositional phrase relative to the verb phrase. Prepositional phrases headed by English-derived prepositions can additionally be placed after a head noun, as a complement of a noun phrase. The only exception is *ôf* ‘of’, which can only be in the complement position of a noun phrase, and not in an adverbial position (e.g., before the verb phrase). The only Tagalog-derived preposition, *ng* ‘of’, also only occupies the complement position.

Prepositions of general location, orientation, and range/path come from Hokkien whereas prepositions of accompaniment, specific location, temporal relations, and conceptual relations come from English. However, there is some variation in the selection of specific prepositions from a specific language. For example, infrequently, speakers use the Tagalog-derived *sà* /sa⁵¹/ ‘at’ or English-derived *ât* /ʔat⁵⁵/ ‘at’ instead of the Hokkien-derived prepositions of general location (e.g., *tī* ‘at’). Speakers also sometimes use English-derived *tò* /tu⁵¹/ ‘towards’ instead of the Hokkien-derived preposition of orientation *hiông* ‘towards’ or *tuī* ‘towards’. I provide an in-depth analysis of prepositions and the variation in their use in Chapter 5.

(379) ta²²pos⁵⁵ hwaj³⁵ ʔu²²we³⁵le⁵⁵ ʔak²²tiv⁵⁵ sa²² tsu²²he²²so⁵⁵
Tapôs *huai* *uwé* *lê* *active* *sa* *Tsuhesô*.
 and DEM.PL some PROG active at Gospel.Center
 ‘And then there are those who are being active [are active] at Christian Gospel Center.’
 (PC0126-CLIN19)

(380) ʔi⁵⁵ tsa⁵⁵ ʔwejv⁵⁵gud²²ʔbaj⁵¹ tu²² hwaj³⁵ ts^han²² ka²²ba⁵⁵
Î *tsâ* *wave* *goodbyè* *to* *huai* *tshankabâ*.
 3.SG then wave goodbye towards DEM.PL frog
 ‘They then waved goodbye towards those frogs.’
 (PC0071-FRST18)

I summarize the prepositions discussed in this section alphabetically along with a description, the preposition type, source language, and distribution of the prepositional phrase in the table below.

Table 15. The prepositions of Lánnang-uè

Preposition	Gloss	Description	Type	Source	Distribution of PP
<i>abovê</i>	‘above’	indicates that the entity is in a higher point relative to another entity; both entities do not touch each other	specific location	English	__ VP VP __ N __

<i>acrôss</i>	‘across’	indicates that the entity is one side to the other of something with clear limits, such as an area of land, a road, or a river	specific location	English	$\overline{\text{VP}}$ VP __ N__
<i>afèr</i>	‘after’	denotes ‘at or during a time later than the time or event mentioned’	temporal relation	English	$\overline{\text{VP}}$ VP __ N
<i>agaînst</i>	‘against’	indicates that the entity is next to and touching or being supported by something	specific location	English	$\overline{\text{VP}}$ VP __ N__
<i>alông</i>	‘along’	indicates that the entity is in a line next to something long	specific location	English	$\overline{\text{VP}}$ VP __ N
<i>amông</i>	‘among’	indicates that the entity is in the middle of or surrounded by other entities	specific location	English	$\overline{\text{VP}}$ VP __ N__
<i>ân</i>	‘at’	indicates the general area or region in a physical, conceptual, or temporal space the entity is in	general location	Hokkien	$\overline{\text{VP}}$ VP __
	‘from’	marks the point in space or time at which something starts	range/path	Hokkien	$\overline{\text{VP}}$ VP __ N
<i>arouînd</i>	‘around’	indicates that the entity is in a position or direction surrounding another entity	specific location	English	$\overline{\text{VP}}$ VP __ N__
<i>befôre</i>	‘before’	denotes ‘in front of’	specific location	English	$\overline{\text{VP}}$ VP __ N
	‘before’	denotes ‘at or during a time earlier than the time or event mentioned’	temporal relation	English	$\overline{\text{VP}}$ VP __ N__
<i>behînd</i>	‘behind’	denotes ‘at the back of’	specific location	English	$\overline{\text{VP}}$ VP __ N
<i>belôw</i>	‘below’	indicates that the entity is in a lower point relative to another entity; both entities do not touch each other	specific location	English	$\overline{\text{VP}}$ VP __ N__
<i>besîde</i>	‘beside’	denotes 'at the side of, next to'	specific location	English	$\overline{\text{VP}}$ VP __ N

<i>between</i>	‘between’	indicates that the entity is in or into the space that separates two other entities	specific location	English	$\overline{\text{VP}}$ VP __ N __
<i>hiòng</i>	‘towards’	indicates that the entity is oriented towards or facing the direction of another entity	orientation	Hokkien	__ VP
<i>ìn</i>	‘in’	indicates that the entity is inside a container, place, or area, or surrounded or closed off by something	specific location	English	$\overline{\text{VP}}$ VP __ N __
<i>inside</i>	‘inside’	indicates that the entity is inside a container, place, or area, or surrounded or closed off by something	specific location	English	$\overline{\text{VP}}$ VP __ N __
<i>kau</i>	‘to’	marks the point in space or time at which something ends	range/path	Hokkien	__ VP VP __
<i>kūn</i>	‘near’	marks a general area that is a short distance away from the entity	general location	Hokkien	__ VP VP __
<i>ng</i>	‘of’	expresses the correlative, meronymic, associative, or possessive relationship between the entity and another entity; occasionally used	‘of’	Tagalog	N __
<i>ôf</i>	‘of’	expresses the correlative, meronymic, associative, or possessive relationship between the entity and another entity.	‘of’	English	N __
<i>òn</i>	‘on’	indicates that the entity is in a position above something else and touching it	specific location	English	$\overline{\text{VP}}$ VP __ N __
<i>outside</i>	‘outside’	indicates that the entity is not in a particular enclosure, but near it	specific location	English	__ VP VP __ N
<i>ovèr</i>	‘over’	indicates that the entity is above or higher than another entity	specific location	English	$\overline{\text{VP}}$ VP __ N
<i>through</i>	‘through’	indicates that the entity occupies a space of	specific location	English	__ VP VP __ NP

		an entity, from one end to the other			
<i>tī</i>	‘at’	indicates the general area or region in a physical, conceptual, or temporal space the entity is in.	general location	Hokkien	$\overline{\text{VP}}$ VP _
<i>tiâm</i>	‘at’	indicates the general area or region in a physical, conceptual, or temporal space the entity is in.	general location	Hokkien	$\overline{\text{VP}}$ VP _
<i>tuâ</i>	‘at’	indicates the general area or region in a physical, conceptual, or temporal space the entity is in.	general location	Hokkien	$\overline{\text{VP}}$ VP _
<i>tui</i>	‘towards’	denotes ‘orienting with’, ‘with regard to’, or ‘concerning’	orientation	Hokkien	$\overline{\text{VP}}$ VP _
	‘from’	marks the point in space or time at which something starts	range/path	Hokkien	$\overline{\text{VP}}$ VP _
<i>undèr</i>	‘under’	indicates that the entity is beneath, lower than, or managed by another entity	specific location	English	$\overline{\text{VP}}$ VP _ N _
<i>wìth</i>	‘with’	denotes ‘in the company of’	accompaniment	English	$\overline{\text{VP}}$ VP _ N
<i>wìthin</i>	‘within’	indicates that the entity is inside the limits of another entity	specific location	English	$\overline{\text{VP}}$ VP _ N

3.9 Conjunctions

In this section, I describe all the conjunctions in Lánnang-uè, based on corpus and elicitation data. I first describe coordinating conjunctions (i.e., adversative, cumulative, disjunctive). Then, I describe subordinating ones (i.e., adverbializers, complementizers, and relativizers).

3.9.1 Coordinating conjunctions

To indicate contrast or opposition between two clauses that have equal (syntactic) importance, the Tagalog-derived adversative conjunction *però* /'pe⁵¹ro⁵¹/ ‘but’ or *kasò* /'ka⁵¹so⁵¹/ ‘but it is the case that’ is placed between them. The first is used by default while the second is used if the speaker wants to emphasize that the utterance following it is a statement of fact.

- (381) kaw⁵⁵hwe⁵¹ ʔu³³ le⁵⁵ ko ŋ⁵⁵ la⁵¹ 'pe²²ro²² gwa⁵⁵ si²²
Kaûhuè, *ũ* *lê* *kông* *là* *pero* *guâ* *si*
church PF PROG speak PRT but 1.SG COP
ʔan⁵⁵ pe²²bu⁵⁵ ʔoʔ²²tjoʔ⁵¹
ân *pebû* *oh-tiòh.*
from parent learn-PFV
‘At church, I have been speaking it but I learned it from my parent.’
(PC0005-CLIN18)

- (382) ʔi⁵⁵ ti³³ti³³ ts^{he}⁵¹ 'ka²²so⁵⁵ tsi²²ge²²
În *tītī* *tshè* ... *kasô* *tsige*
3.PL ITER find but.it.is.the.case ART.INDEF.SG

'ʔawl⁵¹ts^{hut}⁵⁵ laj⁵¹
òwl *tshût* *lài.*
owl emerge DIR
‘They kept on looking ... but it is the case that an owl came out.’
(FRST-19-94:14086)

To cumulatively connect words, phrases, or clauses, the cumulative⁴⁵ conjunctions *kâp* /kap⁵⁵/ ‘and’, *kiaû* /kjaw³³/ ‘and’, *tsakà* /tʃa⁵¹'ka⁵¹/ ‘and also’, and *tapôs* /ta⁵¹'pos⁵⁵/ ‘and then’ are used. The first two, derived from Hokkien, conjoin two or more constituents (i.e., NP, VP, PP, AdjP, S), placing equal emphasis on them. They are used interchangeably. I have not found any linguistic factors that condition the use of one over the other.

- (383) ka²²'si⁵⁵ di⁵⁵ kap⁵⁵ ʔaŋ³⁵koŋ³³ 're²²kon²²sajl⁵¹lo¹¹
Kasí *dí* *kâp* *ángkōng* *reconcile* *lo.*
because 2.SG and grandfather reconcile PRT
‘Because you and your grandfather have reconciled.’
(CFH-001)

- (384) ʔu³³ tsi³³ge³³la²²me⁵¹ kjaw²² tsi²²ge²²kaw⁵⁵ le⁵⁵ k^hwa⁵⁵ tsi²²ge²²'fiag⁵⁵
Ū *tsīgē* *lamè* *kiau* *tsige* *kaû* *lê* *khuâ* *tsige* *frôg.*
have ART boy and a dog PROG look ART frog
‘There is a boy and a dog looking at a frog.’
(PC0012-FRST19)

The third, derived from Tagalog, connects two constituents and additionally emphasizes order – that the constituent before it is sequenced prior to the constituent after it.

⁴⁵ Increasing or increased in quantity, degree, or force by successive additions.

- (385) di⁵⁵ kaj²² 'miks⁵⁵ 'solt⁵⁵ 'pe²²pɿ⁵¹ ta²²pos⁵⁵ ma²²ŋa²² 'hɿb⁵⁵
Dî kay- mîx sâlt, peppèr, tapôs mga herb.
 2.SG CAUS mix salt pepper and.then PL herb
 ‘Mix the salt, pepper, and then the herbs.’
 (elicitation, PC0068)

- (386) ʔi⁵⁵ 'fej⁵⁵ hi⁵⁵ge²² tʃ^hju⁵¹ ta²²'pos⁵⁵ 'fej⁵⁵ djaw⁵⁵
Î shâke hîge chiù tapôs shâke diaû,
 3.SG shake ART.DEF.SG tree and.then shake finish

 hi⁵⁵ge²² 'bi²²hajv⁵⁵ fol²² 'ʔawt⁵⁵
hîge beehîve fall ôt.
 ART.DEF.SG beehive fall out

‘They began shaking the tree, and then, finished shaking the tree, after which the beehive fell out.’
 (FRST-20-172:36999)

- (387) gwa⁵⁵ beʔ⁵⁵ ka²² di⁵⁵ koŋ⁵⁵ tsi²² ge²² 'sen²²tens⁵⁵ ko⁵⁵
Guâ bêh ka dî kông tsi ge sentênce kô
 1.SG POS BEN 2.SG say one CLS sentence PRT

 ta²²'pos⁵⁵ di⁵⁵ beʔ⁵⁵ mən²² bun²²twe³⁵
tapôs dî bêh mung buntué.
 and.then 2.SG POS ask question

‘I will tell you one sentence for your benefit, and then you will ask me a question.’
 (FRST-18-2:907)

The fourth one, also derived from Tagalog, is used to connect two constituents, and indicates that the preceding constituent is being emphasized over the constituent following it.

- (388) hjoŋ²²e⁵⁵ ʔin²²na⁵⁵ tʃa²²ka²² 'rej²²ʃal⁵¹ na²²
 ... *hiongê innâ tsaka raciàl na*
 ... DEM.PL kid and.also racial MOD

 'slɿ⁵¹ dɿn⁵¹ ko⁵⁵
slur dîn kô.
 slur also PRT

‘... those kids, (and racial slurs) too’
 (CLIN-19-16:4820)

To present two or more constituents as alternatives, the variety uses the disjunctive conjunction *âsī /ʔa⁵⁵si³³* / ‘or’, derived from Hokkien.

(389)	ka ²² si ⁵⁵	ja ³⁵	tswe ³³	nan ⁵⁵ e ³³	pe ²² bu ⁵⁵	ʔa ⁵⁵ si ³³	ʔaŋ ³⁵ koŋ ³³
	<i>Kasí</i>	<i>yá</i>	<i>tsuē</i>	<i>nân=ē</i>	<i>pebú</i>	<i>âsī</i>	<i>ángkōng</i>
	because	very	many	1.PL=GEN	parent	or	grandfather
	ʔa ³³ ma ⁵⁵	hi ⁵⁵	tsun ⁵¹	laj ¹¹	huj ³³ di ³³ pin ³³		
	<i>āmá</i>	<i>hī</i>	<i>tsùn</i>	<i>lai</i>	<i>Huīlīpīn.</i>		
	grandmother	DEM.SG	time	come	Philippines		

‘Because a lot of our parents or grandfathers/grandmothers came to the Philippines that time.’

(PC0099-CLIN19)

3.9.2 Subordinating conjunctions

Lánnang-uè also has a variety of subordinating conjunctions. It has adverbializers, complementizers, and relativizers. The adverbializers of the variety mark clauses as “having some adverbial function” (Schachter and Shopen 2007:50), particularly the expression of condition, concession, consequence, location, manner, purpose, reason, substitution, and time (Thompson et al. 2007:243).

To introduce a conditional clause, six conjunctions, enumerated below, are used. All of them are Tagalog-derived except the last, which is English-derived.

1.	<i>nā</i>	‘if (general)’	/na ³³ /
2.	<i>(ka)pâg</i>	‘if (restricted)’	/(ka ⁵¹)’ pag ⁵⁵ /
3.	<i>pagkà</i>	‘as soon as’	/pag ⁵⁵ ka ⁵¹ /
4.	<i>hanggât</i>	‘so/as long as’	/haŋ ⁵⁵ ’ gat ⁵⁵ /
5.	<i>kahitnà</i>	‘even if’	/’ka ⁵¹ hit ⁵⁵ ’ na ⁵¹ /
6.	<i>unlêss</i>	‘unless’	/’ʔan ⁵¹ les ⁵⁵ /

The first is used generally – it can indicate a certain or uncertain condition.

(390)	na ³³	di ⁵⁵	kaj ²²	’miks ⁵⁵	hok ²² kjen ⁵¹	ʔi ⁵⁵	ʔe ³³	t ^h ja ²²	bo ³⁵	lo ⁵¹
	<i>Nā</i>	<i>dī</i>	<i>kay-</i>	<i>mīx</i>	<i>Hokkièn...</i>	<i>î</i>	<i>ē</i>	<i>thia</i>	<i>bó</i>	<i>lò...</i>
	if	2.SG	CAUS-mix	Hokkien	3.SG	POS	hear	NEG	PFV	
	‘If/When you caused Hokkien to be mixed... they will not be able to understand.’									
	(CLIN-19-109:17948)									

(391)	gwa ⁵⁵	na ³³	kh ⁱ 55	taj ²²	djok ³⁵	ko ⁵⁵
	<i>Guá</i>	<i>nā</i>	<i>khī</i>	<i>Taidiok</i>	<i>kó..</i>	
	1.SG	if	go	China	PRT	
	‘If/When I go to China...’					
	(CLIN-19-92:13596)					

The second conjunction is the Tagalog-derived *kapâg* /ka⁵¹pag⁵⁵/, which is occasionally shortened to *pâg* /pag⁵⁵. It is only used when a situation is sure to occur. Two examples containing (*ka*)*pâg* from conversations are found below.

- (392) gwa⁵⁵ k^ha⁵⁵ bwe³³ 'ʔan²²dɪ²²stand⁵⁵ hi⁵⁵ge³⁵
Guâ khâ bō =ē *understand* *hîge*
 1.SG CMPV NEG=ABI understand DEM

ka²²' pag²² ʔi⁵⁵ kaj²²' .io²²ma²²najs⁵⁵
kapag *î* *kay-romanîze.*
 if 3.SG CAUS-romanize

'I will not understand that as much when s/he romanizes [it].'
 (PC0071-CLIN18)

- (393) ' pag²² di⁵⁵ ʔaj⁵⁵ ʔjeŋ³³ hwa²²na²²ʔwe⁵¹ ʔa³³si³³
Pag dî aî iēng Huana-uè... āsî
 if 2.SG DES use Filipino also

ʔe²²tswe⁵⁵ ʔjeŋ⁵¹
etsuê *iēng.*
 PER use

'When you want to use Filipino, [you] are also permitted to use it.'
 (CLIN-19-16:4934)

The third conjunction is the Tagalog-derived *pagkà* 'as soon as' /pag⁵⁵ka⁵¹/, which introduces a condition that, when fulfilled, immediately triggers an event or process.

- (394) 'pe²²ro²² pag²²ka²² di⁵⁵ kap⁵⁵ ma²²ma³⁵ pa²²pa³⁵
Pero pagka dî kâp mamá, papá,
 but as.soon.as 2.SG DAT mom dad

kjaw³³ ʔa²²ma⁵⁵ koŋ⁵⁵
kiaū amâ kông....
 and grandmom say

'But as soon as you say [something] to mom, dad, and grandmom...'
 (CLIN-19-93:13882)

Tagalog-derived *hanggât* ‘as/so long as’ /haŋ⁵⁵gat⁵⁵/ introduces a proviso or a condition attached to an agreement.

- (395) haŋ²²gat²² ʔi⁵⁵ bo²² laj³⁵ dan⁵⁵ bo²² bɛʔ⁵⁵ tsjaʔ³⁵
Hanggat *î* *bo* *laí,* *dân* *bo* *bêh* *tsiáh.*
as.long.as 3.SG NEG come 1.PL.INC NEG POS eat
‘As long as they have not come, we won’t eat.’
(elicitation, PC0068)

The fifth conjunction is the Tagalog-derived *kahitnà* ‘even if’ /ka⁵¹hit⁵⁵na⁵¹/, which refers to clauses analogous to the concession-conditional ‘even if’ clauses in English. It can imply that the speaker is frustrated (Thompson et al. 2007:261).

- (396) ka²²hit²²na²² di⁵⁵ k^hi⁵⁵ to²²loʔ³⁵ so³⁵tsaj²²
Kahitna *dî* *khî* *tolóh* *sótsaī...*
even.if 2.SG go where place

ta²²ge²²laŋ³⁵ ja³⁵ ba²²jas⁵⁵
tage *láng* *yá* *biás...*
all person very bias
‘Wherever you go (literally, even if you go wherever) ... everyone is biased...’
(CLIN-19-14:4439)

The conjunction *unlêss* ‘unless’ /ʔan⁵¹les⁵⁵/, derived from English, introduces the only circumstances in which an event the speaker is mentioning will not take place or in which a statement the speaker is making is not true.

- (397) gwa⁵⁵ bo²² k^hwa⁵⁵tjo²² 'lo²²kal²² ko²²'mju²²ni²²ti⁵¹ kon²²'vɿs⁵⁵
Guá bo khuá-tioh local community conversê
 1.SG NEG see-PFV local community converse
- ʔin²² 'man²²da²².in⁵¹ 'ʔan²²les⁵⁵ di⁵⁵ si²² se^ʔ⁵⁵ hi⁵⁵ge²²
in Mandarin unless dí si séh híge
 in Mandarin unless 2.SG COP say DEM.SG
- lan³⁵ si²² twa⁵⁵ fan⁵⁵ton⁵⁵ laj²²e⁵¹
láng si tuá Shântông láí=ē.
 person COP at Shantong come=MOD

‘I have never seen the local community converse in Mandarin, unless you say that that person is from Shantong.’
 (CLIN-19-111:18363)

To introduce a concession, or an idea that is granted in response to the main clause, Lánnang-uè uses Tagalog-derived *kahit* ‘even though’ /ka⁵¹hit⁵⁵/ and *maski* ‘even though’ /mas⁵⁵ki⁵¹/. Both are used interchangeably without changes in meaning.

- (398) ja³⁵ bo²² saŋ³⁵ la⁵¹ hi⁵⁵ge²² nuŋ²² e²² 'kul²²tʃu⁵¹
Yá bo sáŋ là híge nung e culturè
 very NEG same PRT ART two CLS culture
- 'ka²²hit⁵⁵ di⁵⁵ se^ʔ⁵⁵ di⁵⁵ si³³ lan³⁵naŋ³⁵ ko⁵⁵
kahit dí séh dí sī Lánnáng kô.
 even.though 2.SG say 2.SG COP Lannang PRT

‘The two cultures are very dissimilar, even though you say you are Lannang.’
 (CLIN-19-9:3854)

- (399) 'mas³³ki³³ di⁵⁵ si³³ ti³³ tʃ^hi³³laŋ³³ ʔu²²we²²
Maski dí sī tī tshīlāi, uwe
 even.though 2.SG COP PREP inner.city some
- ka²²tjen³⁵ bo²² ka⁵⁵ gi²²na⁵⁵
katiéng bo kâ ginâ.
 family NEG teach kid

‘Even though you are in the inner city, some families don’t teach their kids.’
 (CLIN-19-136:32675)

To mark a clause as a result or effect, the consequence conjunction *kayâ* ‘that is why’ /ka⁵¹ja^ʔ⁵⁵/ is used. It describes the effect of the event expressed in the main clause.

- (400) di⁵⁵ koŋ⁵⁵ di⁵⁵ ʔo²²tj³⁵ si³³ ka⁵⁵ man²²da²².in⁵¹
Dî kông dî ohtúŋ sī kâ Mandarin
 2.SG say 2.SG school COP teach Mandarin
- ka²²ja²² di⁵⁵ bwe²²hja³⁵ koŋ⁵⁵ hok²²kjen⁵⁵ʔwe⁵¹
kaya dî bo=ēhiaú kông Hokkiēnuè...
 that.is.why 2.SG NEG=ABI speak Hokkien

‘You said your school teaches Mandarin, that is why you aren’t able to speak Hokkien...’
 (PC0083-CLIN19.eaf)

The morpheme *kungsaân* ‘where’ /kuŋ⁵⁵sa⁵¹ʔan⁵⁵/, derived from Tagalog, has two functions. When placed before a noun phrase expressing a location, it is a relativizer that denotes ‘where’ (see discussion on relativizers later). When it conjoins a subordinate clause and a main clause, it is the ‘where’ conjunction that marks the subordinate clause as being a place or situation relevant to the main clause.

- (401) kuŋ²²sa²²ʔan²² di⁵⁵ ʔek²²sist⁵⁵ dia²²ma⁵¹ ʔe³³ laj³⁵
Kungsaan dî exist, dramà ē laí.
 where 2.SG exist drama POS come
 ‘Where you exist, drama will come.’
 (elicitation, PC0068)

To introduce clauses referring to a manner relevant to the main clause, three Hokkien-derived conjunction variants are used – *khâlâng* ‘like’ /k^ha⁵⁵laŋ⁵⁵/, *tshĩntshiu* ‘like’ /ts^hin³⁵ts^hju³³/, and *nántshiu* ‘like/as’ /nan⁵⁵ts^hju³³/. The first is produced by some speakers as *khânân* /k^ha⁵⁵nan⁵⁵/. I have yet to find factors that condition the use of one variant over the other.

- (402) k^ha⁵⁵laŋ⁵⁵ di⁵⁵ kaŋ²² gwa⁵⁵ koŋ⁵⁵ di⁵⁵ le⁵⁵ ts^hoŋ⁵¹ ʃa¹¹
Khâlâng dî kang guâ kông dî lê tshòŋ sha...
 like 2.SG BEN 1.SG say 2.SG PROG do what
 ‘(Just) like how you would tell me what you are doing.’
 (FRST-19-132:30787)

- (403) di⁵⁵ ʃju²²tsuj⁵⁵ o³⁵ ts^hin²²ts^hju²² di⁵⁵ le⁵⁵ t^hjaw⁵⁵bu⁵⁵
Dî shiutsuí ó tshintshiu dî lê thiaúbú.
 2.SG swim PRT like 2.SG PROG dance
 ‘You swim like you are dancing.’
 (elicitation, PC0068)

- (404) nan³⁵ts^hju³³ di⁵⁵ ʔe²²tswe⁵⁵ flek²²si²²bol⁵¹ a¹¹
Nántshiū *dí* *etsué* *flexiblè* *a...*
 like 2.SG can flexible PRT
 ‘(Just) like how you are able to be flexible...’
 (CLIN-19-12:4163)

Speakers occasionally use multifunctional *parâng* /pa⁵¹ ɲaŋ⁵⁵/ ‘like’, derived from Tagalog, to interchangeably fulfill the functions of *khâlâng*, *tshīntshiū*, and *nántshiū*. However, it is rarely used as a conjunction; instead, it is used more frequently as a discourse particle or filler (Section 3.11).

The Tagalog-derived conjunction *parà* ‘so that’ /‘pa⁵¹ ra⁵¹/ indicates the purpose of the action or process in the main clause.

- (405) ʔin⁵⁵ si³³ kap⁵⁵ ʔin²² koŋ⁵⁵ hwa²²na³⁵ʔwe⁵¹ pa²²ra²²
În *sī* *kâp* *in* *kông* *Huana-uè* *para*
 3.PL COP DAT 3.PL speak Filipino so.that

 ʔin⁵⁵ ʔe²²hjaw⁵⁵
în *ehiaû....*
 3.PL ABI

‘They speak Filipino to them so that they naturally know how to [speak the language].’
 (CLIN-19-115:19560)

The equivalent of an English clause headed by ‘because’ is expressed by a *Lánnang-uè* clause introduced by Tagalog-derived *kasì* /ka⁵¹ ‘sì⁵⁵/ ‘because’ and *porkêt* /‘poɪ⁵⁵ket⁵⁵/ ‘just because’. The first subordinating conjunction is used by default to introduce a reason for the idea expressed in the main clause. A very small number of speakers occasionally use Tagalog-derived *dahîl* ‘because’ /da⁵¹hil⁵⁵/ instead of *kasì*. To my knowledge, there are no linguistic factors that condition the use of *dahîl* over *kasì*.

- (406) gwa⁵⁵ bo²²hwat⁵⁵nan³³ tsjap⁵⁵fju³³ pa³¹ ka²²'si⁵⁵ gwa⁵⁵ bo³³
Guâ bo=huâtñāng tsiâpshū pà kasi *guâ bō*
 1.SG NEG=ABI accept yet because 1.SG NEG

sa³³pat⁵⁵
sāpât ...
 acquaintance

‘I am not able to accept this yet because I have no acquaintances.’
 (PC0091-CLIN19)

- (407) 'da²²hi²² gwa⁵⁵ be^ʔ⁵⁵ khⁱ⁵⁵ taj²²djok³⁵
 ... *dahil* *guâ bêh* *khî* *Taidiók.*
 because 1.SG POS go China.

‘... because I will be going to China.’
 (CLIN-18-20:6893)

The second conjunction, *porkêt* ‘just because’ /'po⁵⁵ket⁵⁵/, is used to encode an ironic or critical attitude.

- (408) di⁵⁵ bwe³³ʔjeŋ⁵⁵ 'bas²²tus⁵⁵ gun⁵¹
Dî bo=ē-iêng bastôs gùn
 2.SG NEG=PER disrespectful 1.PL.EXC

'po²²ket²² diŋ⁵⁵ ʔu²² tsī³⁵
porket dîn u tsinn.
 just.because 2.PL have money

‘You can’t disrespect us just because you have money.’
 (elicitation, PC0068)

To signal the substitution of an expected event (the clause following it) with an unexpected one (the main clause), Lánnang-uè uses the subordinating conjunction *kaysà* /kaj⁵⁵sa⁵¹/ ‘instead of’, derived from Tagalog. It is placed before a verb phrase or a clause.

(409)	gun ⁵⁵	ti ²²	ts ^h u ⁵⁵ laj ³³	'baɪ ²² bi ²² kju ⁵¹	ma ^ʔ ⁵⁵
	<i>Gún</i>	<i>ti</i>	<i>tshùlāi</i>	<i>barbecuè</i>	<i>mâh</i>
	1.PL.EXC	PREP	house	barbecue	meat
	'kaj ²² sa ²²	kh ⁱ ⁵⁵	gwa ²² bin ³⁵	tsja ³⁵	
	<i>kaysa</i>	<i>khî</i>	<i>guabín</i>	<i>tsiá.</i>	
	instead.of	go	outside	eat	

‘We barbecued meat at home instead of going out to eat.’
(elicitation, PC0068)

The subordinating temporal conjunctions are the following:

- | | | | |
|----|----------------|------------|--|
| 1. | <i>afèr</i> | ‘after’ | /ʔaf ⁵⁵ tɪ ⁵¹ / |
| 2. | <i>núŋ</i> | ‘when’ | /nuŋ ⁵⁵ / |
| 3. | <i>bagò</i> | ‘before’ | /ba ⁵¹ go ⁵¹ / |
| 4. | <i>hanggâŋ</i> | ‘until’ | /haŋ ⁵⁵ gaŋ ⁵⁵ / |
| 5. | <i>habâŋ</i> | ‘while’ | /ha ⁵¹ baŋ ⁵⁵ / |
| 6. | <i>tuwîŋ</i> | ‘whenever’ | /tu ⁵¹ wiŋ ⁵⁵ / |

All are derived from Tagalog except the first, which is derived from English. The first simply describes the time relationship.

(410)	ʔaf ²² tɪ ²²	gun ⁵⁵	kh ^{an} ³³ ts ^h ju ⁵⁵		
	<i>After</i>	<i>gún</i>	<i>khāntshiû...</i>		
	after	1.PL.EXC	marry		
	gwa ⁵⁵	gwan ³³ tsai ³³	tsjaw ⁵⁵	ʔi ⁵⁵ e ³³	tjaw ³³ kuj ³³
	<i>guâ</i>	<i>guāntsaī</i>	<i>tsiaû</i>	<i>î=ē</i>	<i>tiaûkuī.</i>
	1.SG	still	follow	1.SG=GEN	stipulation

‘After we married... I still followed their stipulation.’
(CLIN-19-68:39942-39945)

The second, on the other hand, is a multifunctional morpheme. When placed before a clause that modifies a noun phrase related to time, it functions as the relativizer ‘when’ (see discussion on relativizers below). When used to concatenate clauses to express temporal relations between them, it functions as a conjunction that indicates that once the event or process in the subordinate clause occurs, the event in the main clause immediately happens.

- (411) nuŋ²² hi⁵⁵ge²² gi²²na⁵⁵kjaŋ³³ hi⁵⁵ tsjaŋ⁵⁵ kaw⁵⁵ k^hi⁵⁵ k^hun⁵¹ lo¹¹
Nung hige ginâ kiaũ hî tsiâh kaũ khî khùn lo,
 when ART kid and DEM CLS dog go sleep PFV
- ʔi⁵⁵ tsa⁵⁵ ʔan⁵⁵ ʔi⁵⁵e³³ jaŋ⁵¹ le⁵⁵ ts^hu⁵⁵laŋ⁵¹ a¹¹
 ... î tsâ ân î=ē jâr lê tshûlai a.
 3.SG then PREP 3.SG=GEN jar PROG emerge PRT
- ‘When the kid and that dog went to sleep, it then started coming out of the jar.’
 (FRST-19-114:19407-19408)

Bagò ‘before’ is used to temporally situate the event or process in the clause before the event.

- (412) ba²²go²² di⁵⁵ khi⁵⁵ hja³⁵
Bago dî khî hiá...
 before 2.SG go DEM
 ‘Before you go there...’
 (PC0068)

Hanggâng indicates that the event or process in the subordinate clause marks the end of the event or process in the main clause.

- (413) ʔin⁵⁵ ts^he⁵¹ haŋ²²gaŋ²² hi⁵⁵ tsjaŋ⁵⁵ kaw⁵⁵ ho²² ʔi⁵⁵e³³
În tshè... hanggang hî tsiâh kaũ ho î=ē
 3.PL search until DEM CLS dog CAUS 3.SG=GEN
- t^haw³⁵ ts^hŋ⁵⁵ dik²² k^hi⁵⁵ glas⁵⁵e²² ʔa²²kwaŋ²²jum⁵¹
thau tshung dik khî glass=e aquariùm.
 head insert enter DIR glass=MOD aquarium

‘They searched until that dog inserted his head into the glass aquarium.’
 (FRST-19-14:4721)

Habâng ‘while’ and *tuwîng* ‘whenever’ indicate that the events in both clauses share the same temporal window. *Habâng* simply indicates that the event or process in the subordinate clause is happening alongside the event or process in the main clause. *Tuwîng*, on the other hand, indicates that event or process in the subordinate clause occurs whenever the event or process in the main clause occurs.

- (414) 'ha²²baŋ²² hi⁵⁵ge²² 'bi⁵¹ le⁵⁵ 'ha²²bol⁵⁵ hi⁵⁵ge²²kaw⁵⁵ o³⁵
Habang hîge beè lê haból hîge kaû ó
 while ART bee PROG chase ART dog PRT
- yuy²² ki²²na⁵⁵ pwaʔ²²tjoʔ⁵¹
 yung kinâ puah-tiòh.
 DEM kid fall-PFV

‘While the bee was chasing the dog, that kid fell.’
 (FRST-19-95:14213)

- (415) 'tu²²wiŋ²² gwa⁵⁵ le⁵⁵ koŋ²²ʔwe⁵¹ ʔi⁵⁵ le⁵⁵ dim³³ ka²²'pe⁵⁵
Tuwing guâ lê kong-uè, î lê dîm kapê.
 whenever 1.SG PROG speak 3.SG PROG drink coffee
 ‘Whenever I am speaking, they are drinking coffee.’
 (elicitation, PC0068)

The Tagalog-derived morpheme *nà* /na⁵¹/ has multiple functions. In the domain of the noun phrase, it can function as either a complementizer or a relativizer, depending on the nature of the clause it attaches to. If the clause is necessary to complete the meaning of the noun phrase, it functions as a complementizer. It is placed before the subordinate clause.

- (416) lan⁵⁵ hwi³³li³³pin³³e²² kul²²tʃur⁵¹ ka²²si⁵⁵ ja³⁵
Lân *Huīlīpīn=ē* *cultùre* *kasí* *yá*
 1.PL.INC Filipino=GEN culture because very
- ʔak²²sep²²tiŋ⁵¹ la¹¹ hi⁵⁵ge²² ʔaj²²dí²²ja⁵¹ na²²
accepting *la* *hîge* *idea* **na**
 accepting PRT ART idea CMPL
- la²²me⁵¹ ko⁵⁵ ʔu³³ ka²²la²²'gu²²jo⁵¹
lamè *kô* *ū...* *ka-laguyò.*
 male PRT have COLL-close.friend

‘The culture of us Filipinos is very accepting of the idea that males have colleagues that they are very intimate with.’
 (CLIN-18-5:2363)

If the subordinate clause modifies the noun phrase, *nà* functions as a general relativizer used to link any type of relative clause or phrase to the noun phrase preceding it. If speakers desire to be

more specific about the type of relative clause (i.e., location, time, reason), they use *kungsaân* ‘where’, *nûng* ‘when’, and *kumbakît* ‘why’, discussed later.

- (417) gwa⁵⁵ kaj²²tan⁵¹ hi⁵⁵ge²² 'plant⁵⁵na³³ jo²²be⁵¹ ho²² gwa⁵¹
Guâ kay-tàn hîge plânt nā shobè ho guà.
 1.SG CAUS-throw ART plant REL little.sister give me
 ‘I caused the plant that little sister gave me to be thrown.’
 (elicitation, PC0068)

- (418) hi⁵⁵ge²² laŋ³⁵ na²² be²² pa³⁵la³⁵ ʔi⁵⁵e³³ ts^hja³³
Hîge lāng nā be pālā î=ē tshîā
 ART person REL NEG pay 3.SG=GEN car

 ti³³ti³³ ko⁵¹ ʔi²² ba³⁵
tītī kò i bá?
 ITER guard 3.SG Q

‘Is the person that hasn’t paid for their car continuing to guard them?’
 (PROT-16-NA:38436)

To highlight the contrast between the complementizer and relativizer roles, I provide an example containing two *nàs* with those roles.

- (419) ʔi⁵⁵ kaj²² 'dis²².i²²ga^{id}hi⁵⁵ge²² 'klejm⁵¹na²² ʔi⁵⁵ si³³ tsi³³ge³³
Î kay- disregârd hîge claim nā î sî tsîgē
 3.SG CAUS-disregard ART claim CMPL 3.SG COP ART

 'man²²du²².i²²kot⁵⁵ tsi²²ge²² 'klejm⁵¹na²² laŋ³⁵ loŋ²²tsoŋ⁵¹ 'fej⁵¹
mandurukôt – tsige claim nā lāng longtsòng share.
 thief ART claim REL person all share

‘They caused the claim that they are a thief – a claim that all people shared – to be disregarded.’
 (elicitation, PC0068)

Some speakers occasionally have another function for *nà* – as a pre- and post-modifier particle. They use it as a modifier particle only when it attaches to an adjective. This particle can be used for pre- and post-modification (see Section 3.4.7.1 for a more in-depth discussion).

(420) hi⁵⁵ge²² 'blak⁵⁵ na²² 'se¹²²fown⁵¹ ti³³ to²²lo³⁵ a⁵⁵
Hige blâck na cellphone tī tolóh â?
 ART.DEF.SG black MOD cellphone PREP where PRT
 'Where is the black mobile phone?'
 (elicitation, PC0068)

(421) gwa⁵⁵ ʔu²² 'fiends⁵⁵ na²² la³⁵naŋ³⁵
Guâ u friênds na Lánnáng.
 1.SG have friends MOD Lannang
 'I have Lannang friends.'
 (PC0095-CLIN19)

In the domain of the verb phrase, *nà* functions as a complementizer. It is first placed before the subordinate clause. Then, the resulting *nà*-headed complement phrase is placed after the verb phrase.

(422) gwa⁵⁵ kaj²²i²²'poit⁵⁵ na²² gwa⁵⁵e³³ ts^hju³⁵ki³³ t^hwe²² tjo⁵¹
Guâ kay-repôrt na guâ=ē tshiúkī thueh-tiòh.
 1.SG CAUS-report CMPL 1.SG=GEN phone grab-PFV
 'I reported that my phone was stolen (by someone).'
 (PROT-16-NA:38207)

(423) 'pe²²ro²² gwa⁵⁵ 'fist²² 'dej⁵¹ gwa⁵⁵ kaj²² di²²'sajd⁵⁵
Pero guâ first dàj guâ kay- decide
 but 1.SG first day 1.SG CAUS decide

 na²² ʔu²² laŋ³⁵ ʔa⁵¹ gwa¹¹
na u lánng ài gua
 CMPL have person love 1.SG

'But on the first day, I decided that someone loves me.'
 (CFH-001)

Other than the default relativizer *nà*, Lánnang-uè also has the following Tagalog-derived relativizers: *kungsaân* /kuŋ⁵⁵sa⁵¹ʔan⁵⁵/, 'where' *núng* /nuŋ⁵⁵/ 'when', and *kumbakít* 'why' /kum⁵⁵ba⁵¹kit⁵⁵/. The first two are multifunctional. When they do not function as subordinating conjunctions of location and time (see discussion earlier), they link relative constituents of location and time to the noun phrase, as exemplified below:

(424)	hi ⁵⁵ ge ²²	'ju ²² ni ²² vɿ ²² si ² ti ⁵¹	'kuŋ ²² sa ²² ʔan ²²	gwa ⁵⁵ e ³³	ʔa ²² tsi ⁵⁵
	<i>Hige</i>	<i>university</i>	<i>kungsaan</i>	<i>guâ=ē</i>	<i>atsî</i>
	ART	university	where	1.SG=GEN	big.sister
	le ⁵⁵	t ^h ak ²² ts ^h e ^ʔ ⁵⁵	si ³³	ti ³³	tʃi ²² 'ka ²² go ⁵¹
	<i>lê</i>	<i>thaktshêh</i>	<i>sī</i>	<i>tī</i>	<i>Chicagò.</i>
	PROG	study	COP	PREP	Chicago

‘The university where my sister is studying is in Chicago.’
(elicitation, PC0068)

(425)	gwa ⁵⁵	bwe ³³	ki ⁵⁵	hi ⁵⁵ ge ²² si ³³ kan ³³	nuŋ ²²	gwa ⁵⁵
	<i>Guâ</i>	<i>bō=ē</i>	<i>kī</i>	<i>hige sīkân</i>	<i>nung</i>	<i>guâ</i>
	1.SG	NEG=ABI	remember	ART time	when	1.SG
	p ^h ah ⁵⁵ si ⁵¹ tjoh ¹¹	tsi ²² ge ²²	'pa ²² la ²² ka ^ʔ ⁵⁵	lo ⁵¹		
	<i>phâhsî-tioh</i>	<i>tsige</i>	<i>palakâ</i>	<i>lò.</i>		
	kill-PFV	ART	frog	PFV		

‘I forgot the time when I successfully killed a frog.’
(elicitation, PC0068)

Kumbakî only has the relativizer function. It links the relative constituent of reason to the noun phrase preceding it.

(426)	ʔi ⁵⁵ e ²²	da ²² hi ²² lan ⁵⁵	kum ²² ba ²² kit ²²	ʔi ⁵⁵	laj ³⁵	si ³³	hi ⁵⁵	ge ³⁵
	<i>Î=e</i>	<i>dahilân</i>	<i>kumbakî</i>	<i>î</i>	<i>laí</i>	<i>sī</i>	<i>hī</i>	<i>gé.</i>
	3.SG=GEN	reason	why	3.SG	come	COP	DEM	CLS

‘Their reason why they did not come is that.’
(elicitation, PC0068)

3.9.3 General patterns & summary

All conjunctions can be placed before clauses, but only cumulative and disjunctive coordinating conjunctions, subordinating conjunctions of condition and substitution, and relativizers can be placed before the verb phrase to be conjoined. Only cumulative and disjunctive coordinating conjunctions can be situated before the pre-conjoined noun phrase.

Disjunctive, non-emphatic cumulative, general conditional, and manner conjunctions are derived from Hokkien. The conditional conjunction meaning ‘unless’ and the time conjunction meaning ‘after’ are both sourced from English while the rest are derived from Tagalog. Although a distribution for the conjunctions exists, some speakers occasionally vary in the selection of specific conjunctions from a specific language. For instance, some speakers use

adversative/contrastive Hokkien-origin conjunction *tānsī* /tan³³si³³/ ‘but’ instead of the Tagalog-origin *però* /pe⁵¹.io⁵¹/ ‘but’. Others use the disjunctive English-origin *òr* /ʔoɪ⁵¹/ ‘or’ instead of Hokkien-origin *ásī* /ʔa⁵⁵si³³/ ‘or’.

- (427) gwa⁵⁵ ja³⁵ kaŋ²²k^ho⁵⁵ t^hja³³ tan³³si³³gwa⁵⁵ tjo^{ʔ22} ʔo^{ʔ22} diŋ⁵¹ ta²²la²²ga⁵¹
Guâ yá kangkhô thiā tānsī guâ tioh óh diŋ talagà.
 1.SG very hard hear but 1.SG NEC learn also really
 ‘I have a very hard time listening to it but I should really learn.’
 (PC0071-CLIN18)

- (428) fi²²li²² pi²²no²² tʃ^haj²²nis⁵⁵ e²² tsjap⁵⁵ʃu³³ fi²²li²² pi²²no²² tʃ^haj²²nis⁵⁵
Filipino-Chinése e tsiâpshū Filipino Chinése
 Filipino-Chinese POS accept Filipino Chinese

ʔoɪ²² tʃ^haj²²nis²²fi²²li²² pi²²no⁵¹
or *Chinese Filipinò.*
 or Chinese Filipino

“The Filipino-Chinese will accept the term ‘Filipino Chinese’ or ‘Chinese Filipino’.”
 (PC0072-CLIN18)

Other examples of variation can be found below. The first conjunction *ând* – a non-emphatic cumulative conjunction – would typically have been derived from Hokkien while the others (in bold) – conjunctions of reason and general temporal relations – would typically have been derived from Tagalog.

- (429) 'dis²²tii²² bju²²ted⁵⁵ la⁵¹ 'ʔend²² tsa⁵⁵ʔjũ⁵¹ koŋ⁵⁵ a⁵¹
*Distributêd là **and** ... tsâ-iùnn kông a ?*
 distribute PRT and ... how say PRT
 ‘It was distributed and ... how do you say this?’
 (PC0005-CLIN18.eaf)

- (430) gwa⁵⁵ t^he⁵⁵po⁵¹ ʔin²²wi³³ gwa⁵⁵e³³ ma²²ma³⁵
*Guâ thêpò **īnwī**⁴⁶ guâ=ē mamá*
 1.SG deteriorate because 1.SG=GEN mom

bo²² ʃa⁵⁵ ʔe²²hjaw²² koŋ⁵⁵
bo shâ ehiaú kông.
 NEG really ABI speak

‘My knowledge deteriorated because my mom did not really know how to speak.’

⁴⁶ This word is derived from Hokkien.

(PC0072-CLIN18)

- (431) bo³³ 'stan²²daɪd⁵⁵ si³³ bwe³³jaw³³kin⁵⁵ bi²²'kos²² tswaj³⁵ lan³⁵nan³⁵
Bō standard sī buēyaūkîn because tsuai Lánáng
 no standard COP okay because DEM Lannang
- si³³ twa⁵⁵ pat²²tja⁵⁵ pat²²tja⁵⁵ laj³⁵ din³⁵ ko⁵⁵
sī tuâ pattiâh pattiâh laí dîn kô.
 COP from other.place other.place come also PRT

‘It’s okay not to have a standard because these Lannangs come from different places too.’

(PC0098-CLIN19)

- (432) 'bi²²foɪ²² gwa⁵⁵ si⁵⁵
Before guâ sí...
 before 1.SG die
 ‘Before I die...’
 (PC0068-CLIN19)

I provide an in-depth analysis of conjunctions and the variation in their use in Chapter 5. I summarize the conjunctions discussed in this section alphabetically along with a description, the conjunction (sub)type, source language, and their distribution in Table 16.

Table 16. The conjunctions of Lánang-uè

Conjunction	Gloss	Type	Subtype	Source	Dist.	Description
<i>afèr</i>	‘after’	subord.	time	English	_S	indicates that the clause contains/describes an event that is temporally sequenced before the event in the main clause
<i>ásī</i>	‘or’	coord.	disjunctive	Hokkien	_NP/ _VP/ _PP/ _AdjP/S	presents two or more constituents as alternatives
<i>bagò</i>	‘before’	subord.	time	Tagalog	_S	indicates that the clause contains/describes an event that is temporally sequenced before the event in the main clause
<i>habâng</i>	‘while’	subord.	time	Tagalog	_S	indicates that the clause contains/describes an event that is happening alongside the event in the main clause
<i>hanggâng</i>	‘until’	subord.	time	Tagalog	_S	indicates that the event or process in the clause marks the end of the event or process in the main clause

<i>hanggât</i>	'so/as long as'	subord.	condition	Tagalog	_S	introduces a proviso or a condition attached to an agreement
<i>kahît</i>	'although'	subord.	concession	Tagalog	_S	introduce a concession, or an idea that is granted in response to the main clause
<i>kahitnà</i>	'even if'	subord.	condition	Tagalog	_S	refers to clauses analogous to 'even if' clauses in English, coding the relation 'frustrated implication'
<i>kâp</i>	'and'	coord.	cumulative	Hokkien	_NP/ _VP/ _PP/ _AdjP/ _S	connects words, phrases, or clauses that are to be taken cumulatively or collectively
<i>kapâg/pâg</i>	'if'	subord.	condition	Tagalog	_S	introduces present, habitual/generic, and predictive conditions
<i>kasi</i>	'because'	subord.	reason	Tagalog	_S	links two clauses, where one clause contains/describes the effect and the other, the cause; introduces the cause
<i>kasò</i>	'but it is the case that'	coord.	adversative	Tagalog	_S	introduces a statement that contrasts with or seems to contradict a statement that was said previously; emphasizes that the utterance following it is a statement of fact.
<i>kayâ</i>	'so/that is why'	subord.	consequence	Tagalog	_S	links two clauses, where one clause contains/describes the cause and the other, the effect; introduces the effect
<i>kaysà</i>	'instead of'	subord.	substitution	Tagalog	_S/ _VP	indicates choice or negation; indicates that something is done in place of something else
<i>khalâng/ khânân</i>	'like/as'	subord.	manner	Hokkien	_S	introduces clauses referring to a manner
<i>kiaū</i>	'and'	coord.	cumulative	Hokkien	_NP/ _VP/ _PP/ _AdjP/ _S	connects words or phrases that are to be taken collectively
<i>kumbakît</i>	'why'	subord.	relativizer	Tagalog	_S	used to connect the relative constituent of reason to the the noun phrase
<i>kungsaân</i>	'where'	subord.	location	Tagalog	_S	indicates that the clause following it contains/describes a reference to a place or situation
	'where'	subord.	relativizer	Tagalog	_S	used to connect the relative constituent of location to the the noun phrase

<i>maskì</i>	‘although’	subord.	concession	Tagalog	_S	introduce a concession, or an idea that is granted in response to the main clause
<i>nà</i>	‘that’	subord.	relativizer	Tagalog	_S	used to connect the relative clause or verb phrase to the noun phrase
<i>nà</i>	‘that’	subord.	complementizer	Tagalog	_S	marks an embedded clause as functioning as a complement
<i>nā</i>	‘if’	subord.	condition	Hokkien	_S	introduces any condition
<i>nántshiū</i>	‘like/as’	subord.	manner	Hokkien	_S	introduces clauses referring to a manner
<i>nūng</i>	‘when’	subord.	time	Tagalog	_S	indicates that once the event or process in the clause happened, the event in the main clause immediately happens.
	‘when’	subord.	relativizer	Tagalog	_S	used to connect the relative constituent of time to the the noun phrase
<i>pagkà</i>	‘as soon as’	subord.	condition	Tagalog	_S	introduces a condition that, when fulfilled, immediately triggers an event or process
<i>parà</i>	‘so that’	subord.	purpose	Tagalog	_S	indicates the purpose for the action in the main clause
<i>però</i>	‘but’	coord.	adversative	Tagalog	_S	introduces a statement that contrasts with or seems to contradict a statement that was said previously
<i>porkèt</i>	‘just because’	subord.	reason	Tagalog	_S	used when the speaker hopes to say that a particular situation should not necessarily make one come to a particular conclusion (the main clause)
<i>tapôs</i>	‘and then’	coord.	cumulative	Tagalog	_NP/ _VP/ PP/ AdjP/ S	connects two constituents; emphasizes sequence, with the preceding constituent coming first and the succeeding one coming second
<i>tsakà</i>	‘and also’	coord.	cumulative	Tagalog	_NP/ _VP/ PP/ AdjP/ S	connects two constituents; indicates that the preceding constituent carries a stronger emphasis than the constituent following it.
<i>tshīntshiū</i>	‘like/as’	subord.	manner	Hokkien	_S	introduces clauses referring to a manner
<i>tuwīng</i>	‘when-ever’	subord.	time	Tagalog	_S	indicates that event or process in the subordinate clause occurs whenever the event or process in the main clause occurs.

<i>unlèss</i>	'unless'	subord.	condition	English	_S	introduces the only circumstances in which an event you are mentioning will not take place or in which a statement you are making is not true
---------------	----------	---------	-----------	---------	----	---

3.10 Interjections

Lánnang-uè has interjections – exclamatory forms that “express states of mind and do not enter into specific syntactic relations” with other constituents (Matthews 2007:474). Some of the interjections are sourced from Hokkien and English, but most of them are derived from Tagalog.

I identify and describe the major ones in

Table 17 below. Some examples are given after the table.

Table 17. Common interjections in Lánnang-uè (with linguistic source)

Interjection	IPA	Gloss	Description	Source
<i>yúck</i>	/ˈjak ⁵⁵ /	‘yuck’	exclamation of disgust	English
<i>arây (ko)</i>	/ʔa ⁵¹ ˈaj ⁵⁵ / /ʔa ⁵¹ ˈaj ⁵⁵ ko ⁵¹ /	‘ouch’	exclamation of pain	Tagalog
<i>ây</i>	/ˈʔaj ⁵⁵ /	‘oh’	exclamation of surprise	Tagalog
<i>sâyâng</i>	/ˈsa ⁵¹ ˈjaŋ ⁵⁵ /	‘what a waste’	exclamation of waste	Tagalog
<i>(hay)nakù</i>	/ˈhaj ⁵¹ na ⁵¹ ˈko ⁵¹ / /na ⁵¹ ˈko ⁵¹ /	‘come on’	exclamation of frustration	Tagalog
<i>patây</i>	/pa ⁵¹ ˈtaj ⁵⁵ /	‘oh no’	exclamation of concern for error or problem	Tagalog
<i>sikuē</i>	/si ³⁵ kwe ³³ /	‘oh no’	exclamation of concern for error or problem	Hokkien
<i>wà</i>	/wa ⁵¹ /	‘wow’	exclamation of wonder	Tagalog
<i>ùy</i>	/ˈʔuj ⁵¹ /	‘hey’	exclamation to call one’s attention	Tagalog
<i>bwisît</i>	/ˈbwi ⁵¹ sit ⁵⁵ /	‘damn it’	exclamation of annoyance	Tagalog
<i>halà</i>	/ha ⁵¹ ˈla ⁵¹ /	‘someone’s in trouble’	exclamation of warning	Tagalog
<i>há</i>	/ˈha ³⁵ /	‘huh’	exclamation of confusion	Tagalog

(433) ʔuj⁵¹ di⁵⁵ si³³ hwa²²na⁵⁵
Ûy! Dî sî Huānnâ?
 hey 2.SG COP Filipino
 ‘Hey! You’re Filipino?’
 (PC0092-CLIN19)

(434) ʔaj⁵⁵ ts^ham ta²²ˈga²²log³⁵ kap⁵⁵ ʔjeŋ³³bun³⁵
Ây. Tshām Tagalóg kâp Iēngbún?
 oh mix Tagalog and English
 ‘Oh, you mean, mixing Tagalog and English?’
 (PC0094-CLIN19)

- (435) 'ta²²pos⁵⁵ 'pa²²raŋ²² le⁵⁵ tjam⁵⁵ 'ʔe²²le²²'vej²²toɪ⁵¹ le⁵⁵
Tapôs *parang* *lê* *tiâm* *elevator* *lê*
and.then like PROG at elevator PROG
- ʔu²²mi²²ʔi²²hi⁵⁵ ʔa⁵⁵ni²² 'jak⁵⁵
umihî *âni.* *Yûck!*
urinate like.that yuck

‘And then they are like loitering around the elevator and urinating. Yuck!’
(PC0095-CLIN19.eaf)

- (436) ha²²'la⁵¹ pa²²'taj⁵⁵ so²² gwa⁵⁵ ʔe²²ʔiŋ⁵⁵
Halâ. *Patây.* *So* *guâ* *e-îng*
someone's.in.trouble. oh.no so 1.SG PER
- pa²²'ʔu²²so⁵¹ 'kwen²²to⁵¹ o¹¹
pausò *kwentò* *o?*
popularize story PRT
‘Oh no. So are you saying that I can popularize a story?’
(PC0005-FRST18.eaf)

- (437) ha³⁵ di⁵⁵ tsja²² ja²² tswe⁵¹ a¹¹ hi⁵⁵ge²² 'lantʃ⁵⁵
Há? *Dî* *tsiah* *yá* *tsuè* *a* *hîge* *lûnch*
huh 2.SG eat very many PRT ART lunch
‘Huh? You ate a lot for lunch.’
(PC0005-CLIN18.eaf)

3.11 Discourse particles

In this section, I describe all Lánnang-uè discourse particles – particles that manage the flow and structure of discourse rather than that of individual sentences or clauses (Matthews 2007:272). These particles – the majority of which are derived from Tagalog – convey “more than what is said in an actual utterance... in such a way as to convey a speaker’s attitude or stance, and to guide the hearer towards the speaker's intended meaning” (Tay et al. 2016:482). I discuss them alphabetically, beginning with *à*.

The Hokkien-sourced particle *à* /ʔa⁵¹/, placed clause-finally, indicates confirmation of the situation expressed in the clause with undertones of insistence.

- (438) ʔin⁵⁵ si²² 'part⁵⁵ a⁵¹
În si pârt à.
 3.SG COP part PRT
 'They are a part [of something]. Really.'
 (CLIN-19-82:11515)

The Tagalog-derived clause-final particle *há* [ha³⁵] is used to seek agreement with the listener. It is sometimes pronounced as *á* [a³⁵]. To my knowledge, these two variants are in free variation.

- (439) tsi⁵⁵ tsja^{ʔ55} ho²² gwa⁵¹ ha³⁵
Tsí tsiâh ho guà há?
 DEM CLS give 1.SG PRT
 'Give this one to me, okay?'
 (elicitation, PC0068)

- (440) gwa⁵⁵ si³³ lan³⁵naŋ^{35e33} hwi²²di²²pin²² laŋ³⁵ a³⁵
Guâ sî Lánáng=ē Huīlīpīn láng á.
 1.SG COP Lannang=MOD Philippines person PRT
 'I am Lannang Filipino, okay?'
 (CLIN-19-114:19079)

Bà /ba^{X1}/, derived from Tagalog, is used to intensify questions and is used to express annoyance at the addressee. It is placed clause-finally.

- (441) ʔu³³ ʃam³³mi^{ʔ55} 'pa⁵¹ 'ba¹¹
Ū shāmmîh pà bà?
 have what yet PRT
 'What else do you have?!?'
 (CLIN-19-68:3842)

- (442) to²²lo^{ʔ35} ba⁵¹
Tolóh bà?
 where PRT
 'Where?!?'
 (elicitation, PC0068)

Bali 'so' /ba⁵¹le⁵¹/, derived from Tagalog, emphasizes that the utterance following it is relevant to or connected to a prior utterance. It is placed before the utterance, as in:

- (443) ba²²li²² sjen³³sĩ³³ ʔe²² ka⁵⁵ ko²²gi⁵⁵ ko⁵⁵
Bali *siēnsīnn* *e* *kâ* *Kogĩ* *ko...*
 PRT teacher POS teach Mandarin PRT
 ‘So the decision is that teachers will teach Mandarin.’
 (CLIN-18-5:2205)

Bastà [bas²²ta⁵¹], also derived from Tagalog, is used to indicate that the (action in the) utterance is important, no matter the circumstances. It can be placed before verb, prepositional, and noun phrases to highlight their importance. It can also be placed before or after the clause. To my knowledge, there are no factors that condition the use of clause-initial *bastà* over clause-final one, although the clause-initial one is the one more popularly used.

- (444) bas²²ta⁵¹ lan⁵⁵ tsju³³ si³³ tso^{ʔ55} sit⁵⁵
Bastà *lân* *tsiū* *sĩ* *tsôh* *sít.*
 PRT 1.PL.INC at.once COP make thing
 loosely: ‘No matter what the circumstance, it is imperative that we work immediately.’
 (CLIN-19-141:35938)

- (445) ʔi⁵⁵ bas²²ta⁵¹ k^ha⁵⁵ bi²²hejv⁵⁵ laŋ⁵⁵ la⁵¹
Í *bastà* *khâ* *behavê* *lâng* *la...*
 3.SG PRT CMPV behave only PRT
 loosely: ‘What is important about them, no matter the circumstance, is that they are only more behaved...[compared to someone]’
 (CLIN-19-126:26765)

- (446) bas²²ta⁵¹ tjoŋ³³kok⁵⁵ hoŋ²²sjok³⁵
bastà *Tiōngkôk* *hongsiók*
 PRT China culture
 loosely: ‘Chinese culture is important, no matter what the circumstance.’
 (CLIN-19-52:9193)

- (447) pa²²raŋ²² tjam⁵⁵ to²²lo^{ʔ22} to²²lo^{ʔ35} pa²²raŋ²²
Parang *tiâm* *tolóh* *tolóh* *pārāng*
 like PREP where where like

 ta²²pun⁵⁵ hi⁵⁵ge²² ba²²su²²ja⁵¹ bas²²ta⁵¹
tapôn *hĩge* *basurà* *bastà.*
 throw ART.DEF.SG trash PRT

loosely: ‘(They) like throw like trash wherever. This is important, no matter what the circumstance.’
 (CLIN-19-95:14125)

The Tagalog-derived particle *dâw* /daw⁵⁵/ is placed after the clause or constituent when the speaker wants to express or introduce quotations or reported speech.

- (448) ka²²'si⁵⁵ tjeŋ³³maj⁵⁵ ʔu²² kaŋ³³k^ho⁵⁵ 'daw⁵⁵
Kasî *tiēngmaî* *u* *kāngkhô* ***dâw*** ...
 because before have hardship PRT .
 'Because (they) had hardship before, I heard.'
 (PC0070-CLIN18)

To ask for confirmation, the Tagalog-derived particle *dibá* /di⁵¹'ba³⁵/ is used. It is placed before or after the clause interchangeably to request confirmation for the whole clause, or a before a phrase if the speaker wants to only confirm the phrase.

- (449) ka²²'si⁵⁵ di²²ba³⁵ 'si²²ŋa²²po¹51 ʔu³³ 'siŋ³³li⁵⁵
Kasî ***dibá*** *Singaporè* *ū* *Sīnglīsh?*
 because PRT Singapore have Singlish
 'Singapore has Singlish, right?'
 (CLIN-18-5:1877)

- (450) hwi³³di³³pin³³di²²ba³⁵ ʔu³³ le⁵⁵ 'se²²le²²b.rejt⁵⁵
Huīlīpīn ***dibá*** *ū* *lê* *celebratê*
 Philippines PRT PF PROG celebrate

 'tʃ^haj²²nis²² 'nju²² 'ji⁵¹
Chinese *New* *Yeàr?*
 Chinese New Year
 'The Philippines habitually celebrates Chinese New Year, right?'
 (CLIN-18-5:2133)

- (451) ʔa²²me²².i²²kan⁵¹ bwe²² 'ʔan²²dɪ²²stand⁵⁵ gua⁵¹ di²²ba³⁵
Americàn *bo=e* *understând* *guà* ***dibá?***
 American NEG=ABI understand 1.SG PRT
 'Americans are not able to understand me, right?'
 (CLIN-18-5:1896)

To indicate that the clause is a reason, the particle *êh* /e⁵⁵/ is interchangeably placed before or after it. In some cases, clauses may have both a clause-initial and clause-final *êh* without known changes in discourse structure. Unlike the conjunctions of reason, the *êh* particle may be used in independent clauses.

(452) gwa⁵⁵ bo²² kaj²²'pa²²si²²'ka²²so⁵¹ e⁵⁵
Guâ bo kay-asikasò êh.
 1.SG NEG CAUS-handle PRT
 'The reason is because I did not cause [something] to be handled.'
 (CLIN-18-5:2617)

(453) tsi⁵⁵ ge²² lan³⁵ ja³⁵ 'di²².ii²?⁵⁵ e⁵⁵
Tsî ge lán yá dirî êh.
 this CLS person very disgusting PRT
 'Why? Because this person is very disgusting.'
 (CFH-001)

(454) e²² ka²²'si⁵⁵ tjeŋ²²ke⁵¹ lan⁵⁵ ja³⁵ kjã taj²²djok³⁵e⁵⁵
Êh kasî tiengkè lân yá kiãnn Taidiok êh.
 PRT because before 1.PL.INC very fear China PRT
 'The reason is because before, we feared China.'
 (CLIN-18-5:2603)

The Hokkien-derived *kô* /ko⁵⁵/ is an utterance chunker – it facilitates the listening comprehension of the addressee in an explanation or process that is perceived to be complex and difficult to comprehend. It is used in contexts of demonstration or instruction.

(455) di⁵⁵ kaj³³ki³³ ts^hoŋ⁵⁵ ko⁵⁵si⁵¹ ko⁵⁵
Dî kaikî tshông kôsî kô...
 2.SG self make story PRT
 'First, you yourself should make a story...'
 (PC0020-FRST18)

(456) ta²²'pos⁵⁵ ko⁵⁵ hi⁵⁵ge 'se²²kond²² 'step⁵⁵ ko⁵⁵ si³³
Tapôs kô hîge second stêp kô sî
 then PRT ART second step PRT COP

 'pɿe²²pa²²'ɿej²²ʃon⁵¹ ko⁵⁵
preparatiòn kô...
 preparation PRT

'Then, the second step is preparation...'
 (elicitation, PC0068)

The Hokkien-derived *là* /la⁵¹/ particle is analogous to the English 'I am telling you'. It emphasizes that what the speaker is saying is true and should be believed. It has undertones of dismissiveness and is placed after the clause.

(457) ʔi⁵⁵ bo²² suj⁵⁵ la⁵¹
Í bo suí là.
 3.SG NEG pretty PRT
 ‘I am telling you. She’s not pretty.’
 (elicitation, PC0068)

(458) laŋ²²gwedʒ⁵⁵ ja³⁵ flu²²wid⁵⁵ din⁵⁵ la⁵¹
Languâge yá fluíd dîn là.
 language very fluid also PRT
 ‘I am telling you. Language is also very fluid!’
 (CLIN-18-7:3559)

(459) 'pe²²ro²² hi⁵⁵ ge³⁵ bo²² səŋ⁵⁵ hwaj³⁵
Pero hí gé bo sŋng huái
 but that CLS NEG count ART.DEF.PL

 ma²²ŋa²² sin³³kjaw³⁵ tse²²tsun⁵¹ dik³³ laj⁵¹ la¹¹
mga sŋkiaú tsetsùn dik lài là.
 PL new.immigrants now enter DIR PRT
 ‘But that doesn’t count the new immigrants who have entered.’
 (PC0005-CLIN18.eaf)

Hokkien-derived *lè* /le⁵¹/, on the other hand, is used in queries, especially in order to express doubts or to check validity or accuracy. It is used by the addressee to bounce questions back to their interlocutor. It is used only in the context of questions and can only be placed after phrases.

(460) tsi⁵⁵ ge²² le⁵¹
Tsí ge lè?
 DEM CLS PRT
 ‘How about this?’
 (FRST-19-132:30838)

To indicate that the clause is informative and something that the listener(s) should know, the particle *nâ* /na⁵⁵/ is placed after the clause.

(461) ʔi⁵⁵ ja³⁵ hwa²²hi⁵⁵ na⁵⁵
Í yá huahî nâ.
 3.SG very happy PRT
 ‘They are very happy, you know.’
 (PC0002-FRST18)

The Tagalog-derived particle *namân* /na⁵¹man⁵⁵/ has five functions. First, it expresses dissimilarity or contrast from an entity or idea mentioned or established previously. It is very

similar to ‘on the other hand’ used in English. When used for this function, it is placed only after the noun phrase with a subject role.

(462) hi⁵⁵ge²² nuŋ³³ e³³ 'ku¹²²tʃu⁵¹ na²²'man⁵⁵ ko⁵⁵ o³⁵
Hige *nūng* *ē* *cultùre* ***namân*** *kô* *ó*
 ART.DEF two CLS culture PRT PRT PRT

ʔe²² ʔa²²'pli²²ka²²bo¹⁵¹
e *applicablè...*
 POS applicable

‘The two cultures, on the other hand, will be applicable...’
 (CLIN-19-9:3782)

Second, it can soften an utterance and/or add hints of reservation to make it more casual, friendly, and/or polite. In answers to questions about attributes, the particle qualifies positive attributes (e.g., *bo=e phaî* ‘not bad’, *okày* ‘okay’, *hosè* ‘good’, *spacioûs* ‘spacious’), resulting in a politeness reading. When used for this function, it is placed only after the clause or an adjective.

(463) laŋ²²gwejd⁵⁵ tsju³³ si³³ ts^hin⁵⁵ts^haj⁵⁵ laj³⁵ na²²'man⁵⁵
Languâge *tsiū* *sī* *tshîntshai* *lai* ***namân***.
 language at.once COP haphazardly come PRT
 ‘Languages immediately and haphazardly come about, I guess.’
 (CLIN-19-16:4928)

(464) ja³⁵ kwaj³³ na²²'man⁵⁵
Yá *kuaī* ***namân***.
 very good PRT
 ‘(They are) very good, I guess.’
 (elicitation, PC0068)

(465) Q: di⁵⁵ tʰak²²tsʰeʔ⁵⁵ tsaj⁵⁵ʔjũ⁵¹
Dî thaktshêh tsâiunn?
 2.SG study how
 ‘How’s your studying?’

A: ham⁵⁵ham⁵¹ na²²man⁵⁵
Hâmhâm namân.
 so.so PRT
 ‘So so, I guess.’

(CLIN-18-5:2160)

Third, it encodes both politeness and mild reproach or complaint in imperatives.

(466) di⁵⁵ kaj²² khjɔʔ⁵⁵ ho²²se⁵¹ na²²man⁵⁵ o⁵¹
Dî kay- khiôh hosè namân ò.
 2.SG CAUS pick.up well PRT PRT
 ‘Pick it up well.’
 (PROT-16-NA:38683)

Fourth, the particle is used to encode politeness in questions related to the addressee’s well-being.

(467) ka²²mus²²ta⁵¹ na²²man⁵⁵
Kamustà namân?
 how.are.you PRT
 ‘How are you?’
 (PROT-16-NA:37549)

Finally, it is used to emphasize the intense feelings or attitudes of the speaker(s), but only in exclamatory utterances with an adjective as a predicate or utterances that contain an adjective. It is placed after the adjective or phrase and can be used sarcastically or jokingly.

(468) swɛɪ²²te⁵¹ na²²man⁵⁵ a⁵¹ ʔa⁵⁵ni³³
Swertè namân à ânī.
 lucky PRT PRT like.that
 ‘How lucky is that.’
 (PROT-16-NA:37763)

(469) hi⁵⁵ ge³⁵ swe⁵⁵ nja⁵⁵ na²²man⁵⁵ o⁵¹
Hî gé, suê niâ namân ò!
 DEM CLS small CLS PRT PRT
 ‘How small is that piece of clothing!’
 (PROT-16-NA:38528)

The Tagalog-derived *ngâ* /ŋaʔ⁵⁵/ is used to emphasize something with a tinge of annoyance. It can also be used in the context of a dare or a challenge, especially when the speaker is slightly annoyed. It is placed after the clause.

(470) hi⁵⁵ge²² 'wej⁵¹ naŋ²² ko²²'mju²²ni²²'kej²²ʃon⁵¹
Hîge wày ng communication
 ART.DEF.SG way of communication

k^ha⁵⁵ kin⁵⁵ 'ŋa^ʔ⁵⁵
khâ kîn **ngâ.**
 CMPV fast PRT

'I told you that the way of communication is faster.'
 (PC0009-CLIN19)

(471) bo²² mŋ⁵¹ ʔi¹¹ pa¹¹ 'ŋa^ʔ⁵⁵
Bo mng i pa **ngâ.**
 NEG ask 3.SG still PRT
 'I told you. I still haven't asked them.'
 (PROT-16-NA:37730)

(472) kaj³³ki³³ hwa³³ ts^hja³³ k^hi⁵¹ 'ŋa^ʔ⁵⁵
Kaīkī huā tshīā khì **ngâ.**
 self hold car DIR PRT
 'Drive solo. I dare you.'
 (PROT-16-NA:38973)

The Tagalog-derived *nóh* /no³⁵/ is used after a clause when the speaker wants to ask for confirmation for something expressed in it.

(473) gun⁵⁵ laj³³ hwi³³di³³pin³³ 'no³⁵
Gún laī Huīlīpīn **nóh?**
 1.PL.EXC come Philippines PRT
 'We came to the Philippines, right?'
 (CLIN-19-125:25837)

To convert a non-exclamatory utterance into an exclamatory one, the Hokkien-derived *ò* /ʔo⁵¹/ is placed after the clause. It can additionally express wonder or surprise.

(474) laj²² o⁵¹
Lai ò!
 come PRT
 ‘Come!’
 (PROT-16-NA:37179)

(475) ja²² suj⁵⁵ o⁵¹
Ya suí ò!
 very pretty PRT
 ‘Wow. So pretty!’
 (elicitation, PC0068)

The Hokkien-derived particle *hó* [ho³⁵], also pronounced as *ó* [o³⁵], is used as an attention-grabbing device after the clause or noun phrase. It has an emphatic reading as well. In terms of function, both variants are identical and are used interchangeably. To my knowledge, there are no factors that condition the use of one variant over the other.

(476) gwa⁵⁵ kam²²kak⁵⁵ taj²²wan³⁵ le⁵⁵ 'gow⁵¹ 'ʔin²²to²²
Guâ kamkâk Taiwán lê gò into
 1.SG feel Taiwan PROG go into

tia²² di²²ʃo²²nal⁵¹ o³⁵
 traditional PRT
traditionàl ó, ...

‘I feel that Taiwan is shifting into traditional [ways],’
 (CLIN-19-133:30977)

(477) ʔi⁵⁵ twa⁵⁵ hja³⁵ t^hak³⁵ ho³⁵
Î tuâ hiá thák hó...
 3.SG PREP DEM read PRT
 ‘They read there, right?’
 (CLIN-18-4:1591)

(478) ʔi⁵⁵ ho³⁵ twa⁵⁵ hja³⁵ k^hun⁵¹ ho³⁵
Î hó tuâ hiá khùn hó...
 3.SG PRT PREP DEM sleep PRT
 ‘They, right, slept there, right?’
 (elicitation, PC0068)

- (479) 'tʃ^haj²²nis⁵⁵ o³⁵ laj³³ gun⁵⁵ kaw⁵⁵tɕ³⁵
Chinesê ó, laī gún kaúting.
 Chinese PRT come 1.PL.EXC church
 'The Chinese came to our church.'
 (CLIN-19-134:31359)

To mark surprise, the Tagalog-derived *palâ* /pa⁵¹laʔ⁵⁵/ is placed after the clause.

- (480) hi⁵⁵ ge²² 'kul²²tʃu⁵¹ ja³⁵ 'ja²²kiŋ⁵¹ pa²²'laʔ⁵⁵
Hî ge cultùre yá shocking palâ.
 DEM CLS culture very shocking PRT
 'It came as a surprise to me that the culture is very shocking.'
 (CLIN-19-41:7246)

The Tagalog-derived particle *parâng* /pa⁵¹raŋ⁵⁵/ denotes '(it is) like'. When not used as a conjunction of manner, it is used to qualify or hedge a clause when placed at the beginning of a clause. In all other contexts, it is used as a discourse filler.

- (481) 'pa²².Iaŋ²² ʔin⁵⁵ ʔu³³ 'slajt⁵⁵ laŋ⁵⁵ ko⁵⁵
Parang in ū slight lāng kô
 PRT 3.PL have slight only PRT
 'It appears that they only have a slight [of something].'
 (CLIN-19-9:3817)

- (482) ʔu²² tsi²²ge²² so³⁵tsaj³³ hi⁵⁵ge²² 'pa²².Iaŋ²² ma²²ma²²
U tsige sótsai híge parang Mama
 have a place ART PRT mother
- me²².i⁵¹ ʔan⁵⁵ hja⁵⁵ ja²² suj⁵⁵ la⁵¹
Mary an hiá yá suí là
 Mary PREP there very beautiful PRT

'There is a place where, like, Mother Mary is being beautiful there.'
 (E-004)

- (483) ka²²'si⁵⁵ ʔin⁵⁵ 'pa²².Iaŋ²² ho²² ʔin⁵⁵ 'swel²²do⁵¹ a¹¹
Kasí in parang ho in sweldò a...
 Because 3.PL PRT give 3.PL salary aRT
 'Because they like gave them salary. Really.'
 (CLIN-19-126:26702)

To mark hope for a whole clause, the Tagalog-derived particle *sanà* /sa⁵¹na⁵¹/ is placed before or after the clause interchangeably. There is not, to my knowledge, any factor that conditions the clause-initial position over the clause-final one.

(484) gwa⁵⁵ ʔaj⁵¹ 'sa²²na⁵⁵
Guâ ài sanà.
 1.SG ABI PRT
 'I want to, hopefully.'
 (CLIN-19-101:15766)

(485) tʃaŋ²²ka²²ba⁵⁵ 'sa²²na⁵¹ ʔi⁵⁵ ʔe³³ ts^he⁵¹tjo^ʔ¹¹
Tshangkabâ, sanà î ē tshè-tioh.
 frog PRT 1.SG ABI find-PFV
 'Hopefully they will be able to find the frog successfully.'
 (FRST-20-19:6600)

If speakers want to mark hope on the verb phrase, then the particle is placed before the verb phrase.

(486) 'jaŋ²²gɿ²² 'dʒe²²ne²²'jej²²ʃon⁵¹ 'sa²²na⁵¹ ʔu³³ tsi³³ge³³
Younger generation sanà ū tsīgē
 younger generation PRT have ART

stan³³daɪd⁵⁵
stāndârd.
 standard

'The younger generation hopefully should have a standard.'
 (CLIN-19-41:7575)

A summary of the discourse particles, including a brief description of them, their distribution, and source language, is presented alphabetically in Table 18.

Table 18. Discourse particles

Particle	Description	Distribution	Source
<i>à</i>	indicates confirmation of a situation with undertones of insistence	S __	Hokkien
<i>(h)á</i>	used to seek agreement	S __	Tagalog
<i>bà</i>	an intensifier used in questions, can be used to express annoyance	S __	Tagalog
<i>balì</i>	indicates that the utterance following it is relevant to or connected to a previous utterance.	__S	Tagalog
<i>bastà</i>	indicates that the utterance or action is important, no matter the circumstances	__ S/VP/ PP/NP S __	Tagalog

<i>dâw</i>	reportative particle used when one is trying to quote a person	S/VP/ PP/NP	Tagalog
<i>dibá</i>	a question particle used to ask for confirmation	__ S/VP/ PP/NP S	Tagalog
<i>éh</i>	indicates that the utterance is a reason	__ S S__	Tagalog
<i>hó</i>	used to ask for confirmation	S/NP__	Hokkien
<i>kô</i>	guides the listener by breaking down complex utterances; used in contexts of demonstration or instruction	utterance __	Hokkien
<i>la</i>	emphasizes that what the speaker is saying is true and should be believed with undertones of dismissiveness	S__	Hokkien
<i>lè</i>	used in a queries or questions about something, especially in order to express one's doubts about it or to check its validity or accuracy; used to bounce questions back when you have just been asked them	VP/PP/NP__	Hokkien
<i>nâ</i>	indicates that the clause is informative and something that the listener(s) should know.	S__	Hokkien
<i>namân</i>	expresses dissimilarity or contrast from an entity or idea mentioned or established previously, similar to 'on the other hand'	NP__	Tagalog
	softens the utterance and/or adds hints of reservation to make it more casual, friendly, and/or polite; in answers to questions about attributes, qualifies only positive attributes	S/Adj__	Tagalog
	encodes both politeness and mild reproach or complaint in imperatives	S__	Tagalog
	encodes politeness in questions related to the addressee's well-being	S__	Tagalog
	emphasizes the intense feelings or attitudes of the speaker(s) in exclamatory utterances, with an adjectives in/as a predicate; can be used sarcastically or jokingly with hints of jealousy	S__	Tagalog

<i>ngâ</i>	used to emphasize something again with a tinge of annoyance; used in challenges and dares	S __	Tagalog
<i>nóh</i>	used to ask for confirmation	S	Tagalog
<i>ò</i>	used to mark a non-interjection utterance as exclamatory; can express wonder or surprise	S__	Hokkien
<i>(h)ó</i>	used as an attention-grabbing device	S/NP__	Hokkien
<i>palâ</i>	usually used when stating a fact that has just been discovered (out of surprise)	S__	Tagalog
<i>parâng</i>	roughly means ‘(it is) like’; a common discourse filler, it can also introduce a hedged clause	__ S/VP/NP	Tagalog
<i>sanà</i>	indicates that the utterance is a hope	__ S/VP S__	Tagalog

3.12 Summary and concluding remarks

In this chapter, I described the lexicon as well as the phonological, morphological, and syntactic structure of Lánnang-uè – specifically the variety used in Manila – as an attempt to document it comprehensively for the first time.

To date, only one study, that of Tsai (2017), has attempted to describe Lánnang-uè. However, their study described the phonology and the lexicon of Hokkien and not Lánnang-uè as defined in Chapter 2.2. My study, on the other hand, describes and analyzes Lánnang-uè from the phonological level to the pragmatic level using data collected first-hand. It is, thus, to my knowledge, the first comprehensive description of the variety.

My descriptive analysis yielded four key findings:

1. ***Systematicity***. The variety is far from a random mix of Hokkien, Tagalog, English and Mandarin, as claimed by some speakers during my fieldwork. The existence of conventions, such as the number convention involving place markers (Section 3.4.5), as well as the convention on the placement of adverbs (Section 3.5.9) in particular adverbial positions, supports this claim. The existence of tone sandhi conventions (Section 3.3.3.1), phonotactic constraints (Section 3.3.2), and affix attachment constraints (Sections 3.4.9,

3.4.10) is also evidence of systematicity. The use of particular discourse particles in specific situations is further support for a systematic Lánnang-uè.

I also found that Lánnang-uè consistently sources certain elements from particular languages. For instance, in the lexicon, I observed that Lánnang-uè consistently derives its prepositions of general location, orientation, and range/path from Hokkien, and its prepositions of accompaniment, specific location, and temporal relations as well as the preposition meaning ‘of’ from English (Section 3.8). The variety consistently derives disjunctive, non-emphatic cumulative, general conditional, and manner conjunctions from Hokkien, the conditional conjunction meaning ‘unless’ and the time conjunction meaning ‘after’ from English, and the rest of the conjunctions from Tagalog. The systematicity in sourcing linguistic elements can also be observed at the structural level – for instance, the personal pronoun system (e.g., number and person contrast) and the modality system (with the exception of Tagalog-derived *dapât*) are derived from Hokkien (Section 3.4.1), while the derivational system (Section 3.4.10) and the approximation system (3.4.7.4) in the noun phrase domain, as well as the yes/no question marker, are derived from Tagalog. Altogether, the evidence indicates systematicity in Lánnang-uè.

2. **High levels of spread and stability.** I observed that many of the features described in the chapter were widespread in my sample, at least based on my corpus, elicitation, and judgment data. That is, only a couple of speakers did not use these features at all. I also observed that many of features in Lánnang-uè were stable – speakers who had these features at all used them with high levels of consistency. For example, the aspectual system was observed to be used by all speakers that I randomly sampled in the corpus (the system is widespread); in addition, these speakers followed the system with high levels of consistency (the system is stable).
3. **Correlation of features within speakers.** I found that the features I described were correlated with each other – speakers who, for example, used the general question marking particle *bá* also positioned their *why*-phrases in the sentence-initial position in *wh*-questions. They also used other features enumerated in this chapter.

4. ***Varying degrees of ‘unstructured’ variation.*** Lánnang-uè has features that exhibit minimal ‘unstructured’ variation – patterns of variation that I, at present, could not explain using linguistic factors. One such feature is superlative marking: *té* is almost always used to mark a superlative construction (96% of 369 superlative constructions); *pinakà-* is rarely used (4%) (Section 3.4.7.3). However, the variety also has features that anecdotally exhibit a higher degree of ‘unstructured’ variation than others (e.g., stress in Section 3.3.3.2, distributional patterns of conjunctions and prepositions in Sections 3.8 and 3.9). For example, I observed that 67% of the 2,151 conjunctions in the Lannang Corpus that mean ‘because’ (Section 3.8) in Lánnang-uè were expressed using the Tagalog-derived variant *kasî* ‘because’; 26% were Hokkien-derived (i.e., *īnwī* ‘because’); and 8% were English-derived (i.e., *becaûse* ‘because’).

Overall, my descriptive analysis of Lánnang-uè provides some evidence that the variety is highly systematic and stable. It also shows that many of these elements are widespread within my sample (and, perhaps, the Lannang community). Overall, the first three findings converge on the suggestion that Lánnang-uè has a high degree of ‘languageness.’ It does not seem to be an ephemeral ad-hoc code-switching phenomenon.

In the next three chapters, I examine the fourth finding. I present further investigations of seven features and/or patterns across three independent levels of languages (i.e., prosody, lexicon, and syntax), all of which I anecdotally observed to exhibit relatively higher amounts of variation compared to other features/patterns:

Prosody

- Lexical tone
- Duration-cued stress
- CV/CVT tone distributional pattern
- CVR-English/CVR-Tagalog tone distributional pattern

Lexicon

- Conjunction distributional pattern
- Prepositions distributional pattern

Syntax

- *Wh*-phrase position distributional pattern

I closely examined these features/patterns because they have the potential to undermine my argument for languageness after a more systematic analysis. I wanted to test whether these features have high degrees of spread and stability and whether the variation is structured. The next chapters are important in that they will either support or weaken my current argument that Lánnang-uè has a high degree of languageness. For example, if I find low degrees of feature/pattern spread (i.e., rare use of a feature/pattern within the community, indicative of low degrees of spread) and high rates of unstructured variation, then my argument that Lánnang-uè is language-like is weakened. However, if I find high degrees of spread and (limited) heterogeneity/variation and find that this variation is structured (Weinreich et al. 1968:187–188), then there is more support for my argument that Lánnang-uè is language-like.

Chapter 4 : Stress and Tone Features

4.1 Introduction

In this chapter, I systematically analyze the prosody of Lánnang-uè. I focus on four prosodic features that I anecdotally found to exhibit more variation compared to other features in the variety – (1) lexical stress, (2) lexical tone, (3) the consonant-vowel-obstruent (CVT/CV) tone distributional pattern, and (4) the consonant-vowel-resonant CVR-English/CVR-Tagalog tone distributional pattern, which I discussed in Chapter 3 and summarize in Section 4.2. The likely high rates of variation I observed suggest that these features are not used by many Lánnang-uè speakers and that these features are not consistently used by them. I was also not able to directly pinpoint conditioning factors for the variation with a simple analysis, suggesting that the prosody of Lánnang-uè has unstructured variation in the domain of prosody. The potential absence of feature spread, stability, and structured variation in Lánnang-uè prosody can weaken my earlier claim that Lánnang-uè has a high degree of languageness, as these features are established hallmarks of languagehood (Weinreich et al. 1968:187–188; Cohen et al. 2021). A lack of linguistic independence can also weaken this claim – the variation observed may be due to the speakers’ high proficiency in Lánnang-uè source languages. The speakers’ knowledge of prosodic structures in Hokkien, Tagalog, English, and/or Mandarin may have influenced Lánnang-uè prosodic structure. If such were the case, the relationship between source language proficiency (high proficiency) and variation could be analyzed as evidence against the hypothesis that Lánnang-uè is highly language-like, as varieties that are highly language-like tend not to be influenced by surrounding languages (Lipski 2020). A systematic investigation of Lánnang-uè’s prosodic features and the variation found within them is needed to establish whether the noticeable variation in prosody poses a challenge to the idea of Lánnang-uè being highly language-like.

I address the gap by formally examining possible spread, stability, structured variation, and linguistic independence in the four prosodic features. First, I assess whether the prosody of

the variety exhibits ‘spread’, i.e., whether its stress and tone features are used by most Lánnang-uè speakers. Second, I examine the stability of the features by looking at the consistency of their use within the individual as well as the consistency of individual patterns of variation between speakers (i.e., interspeaker variation). Third, I attempt to find “language-external” factors that will explain the variation (Ghyselen and De Vogelaer 2018:1), using a sociolinguistic framework (Labov 1972; Eckert 1989). Specifically, I want to test whether the variation in Lánnang-uè prosody is sociolinguistically structured like the variation found in established contact languages, such as Singlish (Starr and Balasubramaniam 2019) and Baba Malay (Lee 2014). Finally, I investigate whether the prosody of Lánnang-uè is influenced by knowledge of (prosody in) its source languages Hokkien, Tagalog, English, and Mandarin.

In this chapter, I address the following research questions:

1. How widespread are the prosodic features within the community?
2. How stable are they? In other words, how consistently do individual speakers use the features? And how similar are their patterns of variation from each other?
3. Is the variation structured? Can a significant part of it be accounted for by sociolinguistic factors and potentially express particular (sets of) “social meaning” (Benor 2010:160)?
4. Are the prosodic features influenced by knowledge of Lánnang-uè’s source languages (and their prosodic features)? Will proficiency in the source languages condition the variation in Lánnang-uè prosody?

The rest of the chapter is dedicated to answering the research questions and fulfilling the objectives. Section 4.2 contextualizes the study by briefly describing the prosodic systems of Lánnang-uè’s source languages (Hokkien, Tagalog, English, and Mandarin); it also briefly summarizes what is known about the prosody of Lánnang-uè, focusing on the four prosodic features investigated in this chapter. Section 4.3 introduces the hypotheses for the highlighted prosodic features. It is followed by Section 4.4, which details the methodology. Sections 4.5 and 4.6 contains the results and the discussion, respectively. Some final remarks and identified limitations conclude this chapter (Section 4.7).

4.2 *The prosody of Lánnang-uè and its source languages: A summary*

The source languages of Lánnang-uè have different prosodic systems. Philippine Hokkien is a lexical tone language (Chappell 2019:181; Tsai 2017:107), or a language where “word meanings or grammatical categories ... are dependent on pitch level” (Crystal 2008:467), while Philippine English and Tagalog are lexical stress languages (Lesho 2018; Schachter and Otones 1972; Hwang et al. 2019), or languages that make lexical contrasts based on syllable prominence (e.g., intensity, fundamental frequency, duration, vowel quality) (Crystal 2008:454). English and Tagalog are non-tone languages (i.e., they do not have lexical tone). Mandarin is a lexical tone language but has been documented to also have stress (Jun 2007; Kuo et al. 2008; Chao 1968; Wang 2015; Duanmu 2007; Chow 2016).

Lánnang-uè has both stress and tone. It has at least four prosodic features relevant to these. These features are (1) lexical stress, (2) lexical tone, (3) the CVT/CV tone distributional pattern, and (4) the CVR-English/CVR-Tagalog tone distributional pattern.

In terms of lexical stress, I found that Lánnang-uè words derived from Tagalog and English tended to have stress, which is marked exclusively using syllable duration, based on my observations. Syllables that are lexically stressed in the variety (e.g., the first syllable of the word *robot* in the phrase *hîge robot heâd* [hi⁵⁵ge³³.ɿo³³bot³³hed⁵⁵] ‘the robot head’) are produced long. Syllables that are unstressed in Lánnang-uè (e.g., the second syllable of the word *robot* in the phrase *hîge robot heâd* [hi⁵⁵ge³³.ɿo³³bot³³hed⁵⁵] ‘the robot head’) are produced short. Speakers occasionally do not follow this stress pattern: for example, syllables that are stressed in the variety are occasionally not produced long.

In terms of lexical tone, I discovered that most syllables/words in Lánnang-uè have lexical/phonemic (henceforth, lexical) tone. Speakers used tone in Hokkien- (e.g., *kaû* [kaw⁵⁵] ‘dog’), Mandarin- (e.g., *siaukhaî* [sjaw²²k^haj⁵⁵] ‘calligraphy paper’), English- (e.g., *shampò* [ʼjam²²pu⁵¹] ‘shampoo’), and Tagalog-derived words (e.g., *basù* [ʼba²²su⁵¹] ‘cup’). However, I noticed that some words did not have this: there were Tagalog-derived words like *basù* [ʼba²²su⁵¹] and Hokkien-derived words like *kaû* [kaw⁵⁵] ‘dog’ that were produced without lexical tone.

In terms of the two tonal distributional patterns, I found that lexical tone assignment in Lánnang-uè words derived from Tagalog and English tended to be conditioned by the source language of the word and syllable structure. Speakers’ CVT syllables generally had high tone

(e.g., English-derived *bôt* [bɔt⁵⁵] ‘robot’) while their CV syllables generally had falling tone (e.g., English-derived *ràw* [ɾɔ⁵¹] ‘raw’) – the CVT/CV tone pattern.⁴⁷ I also observed that CVR syllables in their Tagalog-origin words tended to have high tone (e.g., *sayâng* [sa²²jaŋ⁵⁵] ‘what a waste’) while those in English-origin words tended to have falling tone (e.g., *yoùng* [jaŋ⁵¹] ‘what a waste’) – the CVR-Tagalog/CVR-English tone pattern. Many words followed the distributional patterns, but there were also words that did not. For example, I found words that had high tone for CV and CVR-English syllables (e.g., English-derived *râw* [ɾɔ⁵⁵] ‘raw’, *yoùng* [jaŋ⁵⁵] ‘what a waste’) instead of pattern-conforming falling tone.

4.3 Hypotheses

I have four general hypotheses regarding the features described in Section 4.2, anchored on the possibility that Lánnang-uè is highly language-like:

1. Spread. The four features will be highly widespread within the community.
2. Stability. The features will have high degrees of stability. Speakers who have used these features at all will do so at the individual level with high degrees of consistency. They will have patterns of variation that are highly consistent with each other.
3. Structured variation (systematicity). A significant part of variation will be conditioned by at least age or sex. The tendency not to produce tokens that have the prosodic feature will most likely come from younger speakers and/or female speakers if the variation is innovative.
4. Independence from source languages. The patterns of variation will not be influenced by high language proficiency in Lánnang-uè’s source languages. Speakers who have high

⁴⁷A small number of Tagalog- and English- origin words optionally had rising tone regardless of syllable structure. For example, the final closed syllable of the Lánnang-uè word for ‘pig’, originating from Tagalog, either had the high-III tone (*babôy* [ba³³buj⁵⁵]) or a rising tone (*babóy* [ba³³buj³⁵]); the final open syllable of the Tagalog-origin Lánnang-uè word for ‘duckling’ either had the falling-II tone (*bibè* [bi³³be⁵¹]) or a rising tone (*bibé* [bi³³be³⁵]). The rising tone pattern is not structurally conditioned.

proficiency in them (have knowledge of their prosodic structures) will not be less likely to use Lánnang-uè prosodic structure.

The first two hypotheses on spread and stability (henceforth, Hypotheses 1 & 2) were motivated by previous studies on Lánnang-uè (Gonzales 2018; Gonzales and Starr 2020) as well as many features I described in Chapter 3, where I found evidence of high degrees of spread and stability for many features/patterns. In Gonzales (2018), for example, I presented 65 speakers of Lánnang-uè with different linguistic combinations of derivational affixes and roots (e.g., Tagalog-origin prefix + Hokkien-origin root, Hokkien-origin root + English-origin suffix) in identical carrier sentences and found that most speakers consistently rated words with Tagalog-origin prefixes high and words with English-origin suffixes low (Gonzales 2018). In Chapter 3, I found that all speakers derived personal pronouns from Hokkien with high levels of consistency. Overall, I found evidence of spread and stability in Lánnang-uè, which suggests that Lánnang-uè is very language-like. If this is true, then, the four prosodic features will also have high rates of spread and stability, as languages tend to have high rates of spread and stability across their features. The use of all four prosodic features will be widespread and highly consistent; the variation will only be minimal.

The third hypothesis (henceforth, Hypothesis 3) was motivated by findings in previous variationist studies in Lánnang-uè (Gonzales 2018; Gonzales and Starr 2020), where I found that age and sex conditioned phonological and morphological variation found in Lánnang-uè. In Gonzales (2018), for example, I found that age accounted for a significant part of the variation in the acceptability of (non)conventionalized morphemes. This is similar to what I found in Gonzales and Starr (2020), where I discovered that age and sex conditioned most of the variation in the production of monophthongs. The conditioning effects of these factors on the variation in the use of phonological and morphological features/patterns in Lánnang-uè suggest that Lánnang-uè has high degrees of languageness. If this is indeed the case, then there should be evidence of systematicity or structured variation in its prosodic features/patterns, as languages tend to have systematicity across its features/patterns (Weinreich et al. 1968; Ghyselen and De Vogelaer 2018).

What about the directions of the effects of age and sex on the variation patterns? I rely on previous work and theory to motivate my hypotheses involving them. In the context of sound

change, speakers who are young and those categorized as female have often been reported to exhibit high rates of variation and lead innovative practices (Eckert 1989; Sankoff 2006; Maclagan et al. 1999). In the context of Lánnang-uè, I have found similar effects of age and sex on linguistic behavior (Gonzales and Starr 2020; Gonzales 2018) and interpret the patterns of variation as indicative of a changes-in-progress. In a previous acoustic study of vowel monophthongs (Gonzales and Starr 2020), for example, I found that some young and female speakers tended to deviate from the highly widespread vowel monophthong system of Lánnang-uè. Instead of adopting the system where all vowels are realized similarly irrespective of source language, these speakers modified this system by producing certain vowels differently depending on the source language. For example, they produced the vowel [o] lower (higher F1 formant frequency) in Tagalog-sourced words compared to English- and Hokkien-sourced words, suggesting a potential change-in-progress. Given the documented sociolinguistic patterns, it was reasonable to hypothesize that a sizable portion of the tokens that do not conform to the four prosodic conventions/patterns described in Section 4.2 (e.g., unstressed syllables for syllables that are supposed to be stressed) will come from young and female speakers, assuming that the patterns of variation reflect innovation (i.e., ongoing change).

However, not all patterns of variation mark innovation. There may be other reasons for the variation (e.g., variation being embedded into the linguistic system). I kept this possibility in mind when I formulated my hypotheses for age and sex. This meant that I was also open to the idea that a significant part of non-conforming tokens (i.e., tokens that do not reflect the use of the four prosodic features) will not be associated with female or young speakers.

The final hypothesis (henceforth, Hypothesis 4) was motivated by my observations of varieties characterized as ‘languages’ or, in my terms, ‘highly language-like’ – these varieties tend to have structural patterns that are not influenced by the patterns of other languages. For example, the structures of the Topo and Ugsha varieties of Media Lengua – varieties with Spanish- and Quichua-derived elements characterized by Lipski (2020) as stable and language-like – were reported to be independent of (the Media Lengua speakers’ knowledge of) Spanish structure. Lipski (2020) found that Media Lengua speakers in Topo and Ugsha tended to adhere to Media Lengua grammar regardless of their proficiency in Spanish. That is, Spanish language proficiency did not influence the structure of the Topo and Ugsha Media Lengua. If Lánnang-uè has high degrees of languageness like these two varieties of Media Lengua, then I expect the

variation in Lánnang-uè prosodic structure not to be greatly influenced by high proficiency in its source languages. For example, speakers will continue to use stress in Lánnang-uè regardless of their proficiency in Hokkien (a language without stress). Lánnang-uè tokens that do not conform to Lánnang-uè stress patterns will not come from speakers who are highly proficient in Hokkien.

4.4 Methodology

4.4.1 Task

To test my hypotheses, participants of different sociolinguistic backgrounds performed a self-recorded reading task. They were first given a list of 165 words in Lánnang-uè (Appendix C). Then, they were asked to familiarize themselves with the orthography of the words in the list. Once this was done, the participants were asked to turn on the recording device and produce each word in the list in the Lánnang-uè carrier sentence: *Hîgé sî _ bá?* ‘Is that ___?’. For instance, when participants encountered the stimulus ‘hotdog’ on the list, they were expected to respond by saying *Hîgé sî hotdôg bá?* ‘Is that a hotdog?’. Participants were asked to embed the words in the middle of the Lánnang-uè carrier sentence instead of reciting the words in isolation to avoid potential effects of intonation. The participants stopped the recording and submitted the audio file once they have produced all the words (in their carrier sentences) in the list. Upon submission, I conducted 20-minute interviews with my participants. In the first part of the interview, I asked for the participants’ age and sex. I also asked them to rate their proficiency in English, Tagalog, Hokkien, and Mandarin using a 7-point Likert scale. The second part of the interview focused on questions regarding their community, identity, language, and education (Appendix D). The interviews were conducted to collect social data that is needed for the examination of the relationship between social factors and linguistic behavior.

4.4.2 Stimuli

The experimental stimuli were given to the participants in three blocks. Each block featured the same 55 unique words or, in statistical jargon, ‘items’ (in the same carrier sentence) that were distributed across 17 conditions. These conditions involved the linguistic origin of the words, expected stress pattern in non-final and final syllables (acquired from my analysis of the lexicon of Lánnang-uè), and final syllable structure (Table 19). The words under each of the 17 conditions were all disyllabic, to simplify the analysis. Words sourced from Mandarin were

included in the experiment even if Mandarin-origin words are virtually non-existent in the basic vocabulary of Lánnang-uè, as speakers of Lánnang-uè occasionally use ‘technical’ Mandarin-sourced words. To minimize the possibility of failure to identify words, I ensured that all words in the list are common and culturally relevant by asking two native speakers of Lánnang-uè to double check the list of words.

Although each block contained the same set of 55 items, the order in which these items are given to the participants was randomized differently for each block. I exposed the participants to three blocks instead of one to ensure measurement reliability within the speaker (intra-speaker reliability). Overall, every participant went through 165 trials (55 items X 3 blocks) (Appendix C).

Table 19. Distribution of stimuli (prosody)

Condition	# of items	Source language of word	Expected stress pattern		Final syllable structure
			<i>non-final syllable</i>	<i>final syllable</i>	
1	3	Mandarin	unstressed	stressed	CV
2	3	Mandarin	unstressed	stressed	CVR
3	3	Hokkien	NA	NA	CV
4	3	Hokkien	NA	NA	CVR
5	3	Hokkien	NA	NA	CVT
6	3	Tagalog	stressed	unstressed	CV
7	3	Tagalog	stressed	unstressed	CVR
8	3	Tagalog	stressed	unstressed	CVT
9	3	Tagalog	unstressed	stressed	CV
10	3	Tagalog	unstressed	stressed	CVR
11	3	Tagalog	unstressed	stressed	CVT
12	3	English	stressed	unstressed	CV
13	5	English	stressed	unstressed	CVR
14	3	English	stressed	unstressed	CVT
15	3	English	unstressed	stressed	CV
16	5	English	unstressed	stressed	CVR
17	3	English	unstressed	stressed	CVT

4.4.3 Datasets and preparation

The relevant parts of the recordings (i.e., the tokens of interest in the carrier sentences) were manually segmented by word and by syllable using Praat (Boersma and Weenink 2021). I formed three datasets from the segmented data: (1) a dataset containing the words of interest,

used to test my hypotheses on lexical tone in Lánnang-uè words, (2) a dataset containing syllables from Tagalog- and English-origin words, used to test my hypotheses on conformance to stress patterns, and (3) a subset of the previous syllable-level dataset that only contained syllables in word-final position (syllables that have not been subject to tone sandhi),⁴⁸ to test my hypotheses on the adherence to the tone patterns.

For the first dataset, I asked three coders who speak Lánnang-uè and have basic linguistics training to code each word syllable for lexical tone – ‘1’ if they perceived the word to have lexical tone and ‘0’ if they did not. I coded the data for source language. Words derived from Tagalog, for example, were coded as ‘Tagalog’.

For the second dataset, I extracted the duration of each syllable, in seconds, using a Praat script (Styler 2011). Since duration was found to cue stress exclusively in Lánnang-uè (Chapter 3), I interpret duration as ‘degree of stress’. In addition to a continuous duration/stress variable, I also coded the data categorically with the help of one coder from the pool of three mentioned earlier. The coder, looking at the syllable in the context of the disyllabic word, coded the syllable as ‘stressed/long’ if the syllable was perceived to be longer than the other syllable in the word. They coded the syllable as ‘unstressed/short’ if the syllable was shorter than the other syllable. The coding was done binarily – no middle-ground option was given, and my coder had to classify the syllables in one of two categories. I reviewed the coder’s work to ensure that they coded the data accurately. I chose to include both continuous and categorical stress variables in my dataset because the continuous variable is useful for analyses that perform better with finer-grained data (e.g., regression) whereas the categorical variable is useful for analyses that need frequencies (i.e., spread, stability) (Chapter 4.4.5). For this dataset, I included information regarding the expected stress pattern for each syllable (i.e., ‘expected to be stressed’ vs. ‘expected to be unstressed’).

For the third dataset, I removed all tokens that did not have lexical tone. I then extracted the pitch slope⁴⁹ of each syllable using a Praat script (Styler 2011). Because I filtered out

⁴⁸ Tone sandhi in Lánnang-uè is right-headed and operates within the phrase (in this case, within the noun phrase or word-internally). For instance, in a disyllabic word with lexical tone, the first syllable conditions the phonetic realization of the second syllable within the tonal phrase (Chapter 3.3.3.1).

⁴⁹ I use the term ‘slope’ in ‘pitch slope’ in the mathematical sense, where ‘slope’ refers to the ratio of the vertical change between *two* (pitch) points to the horizontal change between the same two points. In other words, pitch slope measures how steep a pitch is. A slope of zero (or near zero) means a level pitch (there is no change from the starting pitch to the ending pitch). A negative slope indicates a falling pitch while a positive slope is indicative of a

toneless tokens, the pitch slopes of this dataset were interpreted as lexical tones.⁵⁰ A pitch slope significantly below zero (negative slope) is interpreted as a falling (contour) tone whereas a slope close to zero is interpreted as a high (level) tone. I also asked three coders to code the data categorically. I asked them to code the syllable as ‘high’ if they perceived the pitch in the syllable as having high tone and ‘falling’ if they perceived it as having falling tone. I reviewed the coders’ work to ensure that they coded the data accurately. I decided to include both continuous and categorical tone variables in my dataset for the same reasons mentioned earlier for the stress variables. I included the information on syllable structure and source language in the third dataset.

Then, for all three datasets, I coded each token for ‘presence of feature/conformance to pattern’:

- In the first dataset,
 - words that coders perceived to have lexical tone were marked as ‘1’ (lexical tone present);⁵¹
 - words that coders perceived not to have lexical tone were marked ‘0’ (lexical tone absent).
- In the second dataset,
 - syllables that were coded ‘stressed/long’ and ‘expected to be stressed’ were marked ‘1’ (conformed to stress pattern);
 - syllables that were coded ‘unstressed/short’ and ‘expected to be unstressed’ were marked ‘1’ (conformed to stress pattern);
 - the rest of the syllables were coded ‘0’ (did not conform to stress pattern).
- In the third dataset,
 - syllables that were coded ‘high’ and ‘CVT’ were marked ‘1’ (conformed to tone patterns);

rising pitch. For this chapter, only the first two (i.e., zero, negative pitch) is relevant, as they directly correspond to the high (level) and falling (contour) tones under investigation in this study.

⁵⁰ Throughout the chapter, I distinguish between the use of the term ‘tone’ and ‘pitch slope’. I refer to ‘tone’ when I am referring to the phonemic feature; I use the latter term to refer to a phonetic feature. This distinction is relevant for this chapter: I extracted acoustic/phonetic features (e.g., pitch slope) from the tokens, which I interpreted as being phonemic or having lexical tone.

⁵¹ The assumption here is if one syllable has lexical tone, then the other should also have lexical tone.

- syllables that were coded ‘falling’ and ‘CV’ were marked ‘1’ (conformed to tone patterns);
- syllables that were coded ‘high’, ‘CVR’, and ‘Tagalog’ were marked ‘1’ (conformed to tone patterns);
- syllables that were coded ‘falling’, ‘CVR’, and ‘English’ were marked as ‘1’ (conformed to tone patterns);
- all other syllables were coded as ‘0’ (did not conform to tone patterns).

4.4.4 *Analytical method*

4.4.4.1 Descriptive analyses

For each of the coded datasets, I conducted descriptive analyses. I measured the degree of feature/pattern adoption within the community (henceforth, ‘spread’) and consistency of the prosodic features of interest by examining the factor ‘presence of feature/conformance to pattern’ (see Section 4.4.3). I used three measures (see Section 4.4.5). Furthermore, I conducted analyses of tokens that did not exhibit the use of the four prosodic features/patterns by providing a breakdown of these tokens.

4.4.4.2 Regression analyses

It is useful to examine frequencies, proportions, and variability in raw data, but only to a certain extent, as this approach does not allow us to isolate the conditioning effects of a particular factor on linguistic behavior. So, in addition to this, I conducted regression analyses to aid in my analysis of spread, stability, and variation. This approach, unlike the first, allows us to single out the (main or interaction) effects of particular factors and test for correlations between these factors and the dependent variables.

To test my hypotheses on stress, I fitted a linear mixed-effects regression model on the second dataset using R (R Core Team 2015; Kuznetsova et al. 2019). The dependent variable was syllable duration (continuous), which I interpret as degree of stress. Expected stress, sex, age, and z-scored language proficiency in languages with stress (i.e., Tagalog, English,

Mandarin)⁵² and z-scored proficiency in languages without stress (i.e., Hokkien) – collected during the interviews (Section 4.4.1) – were included in the model as fixed effects. Because Lánnang-uè has phrase-final lengthening (i.e., the final syllable of the final word in a phrase is always long) (Chapter 3.3.3.3), and this feature conditions duration, I included syllable position as a covariate. This allowed me to test for the effects of expected stress on syllable duration (actual stress) while controlling for phrase-final lengthening. Position (**final** vs. non-final), expected stress (**stressed** vs. unstressed), sex (**female** vs. male), and age (younger vs. **older**) were coded categorically. The reference level, or the level to which the other level is compared, is indicated in boldface. Interaction factors involving stress and the sociolinguistic factors (i.e., sex, age, proficiency in the stress languages, proficiency in the Hokkien) and position were included to test whether these factors condition the variation in adherence to the stress pattern. Random intercepts for participant and item were included in this model.

To test my hypotheses on lexical tone, I fitted a generalized linear mixed-effects model with a logistic link function on the first dataset in the R environment. The dependent variable was ‘lexical tone presence’ (categorical), i.e., whether the word has lexical tone. Source language, sex, age, z-scored language proficiency in languages with lexical tone (i.e., Hokkien and Mandarin) and z-scored proficiency in languages without lexical tone (i.e., Tagalog and English)⁵³ – collected during the interviews (Section 4.4.1) – were included in the model as fixed effects. Sex (**female** vs. male) and age (younger vs. **older**) were coded categorically, with the reference level indicated in boldface. Interaction factors were not included. Random intercepts for participant were included in this model.

To test my hypotheses on the two tone distribution patterns, I fitted a linear mixed-effects regression model on the third dataset, where the dependent variable was pitch slope (continuous), interpreted as use of high/falling lexical tone. Source language, syllable structure, sex, age, and z-scored language proficiency in languages without the tone patterns⁵⁴ were included in the

⁵² I created the factor ‘proficiency in languages with stress’ by running a Principal Components Analysis (PCA) on the z-scored Tagalog, English, and Mandarin proficiency scores and getting the component that is positively correlated with the three scores.

⁵³ I created the factor ‘proficiency in languages with tone’ by running a PCA on the z-scored Hokkien and Mandarin proficiency scores and getting the component that is positively correlated with both scores. I created the factor ‘proficiency in languages without tone’ by running PCA on z-scored Tagalog and English proficiency scores and getting the component that is positively correlated with both scores.

⁵⁴ I created the factor ‘proficiency languages without the tone patterns’ by running a PCA on the z-scored Tagalog, English, Hokkien, and Mandarin proficiency scores and getting the component that is positively correlated with the four scores.

model as fixed effects. Source language (Tagalog vs. **English**), syllable structure (i.e., CVT vs. **CV** and CVR vs. **non-CVR**), sex (**female** vs. male), age (younger vs. **older**) were coded categorically, with the reference level indicated in boldface. Two-level interaction effects between syllable structure (CVT vs. CV) and the sociolinguistic factors (i.e., sex, age, proficiency in the languages without the tone patterns) were included to test whether these factors condition the variation in the adoption of the CVT/CV pattern. Three-level interaction factors between source language (Tagalog vs. English), syllable structure (CVR vs. non-CVR), and the sociolinguistic factors were included to test whether these factors condition the variation in the use of the CVR-Tagalog/CVR-English pattern. Random intercepts for participant and item were included in this model.

In all my regression models, the categorical predictor variables were analyzed after (re)coding the variables using unweighted effect contrast coding conventions (i.e., 1 vs. -1) (Sonderegger 2022).

4.4.4.3 Criteria for hypothesis testing

My hypotheses on spread and stability (Hypotheses 1 and 2) will be supported if I find evidence of them in my data. Regarding my descriptive analyses, if the feature spread scores are above average (i.e., 0.5, or more than half of the population) (see Section 4.4.5), then my hypotheses on spread will be supported. If the features I hypothesized to have high stability have mean intraspeaker feature consistency scores (as measured in Section 4.4.5) that are higher than 0.5 (i.e., the features were used more than 50% of the time, on average), then my hypothesis on stability will be supported. It will be further supported if I find interspeaker pattern inconsistency scores that are below 0.5 (i.e., the patterns of variation among speakers have heterogeneity levels below 50%) (Section 4.4.5).

On my regression analyses of syllable duration/stress and pitch slope/tone, if there is statistically significant evidence for the effects of structural factors (i.e., expected stress pattern, syllable structure, syllable structure and source language) on the dependent variables and if the effects are in the expected direction (e.g., longer/stressed syllables associated with syllables that are expected to be stressed), then my hypotheses (Hypotheses 1 and 2) will be supported as well. I interpret these statistically significant effects, if present, as evidence of both spread and stability. In regression models, the effect of a specified predictor variable becomes statistically

significant if there are consistent correlations between the specified predictor (e.g., structural factors) and the dependent variable for many participants.

My hypotheses on structured variation (Hypothesis 3) (i.e., age and sex conditioning much of the variation) will be supported if I find interaction effects between the hypothesized sociolinguistic variables and the structural variables on the dependent variable in my regression models. I did not directly model ‘variation’; instead, I interpreted the interaction effects as the potential conditioning effect of a sociolinguistic variable on a prosodic feature/pattern (e.g., the potential conditioning effect of age on the relationship between pitch slope/tone and syllable structure). My hypotheses on the direction of the effect (e.g., younger speakers tending to produce tokens without lexical tone) will be supported if I find the expected pattern in an examination of the marginal effects – defined as “predictions generated by a model when one holds the non-focal variables constant and varies the focal variable(s)” (Lüdecke 2018a:1; Lüdecke 2018b) or the effect the individual predictors have on the dependent variable while all other variables are held constant. I used the `ggeffects` package to compute the estimated marginal means (predicted values) for the dependent variable at the margin of specific values or levels from certain model terms (Lüdecke 2018b).

My hypotheses on linguistic independence (Hypothesis 4) will not be supported if I find evidence of negative correlations between the proficiency variables and the structural variables on the dependent variable in my regression models (i.e., high proficiency linked to use of pattern-non-conforming features).

4.4.5 *Measuring spread and stability*

To examine the degree of spread and stability of the prosodic features (and patterns of variation involving these features), I relied on three measures:

1. *Feature spread* – What proportion of my speakers uses the feature at all?
2. *Mean intraspeaker consistency* – How often/consistently do individual speakers use the feature?
3. *Interspeaker pattern inconsistency* – How inconsistent are the patterns of variation between speakers?

I measured feature spread, here operationalized as the degree of feature/variant adoption in my sample, by dividing the number of speakers who produced the feature/variant at least once by the number of all speakers. A high spread score intuitively means that the feature/variant is widespread whereas a low score indicates that the feature/variant is limited to particular (groups of) speakers. A score of zero means that the feature/variant is not used.

Feature spread score = *number of speakers who used the feature at least once / number of all speakers*

I acknowledge that this measure – in isolation – is weak: a single occurrence for a given speaker is not sufficient evidence to conclude that the feature has spread to that speaker. For one, there is the possibility of accident. I am nevertheless using this measure in this chapter (and in the rest of the dissertation) because even weak evidence of spread is relevant to the question of Lánng-uè's languageness.

To quantify stability, I looked at mean/general consistency in the use of the prosodic features within individuals (henceforth, mean intraspeaker consistency) as well as the inconsistency of patterns of variation between speakers (henceforth, interspeaker pattern inconsistency).

I measured mean intraspeaker consistency by first finding the consistency scores for each individual who used that feature (or conformed to the pattern) at least once:⁵⁵ for each individual, I divided the number of tokens where they used the feature or followed the pattern by the number of tokens where they could have done so. After finding the individual scores, I averaged them to find the mean/general intraspeaker consistency score. Intuitively, a high score indicates that the feature/pattern is consistently used/followed within individuals. A low score indicates that speakers rarely use that feature or pattern.

⁵⁵ I only included speakers who have the feature/variant to allow me to distinguish between 'spread' and 'stability' more clearly. By defining 'stability' this way, I avoided confounding both concepts – that is, I allowed for a scenario where only some speakers use the feature/variant (low rates of spread) but consistently do so (high stability).

Individual intraspeaker consistency score = *number of tokens where an individual speaker uses a particular feature / number of tokens where the feature could be used for that individual*

Mean intraspeaker consistency score = *individual intraspeaker consistency scores / number of individuals*

I measured interspeaker pattern inconsistency by finding the mean-normalized standard deviation of the individual patterns of variation (i.e., the individual intraspeaker feature consistency scores mentioned earlier) (Kahn et al. 2011; Shetewi 2018; Verhagen et al. 2020). The standard deviation is hard to interpret on its own, so I normalized it by the mean, deriving the “coefficient of variation” (Pélabon et al. 2020:180). The coefficient of variation can be interpreted as the degree of inconsistency in the patterns of variation between speakers in a percentage-like scale similar to the scale used for the mean intraspeaker feature consistency score. I henceforth refer to this measure as the interspeaker pattern inconsistency score. A low score (i.e., close to zero) indicates more homogeneity between speakers who have the feature/variant (e.g., almost all speakers use the feature 80% of the time). A high score indicates more heterogeneity (e.g., some speakers use the feature 30% of the time, some use it 50% of the time, others, 80% of the time).

Interspeaker pattern inconsistency score = *standard deviation of all individual intraspeaker feature consistency scores / mean of these scores*

4.4.6 Participants

There were 20 participants, all of whom were born and raised in the Philippines, spoke Lánnanguè, and had at least some knowledge of Tagalog, English, Mandarin, and Hokkien. All of these participants were recruited via social media or word of mouth. Out of the 20 participants, 11 self-reported as female while nine self-reported as male. With regard to age, twelve participants (i.e., ‘younger speakers’) fall within the 20 to 33 age range; eight participants (i.e., ‘older speakers’) fall within the 37 to 63 age range. The average z-scored proficiency scores of my 20 participants are the following: Tagalog (0.18), English (0.74), Hokkien (-0.15), and Mandarin (-0.05). A

negative score indicates low proficiency, a positive score indicates high proficiency, whereas a score close to zero indicates average proficiency.

Nine of the participants in this study were participants in my *wh*-question study in Chapter 6; twelve contributed to the lexicon study in Chapter 5. Eight were present in all three variationist studies in this dissertation.

4.5 Results

4.5.1 Lexical stress

4.5.1.1 Descriptive analysis

All of my speakers followed the lexical stress pattern at least once (feature spread score = 1): they produced at least one syllable that is expected to be stressed as stressed/long⁵⁶ (mean = 0.30 seconds, SD = 0.07) and produced at least one syllable that is expected to be unstressed as unstressed/short (mean = 0.24 seconds, SD = 0.07).⁵⁷

Out of the 20 speakers, five (25%) followed the stress pattern 60 to 70% of the time (mean intraspeaker consistency score = 0.683, SD = 0.008); fourteen (70%) followed the pattern 70 to 89% of the time (mean intraspeaker consistency score = 0.753, SD = 0.27). Only one, 60-year-old male, followed the pattern more than 90% of the time (individual intraspeaker consistency score = 0.904). Overall, all 20 speakers occasionally varied in their adherence to the stress pattern (none of the individual consistency scores were 1) (Figure 5).

⁵⁶ In a supplementary investigation using the data of this chapter, I verified that lexical stress is cued by duration exclusively. Two native speakers (not part of the participant pool) and I perceived the participants to exclusively use duration to mark lexical stress. I also fitted a supplementary regression model on the stress data to verify this quantitatively (Appendix E). In this model, duration, fundamental frequency (mean F0, initial F0, F0 pitch), overall intensity (mean intensity), and vowel quality (distilled F1 and F2 formant frequencies using PCA) – acoustic correlates of stress (Gordon and Roettger 2017) – were used to predict perceived stress (‘stressed’ vs. ‘unstressed’, coded by a trained coder who was not explicitly told to pay attention to duration when they classified syllables for stress). In the same model, I also included the sociolinguistic variables of interest in this chapter (i.e., speaker age, sex, proficiency in languages with stress and proficiency in without stress) as well as interactions between these factors and the acoustic correlates of stress, as covariates to account for potential effects of these sociolinguistic variables on the hypothesized links between the correlates and perceived stress. Random intercepts for participant and item were included to account for individual (speaker) variation. My results confirm the preliminary perception results and observations in preliminary work, indicating that perceived lexical stress is exclusively correlated with (cued by) duration ($\beta = 3.27, p < 0.001$) (Appendix E). Overall intensity, fundamental frequency, and vowel quality cannot reliably predict perceived lexical stress in Lánnang-uè words derived from Tagalog and English.

⁵⁷ The duration values are averaged over individual syllables and not by participant.

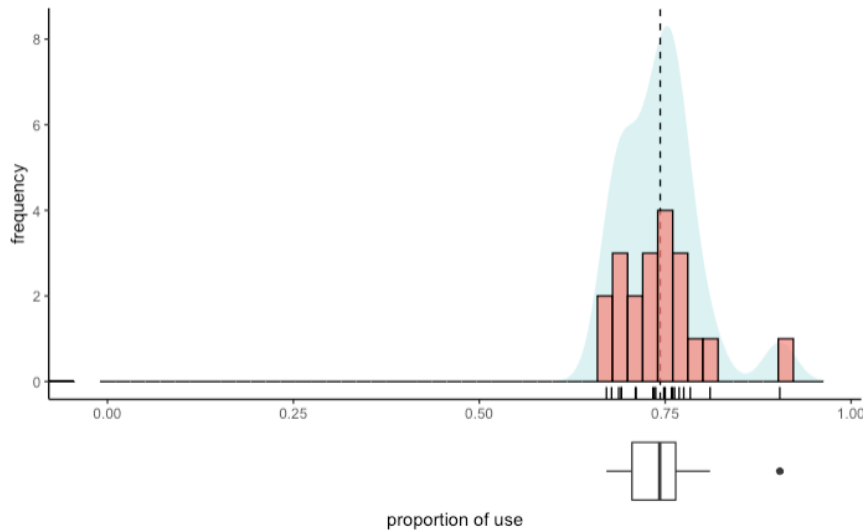


Figure 5. Histogram (frequency), density, and box plot: individual intraspeaker consistency scores (conformance to the stress pattern); broken line indicates mean

The mean intraspeaker consistency score for *all* participants is 0.743 (SD = 0.05). The interspeaker pattern inconsistency score is 0.0725. Overall, the stress pattern is highly stable: individually, speakers consistently stress/lengthen syllables that are stressed and unstress/shorten syllables that are unstressed (mean individual intraspeaker feature consistency score above 0.5). The patterns of variation between speakers were also highly homogenous (inconsistency below 0.5).

The total number of tokens that did not conform to the stress pattern is 1,229 (out of 4,786). Out of these tokens, 664 (54.03%) were unstressed but were produced as long/stressed syllables. The top five words that had the largest proportion of non-conforming tokens in this category are:

- | | | | | |
|----|-----------------|--|-------------------------------------|-----------|
| 1. | <i>bāllpên</i> | [³³ bol ⁵¹ pen ⁵¹] | ‘ballpoint pen’ | 44 tokens |
| 2. | <i>isâw</i> | [²² ʔi ⁵⁵ saw ⁵⁵] | ‘grilled chicken or pork intestine’ | 42 tokens |
| 3. | <i>basebàll</i> | [²² bejs ⁵¹ bol ⁵¹] | ‘baseball’ | 41 tokens |
| 4. | <i>sāndàl</i> | [³³ san ⁵⁵ dal ⁵⁵] | ‘sandal’ | 39 tokens |
| 5. | <i>pancâke</i> | [²² pan ⁵⁵ kejk ⁵⁵] | ‘pancake’ | 39 tokens |

The syllables in bold were frequently stressed or produced as long syllables even if they are unstressed.

Out of the 1,229 tokens that did not conform to the stress pattern, 565 (45.97%) were expected to be stressed but were produced as short/unstressed syllables. The top five words that had the largest proportion of tokens in this category are:

1. <i>isâw</i>	['ʔi ²² saw ⁵⁵]	'grilled chicken or pork intestine'	44 tokens
2. <i>hotdôg</i>	['hat ²² dog ⁵⁵]	'hotdog'	38 tokens
3. <i>lāsèr</i>	['lej ³³ sɿ ⁵¹]	'laser'	34 tokens
4. <i>bikò</i>	['bi ²² ko ⁵¹]	'sticky rice cake'	31 tokens
5. <i>lūgâw</i>	['lu ³³ gaw ⁵⁵]	'rice porridge'	31 tokens

The syllables in bold were frequently unstressed or produced as short syllables even if they are expected to be stressed.

The significant rates of instability (25.7% of the total 4,786 tokens) are partially due to Lánngang-uè's phrase-final lengthening/non-phrase-final shortening feature (Chapter 3.3.3.3). Some unstressed syllables in the phrase-final (word-final) position were long; some stressed syllables in the phrase-final position were short. I am unable to account for this feature using descriptive analyses. I accounted for it in my regression modelling in Section 4.5.1.2 by including syllable position as well as its interaction with expected stress in the model.

I am aware that there are phonetic factors that condition duration/stress, such as vowel quality, syllable structure, segmental make-up (e.g., presence of sonorants, diphthongs), and word structure (e.g., compound word vs. mono-morphemes). However, in the interest of simplifying my analysis, I did not include them in my regression model in the following section. I also do not discuss them further in this chapter.

4.5.1.2 Regression analysis

My model of syllable duration, summarized in Table 20, showed main effects of position, expected stress, and proficiency in Hokkien. They furthermore revealed interaction effects between (1) expected stress and position, (2) expected stress and age, (3) expected stress and sex, and (4) expected stress and Hokkien proficiency. There were no main effects of age, sex, and

proficiency in languages with stress. Interaction effects between expected stress and proficiency in languages with stress on syllable duration were not significant.

Table 20. Linear regression results for syllable duration (observations = 4,786, conditional $R^2 = 0.694$, random intercepts for participant and item)

Predictors	Estimates	SE	CI	<i>p</i>
(Intercept)	0.3168	0.018	0.2816 – 0.3521	< 0.001
position (final vs. non-final)	-0.0579	0.0063	-0.0702 – -0.0456	< 0.001
expected stress (stressed vs. unstressed)	-0.0248	0.0067	-0.0380 – -0.0116	< 0.001
age (younger vs. older)	0.025	0.0195	-0.0133 – 0.0633	0.2
sex (male vs. female)	0.0033	0.0177	-0.0313 – 0.0380	0.851
proficiency (Hokkien)	0.0296	0.0136	0.0030 – 0.0563	0.029
proficiency (languages with stress)	0.0014	0.0009	-0.0004 – 0.0031	0.127
position : expected stress	-0.0523	0.0123	-0.0764 – -0.0283	< 0.001
expected stress: age	-0.0082	0.0031	-0.0144 – -0.0021	0.009
expected stress: sex	-0.0071	0.0029	-0.0127 – -0.0015	0.013
expected stress: proficiency (Hokkien)	-0.0099	0.0023	-0.0143 – -0.0055	< 0.001
expected stress: proficiency (languages with stress)	-0.0018	0.0011	-0.0041 – 0.0004	0.116

An examination of the marginal means revealed the direction of the conditioning effect of expected stress on duration – syllables that were expected to be stressed were consistently associated with longer syllables (i.e., they were stressed) while ones that were expected to be unstressed were associated with shorter syllables (i.e., they were unstressed) (see Figure 38 in Appendix F for individual speaker patterns).

An examination of the marginal means of the interaction terms shows that a sizable number of tokens that did not follow the expected stress patterns, as indicated by more overlap in the marginal means between expected stress contexts in Figure 6, were associated with speakers who are old, female, and not proficient in Hokkien. Many of the syllables, for example, that were expected to be stressed but were unstressed or produced as short syllables came from these speakers. My model predicts that speakers will be less likely to conform to the expected stress patterns in Lánnang-uè if they are old, female, and not proficient in Hokkien.

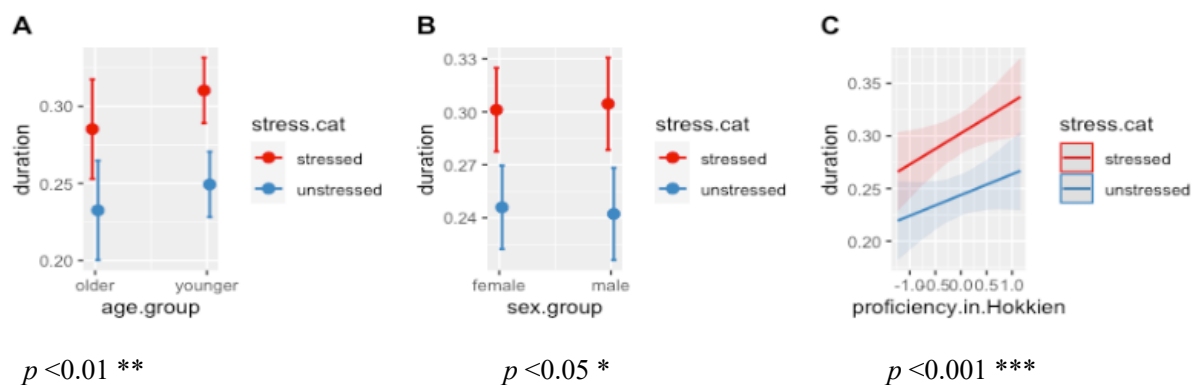


Figure 6. Marginal means/effects of age, sex, and proficiency on duration and expected stress

Focusing on results that are relevant to my hypotheses, I found that speakers produced long syllables (stressed syllables) for syllables expected to be stressed and short syllables (unstressed syllables) for ones expected to be unstressed with high consistency. They suggest that the stress pattern has high degrees of spread and stability (relevant to Hypotheses 1 and 2). I also found that the variation in the adherence to the stress pattern was conditioned by age (older speakers), sex (female speakers), and Hokkien proficiency (low proficiency) (relevant to Hypotheses 3 and 4).

4.5.2 Lexical tone

4.5.2.1 Descriptive analysis

All 20 speakers in my sample had lexical tone in at least one Lánnang-uè word they produced during the experiment (feature spread score = 1). The results indicate that the use of lexical tone is highly widespread.

Out of these speakers, two female speakers aged 33 and 54 had lexical tone in all their words (mean intraspeaker consistency score = 1) and 16 almost always had lexical tone in their production of Lánnang-uè words (mean intraspeaker consistency score = 0.985). Two speakers, one 24-year-old male and one 32-year-old female, had significantly more variation than the rest, with a mean intraspeaker consistency score of 0.92 (Figure 7). Overall, 18 speakers occasionally varied in their use of lexical tone.

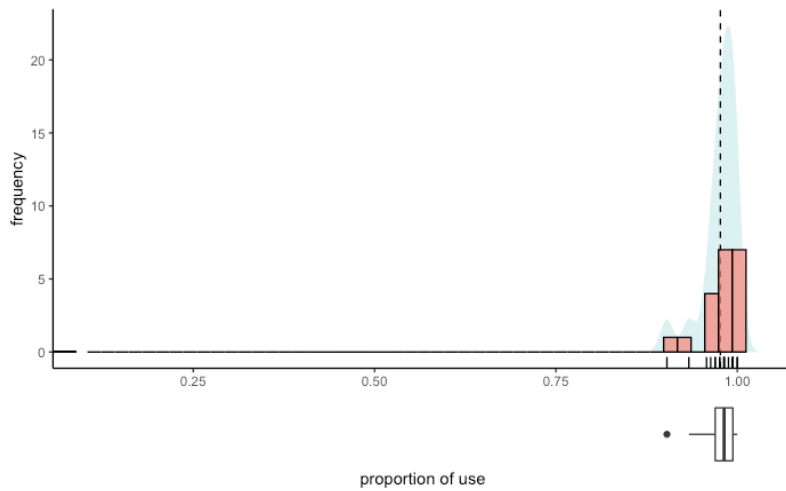


Figure 7. Histogram (frequency), density, and box plot: individual intraspeaker consistency scores (lexical tone in Lánnang-uè words); broken line indicates mean

The mean intraspeaker consistency score for *all* participants is 0.9766 (SD = 0.02).

The interspeaker pattern inconsistency score is 0.0244. Overall, the use of lexical tone is highly stable (instability below 0.5). Speakers who varied in their use of lexical tone at least once only varied occasionally (mean intraspeaker feature consistency score above 0.5) and had patterns of variation that are similar with other speakers (interspeaker pattern inconsistency below 0.5).

A breakdown of the tokens that the 18 speakers did not produce with lexical tone – i.e., 77 (2.34%) out of the 3294 words in the complete dataset – is shown in Table 21. For example, in the first row of the table, I report that 12 (15.58%) of the total words that did not have lexical tone were derived from English. The tokens are *halò* ‘halo’, *hotdôg* ‘hotdog’, and *Beijîng* ‘Beijing’.

Table 21. Frequency of Lánnang-uè words that did not have lexical tone, by source language

Source language	Token (in Lánnang Orthography)	Participant Pronunciation (IPA)	Freq.	Total	% of total variation
English	<i>halò</i> ‘halo’	[hej ⁵⁵ loʔ ²¹]	5	12	15.58%
	<i>hotdôg</i> ‘hotdog’	[hat ⁵⁵ dog ²¹]	1		
	<i>Beijing</i> ‘Beijing’ ⁵⁸	[bej ¹¹ dʒɪŋ ⁵¹]	6		
Tagalog	<i>atây</i> ‘liver’	[ʔa ¹¹ taj ⁵¹]	3	15	19.48%
	<i>putò</i> ‘rice cake’	[pu ⁵⁵ to ²¹]	1		
	<i>siōmaí</i> ‘shomai dimsum’	[ʃo ⁵⁵ maj ²¹]	4		
	<i>siōpaó</i> ‘pork bun’	[ʃo ⁵⁵ paw ²¹]	6		
	<i>turôn</i> ‘banana fritter’	[tu ¹¹ ʔon ⁵¹]	1		
Hokkien	<i>ditsáp</i> ‘twenty’	[di ⁵⁵ tsap ²¹]	2	23	29.87%
	<i>pantôh</i> ‘banquet’	[pan ⁵⁵ toʔ ²¹]	5		
	<i>siēnsī</i> ‘teacher’	[sien ⁵⁵ si ²¹]	1		
	<i>taugé</i> ‘bean sprout’	[taw ⁵⁵ ge ²¹]	7		
	<i>tshiathaú</i> ‘driver’	[tʃa ⁵⁵ t ^h aw ²¹]	8		
Mandarin	<i>shibā</i> ‘eighteen’	[ʃə ⁵⁵ pa ²¹]	8	27	35.06%
	<i>shiqī</i> ‘seventeen’	[ʃə ⁵⁵ tʃi ²¹]	4		
	<i>shisān</i> ‘thirteen’	[ʃə ⁵⁵ san ²¹]	4		
	<i>toúfa</i> ‘hair’	[t ^h ow ⁵⁵ fa ²¹]	5		
	<i>xuexiào</i> ‘school’	[sje ⁵⁵ sjaw ²¹]	3		
	<i>yanjīng</i> ‘eye’	[jen ⁵⁵ tʃɪŋ ²¹]	3		
Total			77		100%

Words without lexical tone that were derived from Tagalog and English had phonetic (non-phonemic) pitch – they sounded just like how they would in the source languages (e.g., Tagalog-sounding *atay* [ʔa¹¹taj⁵¹] ‘liver’ was produced instead of *atây* [ʔa³³taj⁵⁵]). Those derived from Mandarin and Hokkien had Tagalog- and English-like tone (e.g., Tagalog- and English-sounding *shiqi* [ʃə⁵⁵tʃi²¹] ‘seventeen’ was produced instead of word *shiqī* [ʃə³⁵tʃi⁵⁵]).

4.5.2.2 Regression analysis

The results showed a main effect of self-reported proficiency (henceforth, proficiency) in Tagalog and English on likelihood to use lexical tone in a Lánnang-uè word. Age, sex, and

⁵⁸ This word was treated as a Mandarin borrowing in English. Segmentally, participants produced it like English (i.e., [bej dʒɪŋ]) rather than like Mandarin [pej tɛɪŋ]).

proficiency in languages with tone (i.e., Hokkien and Mandarin) did not have significant effects on this likelihood (Table 22).

Table 22. Generalized linear mixed-effects regression (with logistic link function) results – likelihood to use lexical tone in a Lánnang-uè word (observations = 3,294, R^2 Tjur = 0.16, random intercepts for participant included).

Predictors	Log-odds	SE	CI	<i>p</i>
(Intercept)	4.21	0.43	3.37 – 5.05	< 0.001
age (younger vs. older)	-0.66	0.45	-1.55 – 0.23	0.144
sex (male vs. female)	-0.51	0.37	-0.21 – 1.23	0.167
proficiency (Tagalog and English)	-0.39	0.17	-0.72 – -0.06	0.022
proficiency (Hokkien and Mandarin)	-0.07	0.2	-0.46 – 0.33	0.746

An examination of the marginal means of the statistically significant Tagalog and English proficiency factor shows that a sizable portion of tokens without lexical tone (tokens with non-phonemic pitch) is associated with speakers who have high proficiency in Tagalog and English (Figure 8). This model predicts that as proficiency in these languages increases, the likelihood of using lexical tone in Lánnang-uè words decreases. Overall, I found that the variation in the use of lexical was conditioned only by Tagalog and English proficiency (high proficiency).

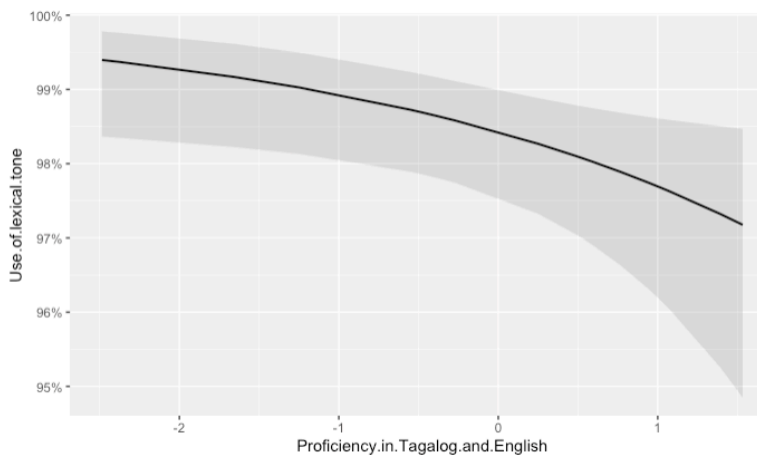


Figure 8. Marginal effects of proficiency in Tagalog and English on likelihood to use lexical tone in a Lánnang-uè word

4.5.3 Tone distributional patterns

4.5.3.1 Descriptive analysis

All participants followed the CVT/CV tone pattern at least once (feature spread score = 1): they used a falling tone (mean pitch slope⁵⁹ = -4.77, SD = 2.5) for CV syllables and a high tone (mean pitch slope = -0.06, SD = 0.48) for CVT syllables in words that were derived from Tagalog and English at least once. The results show that the CVT/CV pattern is widespread.

Out of the 20 speakers, six of the participants (30%) followed the pattern 93% to 95.99% of the time (mean = 94.86%, SD = 0.01). Nine (45%) did 96% to 99.99% of the time (mean = 98.44%, SD = 0.004). Five (25%) always followed the pattern and did not vary at all (Figure 9).

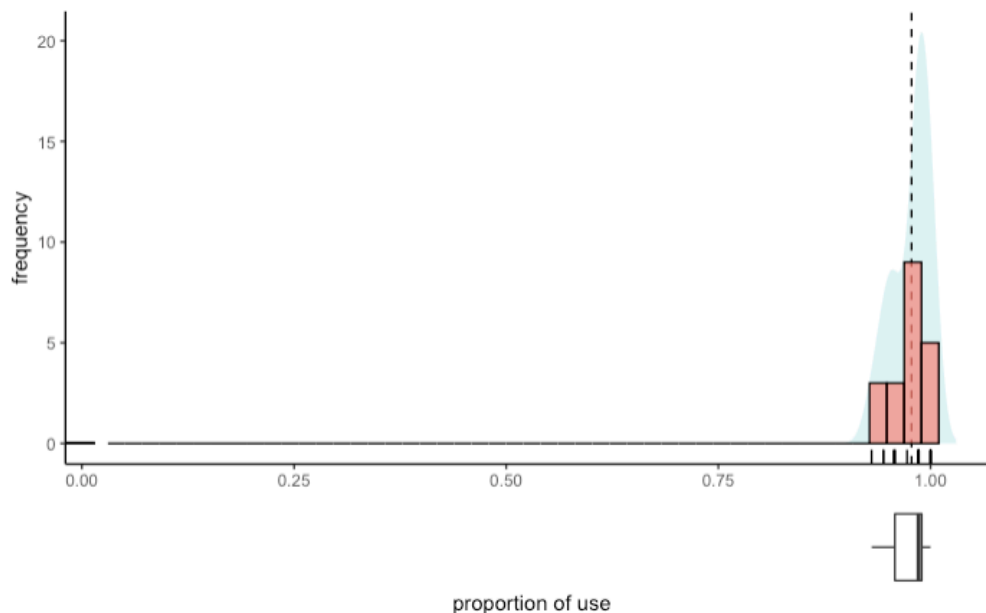


Figure 9. Histogram (frequency), density, and box plot: individual intraspeaker consistency scores (adherence to the CVT/CV tone pattern); broken line indicates mean

The mean intraspeaker consistency score for *all* participants is 0.9776 (SD = 0.02). The interspeaker pattern inconsistency score is 0.0219. Overall, the CVT/CV pattern is stable – speakers produced high tones for CVT syllables and produced falling tones for CV syllables with

⁵⁹ See Chapter 4.4.3 for the rationale behind using pitch slope instead of lexical tone.

high consistency (mean individual intraspeaker feature consistency score above 0.5). The patterns of variation between speakers were also highly homogenous (inconsistency below 0.5).

The total number of tokens that did not follow the stable CVT/CV tone pattern is 32 (out of 1419, 2.25%). Out of these tokens, 8 (25%) had the CVT structure but were coded as ‘falling’ instead of ‘high’, and 24 (75%) had the CV structure but were coded as ‘high’ instead of ‘falling’. A breakdown of the 32 pattern-non-conforming tokens is shown in Table 23.

Table 23. Frequency syllables that did not adhere to the CVT/CV tone pattern, by syllable structure (grave \grave{a} = falling tone, circumflex \hat{a} = high tone)

Syllable structure	Token (in word)	Participant Pronunciation (IPA)	Freq.	Total	% of total variation
CVT	<i>gatà</i> ‘coconut cream’	[ga ²² ta ⁵¹]	1	8	25%
	<i>hotdòg</i> ‘hotdog’	[‘hat ²² dog ⁵¹]	1		
	<i>massàge</i> ‘massage’	[ma ²² sad ⁵¹]	1		
	<i>salàd</i> ‘salad’	[‘sa ²² lad ⁵¹]	1		
	<i>sisìg</i> ‘sizzling diced meat dish’	[‘si ²² sig ⁵¹]	1		
	<i>sukà</i> ‘vinegar’	[‘su ²² ka ⁵¹]	1		
	<i>tahò</i> ‘soft tofu’	[ta ²² ho ⁵¹]	2		
CV	<i>batô</i> ‘rock’	[ba ²² to ⁵⁵]	3	24	75%
	<i>bikô</i> ‘sticky rice cake’	[‘bi ²² ko ⁵⁵]	2		
	<i>bukô</i> ‘coconut’	[‘bu ²² ko ⁵⁵]	1		
	<i>butô</i> ‘seed’	[‘bu ²² to ⁵⁵]	1		
	<i>cookiê</i> ‘cookie’	[‘ku ²² ki ⁵⁵]	1		
	<i>depôt</i> ‘depot’	[‘di ²² po ⁵⁵]	6		
	<i>plateaû</i> ‘plateau’	[pla ²² tu ⁵⁵]	1		
	<i>putô</i> ‘rice cake’	[‘pu ²² to ⁵⁵]	4		
	<i>taxî</i> ‘taxi’	[‘tak ²² si ⁵⁵]	1		
	<i>togâ</i> ‘toga’	[‘to ²² ga ⁵⁵]	4		
Total			32	100%	

Around 25% of the time that the 15 speakers who varied did not adhere to the CVT/CV tone pattern, they used a falling tone for CVT syllables (i.e., *salàd* [‘sa²²lad⁵¹] instead of *salâd* [‘sa²²lad⁵⁵] ‘salad’, *tahò* [ta²² ho⁵¹] instead of *tahô* [ta²² ho⁵⁵] ‘soft tofu’, and *sisìg* [‘si²²sig⁵¹] instead of *sisîg* [‘si²²sig⁵⁵] ‘sizzling diced meat dish’). The rest of the time, they used a high tone

for CV syllables that are expected to be falling (e.g., *depôt* [ˈdi²²po⁵⁵] instead of *depòt* [ˈdi²²po⁵¹] ‘depot’, *togâ* [ˈto²²ga⁵⁵] instead of *togà* [ˈto²²ga⁵¹] ‘toga’, *bukô* [ˈbu²²ko⁵⁵] instead of *bukò* [ˈbu²²ko⁵¹] ‘coconut’, and *batô* [ba²²to⁵⁵] instead of *batò* [ba²²to⁵¹] ‘rock’.

All speakers also followed the CVR-Tagalog/CVR-English tone pattern at least once (feature spread score = 1). They used a high, level pitch (mean pitch slope = -1.03, SD = 1.55, high tone) for CVR syllables in Tagalog-derived words and a falling, contour pitch for CVR syllables in English-derived words (mean pitch slope = -3.61, SD = 1.56, falling tone) at least once. The CVR-Tagalog/CVR-English tone pattern is highly widespread.

Out of the 20 speakers, most (n = 12, 60%) adhered to the pattern 77 to 90% of the time (mean = 85.72%, SD = 0.04). The rest of the participants (n = 8, 40%) followed the pattern 91% to 99% of the time; they rarely varied (mean = 94.61%, SD = 0.02). None of the participants followed the pattern 100%: all speakers occasionally varied (Figure 10).

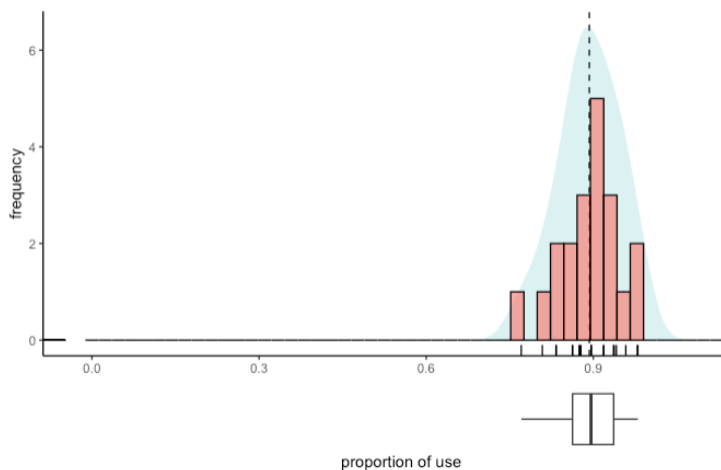


Figure 10. Histogram (frequency), density, and box plot: individual intraspeaker consistency scores (adherence to the CVR-English/CVR-Tagalog tone pattern); broken line indicates mean

The mean intraspeaker consistency score for *all* participants is 0.8928 (SD = 0.05). The interspeaker pattern inconsistency score is 0.0622. Overall, the pattern is highly stable. Speakers produced syllables with level pitch slopes (interpreted as high tones)⁶⁰ for CVR-Tagalog

⁶⁰ The interpretation is warranted after plotting the pitch. The pitch of syllables with near-zero pitch slopes is, on average 190 Hz (starting pitch ~190 Hz, ending pitch ~190 Hz), consistent with a high tone. The pitch of syllables with negative pitch slopes roughly begin at 200 Hz and end at 170 Hz, indicative of a falling tone.

syllables and produced syllables with negative contour pitch slopes (interpreted as falling tones) for CVR-English syllables with high consistency (mean individual intraspeaker feature consistency score above 0.5). The patterns of variation between speakers were also highly homogeneous (inconsistency below 0.5).

The total number of tokens that did not conform to the CVR-Tagalog/CVR-English tone pattern is 100 (out of 972, 10.29%). Out of these tokens, 60 were derived from English but were coded as ‘high’ instead of ‘falling’, and 40 were derived from Tagalog but were coded as ‘falling’ instead of ‘high’. A breakdown of the 100 pattern-non-conforming tokens is shown in Table 24.

Table 24. Frequency of syllables that did not adhere to the CVR-Tagalog/CVR-English tone pattern, by source language (grave \grave{a} = falling tone, circumflex \hat{a} = high tone)

Source language	Token (in word)	Participant Pronunciation (IPA)	Freq.	Total	% of total variation
English	<i>balloôn</i> ‘balloon’	[ba ²² lun ⁵⁵]	5	60	60%
	<i>ballpên</i> ‘ballpoint pen’	[ˈbol ²² pen ⁵⁵]	6		
	<i>basebâll</i> ‘baseball’	[ˈbejs ²² bol ⁵⁵]	7		
	<i>bazaâr</i> ‘bazaar’	[ba ²² za ⁵⁵]	3		
	<i>bouquêt</i> ‘bouquet’	[bu ²² ke ⁵⁵]	3		
	<i>dustêr</i> ‘duster’	[ˈdas ²² tɛr ⁵⁵]	13		
	<i>hotêl</i> ‘hotel’	[ho ²² tɛl ⁵⁵]	2		
	<i>lasêr</i> ‘laser’	[ˈle ²² sɛr ⁵⁵]	9		
	<i>salôn</i> ‘salon’	[sa ²² lon ⁵⁵]	3		
<i>sandâl</i> ‘sandal’	[ˈsan ²² dal ⁵⁵]	9			
Tagalog	<i>atây</i> ‘liver’	[ʔa ²² taj ⁵¹]	3	40	40%
	<i>isàw</i> ‘grilled chicken or pork intestine’	[ˈʔi ²² saw ⁵¹]	2		
	<i>lugàw</i> ‘rice porridge’	[ˈlu ²² gaw ⁵¹]	9		
	<i>siopaò</i> ‘pork bun’	[ˈʃo ²² paw ⁵¹]	2		
	<i>sumàn</i> ‘glutinous rice cake’	[ˈsu ²² man ⁵¹]	11		
	<i>turòn</i> ‘fried banana roll’	[tu ²² jun ⁵¹]	13		
Total			100		100%

Around 60% of the time that the speakers did not adhere to the CVR-Tagalog/CVR-English pattern, they used a high tone for CVR-English syllables instead of a falling tone (e.g., *hotêl* [ho²² tɛl⁵⁵] ‘hotel’, *ballpên* [ˈbol²² pen⁵⁵] ‘pen’, *basebâll* [ˈbejs²² bol⁵⁵] ‘baseball’); the rest of the time (40%) they used a falling tone for CVR-Tagalog syllables instead of a high tone (e.g., *sumàn* [su²² man⁵¹] ‘glutinous rice cake’, *isàw* [ʔi²² saw⁵¹] ‘grilled chicken or pork intestine’, *turòn* [tu²² .ɲun⁵¹] ‘fried banana roll’).

4.5.3.2 Regression analysis

Focusing on regression results⁶¹ that are relevant to my hypotheses, I found evidence that the CVT/CV and CVR-Tagalog/CVR-English tone patterns are highly widespread and stable: the model showed a main effect of structure (CVT vs. CV) and an interaction effect between structure (CVR vs. non-CVR) and source language on pitch slope, effects that appear when many speakers consistently follow the tone patterns (Table 25). Furthermore, I found that these factors interacted with sociolinguistic factors. The CVT/CV factor interacted with age, sex and proficiency in languages without the tone patterns. The CVR interaction factor (i.e., structure and source language) interacted with age.

⁶¹ The model was fitted on a subset of the data with only the final syllables of Tagalog- and English-origin words (n = 2391), to control for syllable position and any potential phonological effect.

Table 25. Linear regression results for final syllable pitch slope (observations = 2,391, conditional $R^2 = 0.41$, random intercepts for participant and item).

Predictors	Estimates	SE	CI	<i>p</i>
(Intercept)	-5.49	0.67	-6.80 – -4.18	< 0.001
structure (CVR vs. non-CVR)	1.85	0.57	0.73 – 2.97	0.001
structure (CVT vs. CV)	3.19	0.27	2.65 – 3.73	< 0.001
source language (Tagalog vs. English)	5	0.68	3.67 – 6.33	< 0.001
age (younger vs. older)	2.27	0.62	1.05 – 3.49	< 0.001
sex (male vs. female)	0.56	0.57	-0.56 – 1.68	0.325
proficiency in languages with no tone distribution	0.02	0.16	-0.29 – 0.34	0.89
source language : age	-2.87	0.47	-3.80 – -1.95	< 0.001
source language : sex	-0.74	0.43	-1.58 – 0.10	0.085
source language : proficiency	-0.12	0.25	-0.62 – 0.37	0.631
structure (CVR vs. non-CVR) : source language	-5.11	0.87	-6.82 – -3.40	< 0.001
structure (CVR vs. non-CVR) : age	-0.86	0.4	-1.64 – -0.09	0.029
structure (CVR vs. non-CVR) : sex	-0.02	0.36	-0.71 – 0.68	0.965
structure (CVR vs. non-CVR) : proficiency	-0.07	0.21	-0.47 – 0.33	0.731
structure (CVT vs. CV) : age	-0.92	0.19	-1.29 – -0.55	< 0.001
structure (CVT vs. CV) : sex	-0.45	0.17	-0.78 – -0.11	0.009
structure (CVT vs. CV) : proficiency	-0.26	0.1	-0.45 – -0.06	0.011
structure (CVR vs. non-CVR) : source language : age	2.99	0.61	1.81 – 4.18	< 0.001
structure (CVR vs. non-CVR) : source language : sex	0.88	0.55	-0.19 – 1.96	0.108
structure (CVR vs. non-CVR) : source language : proficiency	0.12	0.32	-0.51 – 0.75	0.714

An examination of the marginal effects revealed the direction of the effect of syllable structure on pitch slope. CVT syllables were associated with higher, near-zero pitch slopes (which I interpreted as high tones) while CV ones were associated with lower, negative pitch slopes (which I interpreted as falling tones) (see Figure 39 in Appendix F for individual speaker patterns). An examination of pitch slope values in CVR syllables by source language for the same participants showed syllables in English-derived words having relatively smaller pitch slopes (falling tones) than syllables in Tagalog-derived words, indicative of high tone use (see Figure 40 in Appendix F for individual speaker patterns).

I provide a plot of pitch over time by syllable and source language, averaged across all data points, to illustrate the effects of syllable structure on pitch (slope) more clearly (Figure 11). It provides a glimpse of the CVT/CV and CVR-English/ CVR-Tagalog tone patterns, which have high degrees of spread and stability.

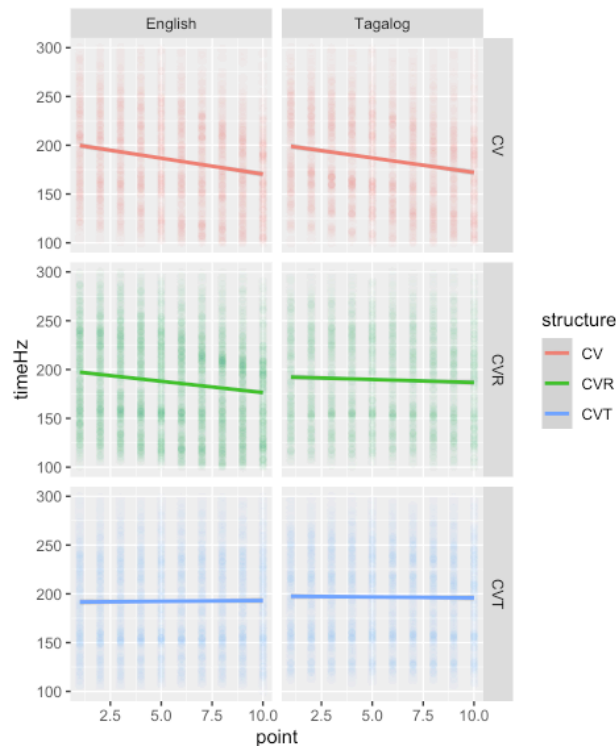


Figure 11. Syllable structure and pitch slope by source language of the word (final syllables of Tagalog- and English-origin words)

My analysis of the variation – particularly the marginal means of the interaction terms related to the CVT/CV structure in my model of pitch slope – revealed that a significant portion of tokens that did not conform to the CVT/CV tone pattern in Tagalog- and English-origin words was associated with young participants, male participants, and participants highly proficient in languages without the CVT/CV pattern (i.e., Hokkien, Tagalog, English, Mandarin). The tendency to not adhere to the pattern can be observed in Figure 12, where the (mean) slope values of the CVT (blue) and CV (red) syllables for the (a) younger, (b) male, and (c) high proficiency groups are closer to each other than the older, female, and low proficiency groups, who have means that resemble the stable distributional pattern more. My model predicts that speakers will be more likely not to conform to the CVT/CV pattern if they are part of any of the

following social groups: younger speakers, male speakers, and speakers highly proficient in languages without the CVT/CV pattern.

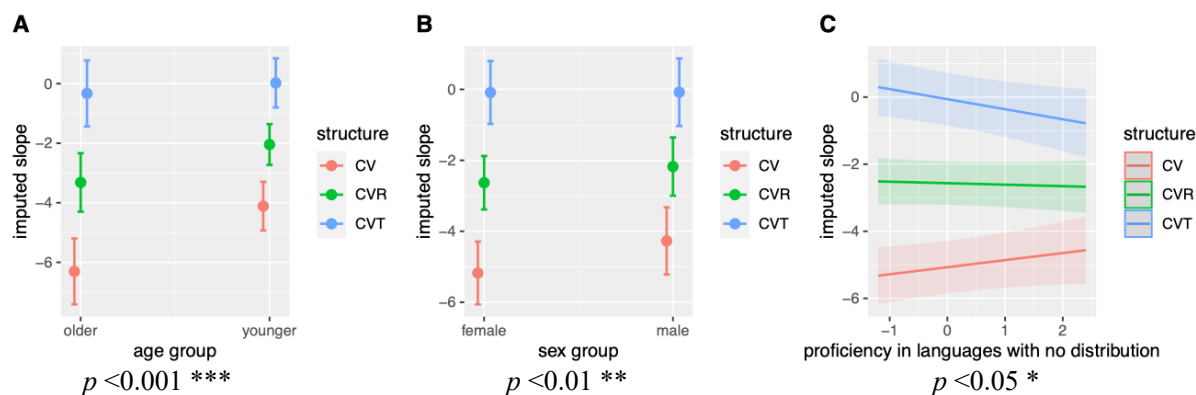


Figure 12. Marginal means/effects of (a) age, (b) sex, and (c) proficiency on the CVT/CV tone pattern in Tagalog- and English-origin words; ignore CVR, which was automatically included in plotting package; slope means close to zero suggest use of high tone, slope means -5 and below suggest use of falling tone, slope means near -3 (e.g., CVR syllables) indicate mixed use of falling and high tone.

Examining the marginal means of the interaction terms related to the CVR-English/CVR-Tagalog variable in my model, I found that most of the tokens that did not conform to the tone pattern were only associated with younger participants. The tendency to vary can be observed in Figure 13, where the (mean) slope values of the English- (blue) and Tagalog-origin (red) syllables for younger participants are visibly closer to each other than the means of the older participants, who have means that resemble the stable tone pattern more. This is unlike the means for the sex and proficiency groups (see plots (b) and (c) in Figure 13), where virtually no differences between groups can be observed. My model predicts that speakers will be less likely to produce tokens that do not adhere to the CVR-English/CVR-Tagalog pattern if they are younger.

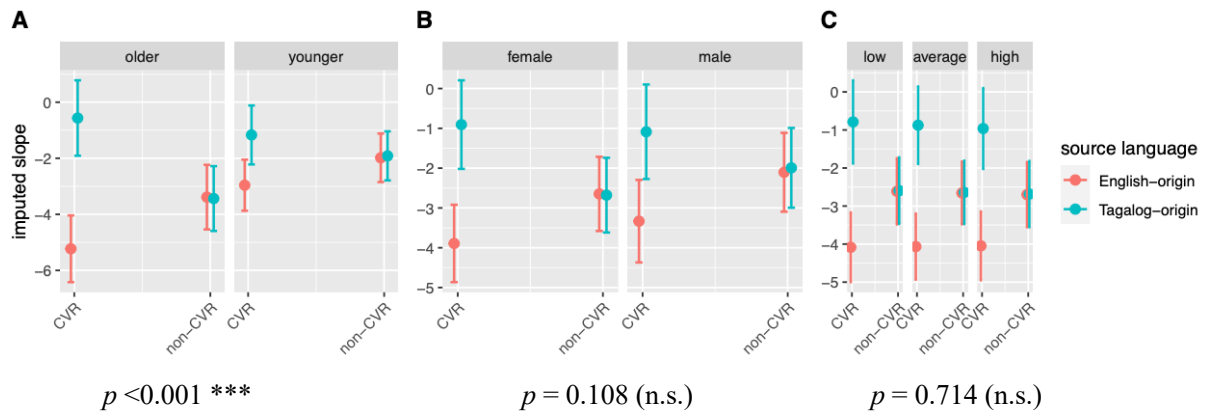


Figure 13. Marginal means/effects of (a) age, (b) sex, and (c) proficiency – the CVR-Tagalog/CVR-English tone pattern; slope means close to zero suggests use of high tone, slope means -5 and below suggests use of falling tone, slope means near -3 (e.g., non-CVR syllables) suggests mixed use of falling and high tone. (n.s. = not significant)

I provide an alternative illustration of the age effect on the CVR-English/CVR-Tagalog pattern – a plot of pitch over time by syllable, source language, and age group, averaged across all data points (Figure 14).

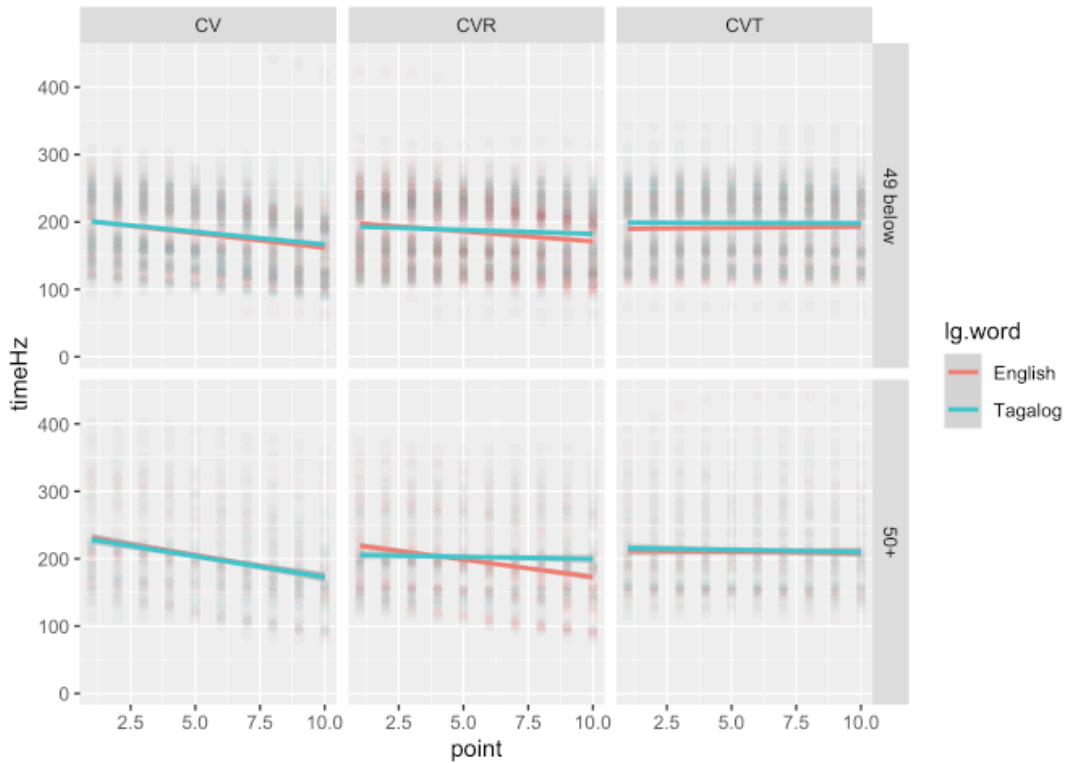


Figure 14. Final pitch slope, structure, age, and source language of the word (Tagalog- and English-origin words)

4.6 Discussion

4.6.1 Hypotheses 1 and 2: Spread and stability

The results of my analyses supported two of my hypotheses: that the four prosodic features – the lexical stress, lexical tone, the CVT/CV pattern, and the CVR-English/CVR-Tagalog pattern – will be highly widespread and stable. I found that most speakers within my sample used the four features with high levels of consistency. I also found that the patterns of variation among speakers are highly homogeneous. The results suggest that the four prosodic features exhibit high degrees of spread and stability within the community of Lánnang-uè speakers.⁶²

⁶² The inclusion of random effects for participant in my model gives me some statistical license to generalize this claim to the population of Lánnang-uè speakers (Konstantopoulos and Hedges 2019:278). It should be noted, however, that this is not a perfect remedy for my small sample (data). While the random effects strategy offers me some license to generalize, the only true measure of whether something is generalizable is through study replication – that is, using the same model to predict the linguistic behavior of participants that are not part of this study, and seeing whether the main and interaction effects remain.

The patterns of spread and stability are expected if Lánnang-uè has a high degree of languageness – an argument I made in previous work (Gonzales 2018; Gonzales and Starr 2020) and in Chapter 3. I was also not surprised to have found high rates of spread and stability in the four prosodic features because I have found evidence of them in other phonological domains (Gonzales and Starr 2020). Despite evidence of spread and stability (two hallmarks of languageness), it is important to acknowledge the possibility that Lánnang-uè is not immune to the influence of external factors (i.e., Lánnang-uè may be viewed as not being fully crystallized), given that it is situated in a complex, multilingual setting. Keeping this possibility in mind, what factors reinforced the spread and stability of the prosodic patterns? Why do these features still exhibit high degrees of spread and stability despite the situation?

4.6.1.1 Possible reinforcers of the stress pattern

One possible reason for the continued spread and stability of the stress pattern is feature “congruence” (Baptista 2020:162; Matras and Sakel 2007). Most speakers, for example, continue to lengthen/stress syllables in syllables that are expected to be stressed in Lánnang-uè and consistently do so perhaps because similar stress features can be found in three of the source languages all of my speakers know – Philippine English, Tagalog, and Mandarin (Lesho 2018; Gonzalez 1970; Chen 2006; Anderson 2006).⁶³ The presence of the feature in many other languages in the Lannang linguistic ecology might have reinforced the use of the stress pattern in the community.

Another reason why the pattern continues to be widespread and stable is because stress encodes lexical meaning. A word that is unstressed (e.g., [ba³³ka⁵¹]) is ambiguous. Applying the

⁶³ While the stress feature may have been selected into Lánnang-uè, it was not simply transplanted from the source languages as is. For one, not all words derived from Tagalog or English retain the original stress placement from the source language (e.g., English-derived *utensil* [ˈju²²ten²²si⁵¹] ‘utensil’ vs. English *utensil* [juˈtensil] ‘utensil’, Tagalog-derived *kasi* [ˈka²²si⁵⁵] ‘because’ vs. Tagalog *kasi* [kaˈsi] ‘because’). Furthermore, Philippine English and Tagalog lexical stress is marked differently compared to Lánnang-uè: in Philippine English, preliminary results indicate that the variety uses “vowel reduction” – a combination of sonority, duration, amplitude – to mark lexical stress (Lesho 2018:11). In Tagalog, lexical stress is realized acoustically by higher F₀, greater intensity, and longer duration (Gonzalez 1970:16). However, the three correlates of stress are not always simultaneously used. Gonzalez (1970:28) reports that in certain syllables with particular segmental properties, speakers may not use a particular cue for Tagalog lexical stress: for example, F₀ is not used to mark stress in syllables that contain a voiced stop; intensity is not used for syllables with low vowels; duration is not used to mark stress in non-final CV syllables. In other words, lexical stress in Tagalog has three cues for stress (i.e., F₀, intensity, and duration), but these are inconsistently used. Unlike Tagalog and Philippine English, Lánnang-uè stress is exclusively and consistently marked using duration, evidenced in production and preliminary perception results (see Appendix E).

stress contrast to it – through lengthening and shortening – allows for the encoding of different unambiguous meanings (i.e., [‘ba³³ka⁵¹] ‘cow’, [ba³³‘ka⁵¹] ‘maybe’). But why is stress still marked exclusively using duration? One reason is because the use of duration for stress does not interfere with other prosodic systems in the variety, such as the tone systems, where duration is irrelevant. Another possible reason is because duration is near-universally successful in marking stress. Cross-linguistically, duration was able to distinguish stress in “85 of 100 (sub-) studies and 65 of 72 languages” (Gordon and Roettger 2017:5). Given other choices to mark stress (e.g., F₀, amplitude), speakers of the variety perhaps want the optimal or most effective marker for stress discrimination.

4.6.1.2 Possible reinforcers of the lexical tone feature

The continued widespread and stable use of lexical tone in Lánnang-uè arguably has a social underpinning – the negotiation of the Lannang identity using lexical tone. In my interviews with my 20 speakers (Section 4.4.1), all of them expressed concerns about their Lannang identity deteriorating – within the Lannang community, my participants felt not Chinese enough. They attributed this situation to Filipinization, a gradual process that saw the removal of Chinese subjects in Lannang schools and the introduction of pro-Filipino policies. While they did not view the “Filipino” aspect of their Lannang identity (Gonzales 2021a:8) negatively, they were alarmed by the fact that it has overshadowed the “Chinese” aspect (Gonzales 2021a:9). I argue that this awareness of attrition – which began long before the present, perhaps in the 1970s when the Filipinization act was institutionalized (Tan 1993:29) – contributed to the continued spread and maintenance (stability) of lexical tone use in Lánnang-uè (their ‘tribal’ variety or in Lánnang-uè speaker’s terms, a “secret language”) (Lannang Corpus, file CLIN-19-10:4076). They continued to use lexical tone – a feature acknowledged by Lánnang-uè-speaking community members to be (distinctively) ‘Chinese’ (ethnographic notes, 2019) – consistently in Lánnang-uè words regardless of source language in order to orient themselves as ‘Chinese’ and sound more Chinese with each other. In other words, the use of tone arguably has “social meaning” for the community (Benor 2010:160; Eckert 2012) – it means countering identity attrition and asserting the Chineseness of their Lannang identity. Overall, I suspect that community-wide stressing of ‘Chineseness’ contributed to the continued widespread and stable

use of lexical tone in Lánnang-uè. However, I am reluctant to commit to this account until I find such a link in an empirical investigation of identity construction and lexical tone.

Another possible reason why the use of lexical tone continues to be widespread and stable within the community is pattern extension. Since most of the vocabulary in Lánnang-uè is sourced from Hokkien (Chapter 3.2), a tone language, speakers of Lánnang-uè might be more inclined to continue extending this tonal pattern to the rest of the vocabulary (e.g., words derived from Tagalog). The consequence of such an extension is a variety with a lexicon that continues to be toned.

4.6.1.3 Possible reinforcers of the CVT/CV and CVR-English/Tagalog tone patterns

I argue that phonetics is a contributing factor, based on findings derived from cross-linguistic research (Kingston 2011; Zhang 2002; Remijsen 2014). Studies have shown, for example, that ‘markedness’ – reflected in greater “articulatory effort” (Yip 2002:190) – as well as sonority, “sonorous space” (Remijsen 2014:674), or “sonorous energy” (Gordon 2001:425) (i.e., the syllable’s phonetic suitability to carry tonal information) can influence tone selection or assignment (Zhang 2002; Gordon 2001; Remijsen 2014; Kingston 2011). The central idea is this: the greater the sonorous portions of the syllable (e.g., bimoraic CV syllables, (Gordon 2001:447)), the more able it is to support contour tones (Gordon 2001:425); the less sonorous energy it has (e.g., CVT syllables), the more likely it is to be assigned a level tone. The exact contour or level tone (i.e., rising vs. falling, rising vs. rising-falling) may be conditioned by the degree of articulatory effort needed to produce certain tonal features (as well as a host of other factors such as perception, coarticulation, inherent pitch of vowels, etc.) (Prince and Smolensky 2008; Yip 2002).

In the case of the Lánnang-uè CVT/CV high-falling pattern, a possible reason for the continued assignment of high tone to CVT syllables and falling tone to (bimoraic) CV syllables is that there is not enough sonority in CVT syllables to support contour tones but there is in CV syllables. Preliminary evidence suggests that this account is plausible: CV syllables in Tagalog- and English-derived Lánnang-uè words have more sonorous portions (mean = 0.12, SD = 0.07, n = 2733) compared to CVT syllables (mean = 0.08, SD = 0.04, n = 895).

Sonority may also partially contribute to the reinforcement of the CVR-Tagalog/CVR-English high-falling pattern if one accepts the proposition that Lánnang-uè speakers analyzed

Tagalog-derived CVR syllables differently from English-derived ones. The vowel-sonorant rime in Tagalog-derived syllables could have been analyzed by speakers as only having one mora or tone bearing unit whereas those from English-derived syllables have two. There is some evidence for this – CVR syllables in Lánnang-uè derived from Tagalog (mean = 0.16 seconds, SD = 0.05, n = 792) have shorter rimes or less sonorous space than ones derived from English (mean = 0.18 seconds, SD = 0.05, n = 366). And it is established cross-linguistically that limited sonorous space constrains the use of a contour tone and that sonorous portions support contour tones (Gordon 2001:435). If Lánnang-uè speakers really did analyze the CVR-Tagalog syllables differently from CVR-English syllables, then sonority is a plausible factor that contributes to the continued adherence to the CVR-Tagalog/CVR-English high-falling tone pattern.

Speakers may have continued to use falling tone for CV and CVR-English syllables instead of rising tone and other contour tones (e.g., falling-rising [515]) partially because it requires the least articulatory effort among the contour tones (Ohala 1978:30; Zhang 2016:429; Cheng 1973).

4.6.2 *Hypothesis 3: Structured variation (Systematicity)*

To recapitulate, the results of my regression analyses partially supported my hypothesis that variation in the use of the four prosodic features is structured – that it will be conditioned by age and/or sex. I found at least one of these factors conditioning the variation in three prosodic features investigated in this chapter (i.e., all except variation in lexical tone). This, again, is suggestive of a high degree of languageness in Lánnang-uè, as established contact languages also have systematicity or sociolinguistically conditioned variation (e.g., Singlish, Baba Malay, Light Warlpiri, Gurindji Kriol) (Starr and Balasubramaniam 2019; Lee 2014; Meakins and O’Shannessy 2010).

The results only partially supported my hypotheses on the direction of the conditioning effects of the sociolinguistic factors. The direction of the effect was not expected for some prosodic features.

I attempt to explain the sociolinguistic patterns found systematically by variable (i.e., age, sex) in the following two subsections. I discuss them keeping in mind that Lánnang-uè may have sociolinguistic variation embedded in its system (i.e., that sociolinguistic variation is a crucial part of the variety or that command of the variety requires skillful manipulation of variants).

4.6.2.1 Age

Age conditioned a sizable portion of the variation in the stress pattern and the variation in the two tone distribution patterns, and there is no evidence that it conditioned the variation in lexical tone. A significant portion of the tokens that did not conform to the tone patterns came from younger speakers, while the bulk of tokens that did not have duration-cued stress, a highly widespread and stable feature in Lánnang-uè, came from older speakers.

The first pattern – younger speakers tending to deviate from widespread (conventionalized) patterns more than older speakers – supports my hypothesis on the direction of the age effect. The pattern is expected given a similar pattern I have found in Lánnang-uè phonology (Gonzales 2018; Gonzales and Starr 2020) and variationist research on the phonology of other linguistic varieties in the context of linguistic innovation (Starr and Balasubramaniam 2019). In a sociophonetic study of the Lánnang-uè vowel system, I discovered, for example, that younger speakers generally followed community norms, but occasionally did not follow them. For example, one norm is the production of the vowel [ʊ] in Tagalog-, English- and Hokkien-derived words with consistent vowel qualities (i.e., similar F1 and F2 values). Unlike older speakers who consistently followed this convention, younger speakers were more likely to produce [ʊ] vowels differently in Tagalog-derived words compared to [ʊ] vowels in English-derived words. They appear to be introducing innovations to the existing system. Research on other linguistic varieties has also shown that young speakers tended to be at the forefront of linguistic change (Sande 2015; Maclagan et al. 1999), as they are “people with energy and enterprise” (Maclagan et al. 1999:19). So, if the use of tokens that did not conform to the two widespread tone patterns is reflective of innovation, then it would not be surprising to have observed younger speakers occasionally patterning differently from other speakers.

The first pattern is likewise unsurprising in contexts beyond language change/innovations. Research has shown that younger speakers can vary their use of language for many reasons, including deliberately countering adult norms (Eckert 1989:264; Hurst 2009; Sande 2015) and engaging in complex social stylistic practice (Eckert 1989:245; Eckert and McConnell-Ginet 2003; Shin 2013) such as avoiding certain language styles that sound ‘old’, among other styles. However, if the pattern-non-conforming tokens associated with the youth indeed emerged out of non-innovative contexts, I have yet to find the exact motivation for them

– did the youth tend to pattern differently from other speakers because they wanted to avoid sounding ‘old’ or for some other reason? Future research can investigate this, as I currently do not have enough evidence from my interviews to commit to a fine-grained account.

Under the lens of linguistic innovation, the second pattern – older speakers tending not to follow the widespread stress pattern – did not support my hypothesis on the direction of the age effect. It is surprising given what is commonly found in sociophonetic research that views variation as linguistic innovation – that younger speakers tend to pattern differently from other speakers by introducing innovative features to the mainstream variety (Sande 2015; Starr and Balasubramaniam 2019), as discussed earlier. The pattern is, however, not surprising from the perspective of stylistic practice (Eckert 1989:245; Eckert and McConnell-Ginet 2003; Shin 2013), where older speakers could use a feature to express a particular style specific to their group. It is also not surprising from the perspective of language development, where the pattern could be regarded as a vestige of change that has already occurred rather than a reflection of ongoing change led by older speakers (Sankoff 2006; Gordon et al. 2004). The behavior of the older speakers may represent a phase of Lánnang-uè when the stress pattern was not as widespread or stable. The increased variation in the older speakers’ speech could be remnants of a potential language change from a variety that did not have the stress pattern to a variety with it. There is evidence of a similar pattern in my previous work on Lánnang-uè derivational morphology (Gonzales 2018), which I also argued to be indicative of change that has occurred. I found that the oldest speakers (speakers in their 90s) had patterns that differed significantly from the stable morphological norms of all other members in the community– unlike the rest of the speakers (20s to 80s), they tended to regard words with short Tagalog-derived prefixes as unacceptable and to avoiding using those prefixes.

I emphasize that my sample size is relatively small. Thus, discussions of change involving this sample should be examined with great caution. Also, in discussions of change here and throughout, I am assuming that Lánnang-uè already has (at least) some stability and degree of conventionalization as a whole, based on preliminary evidence in previous work (Gonzales 2018; Gonzales and Starr 2020).

4.6.2.2 Sex

Sex conditioned most of the variation in the stress pattern and the variation in the CVT/CV tone distributional pattern, and there is no evidence of it conditioning the variation in lexical tone and the CVR-Tagalog/CVR-English tone pattern. Many of the tokens that did not follow the widely-used and stable stress pattern came from female speakers, while most of the tokens that did not conform to the widespread and stable CVT/CV tone pattern came from male speakers.

The first pattern involving female speakers is expected in the context of linguistic innovation and sound change, given the plethora of sociolinguistic literature that found female speakers tending to use innovative patterns (Labov 1972; Eckert 1989; Maclagan et al. 1999; Shin 2013), including previous research on Lánnang-uè (Gonzales and Starr 2020).

In the same sociophonetic study discussed in Section 4.6.2.1, I also found that female speakers generally produced [e] vowels in Tagalog-, English-, and Hokkien-derived words with similar vowel qualities (following the convention) but occasionally produced the vowel differently depending on the source language of the word. They tended to be less consistent in their production of conventionalized phonetic features compared to male speakers, presumably leading the change in an already conventionalized vowel system. The pattern is also well-documented in research on sociolinguistic variation: scholars often reported female speakers tending to exhibit innovative patterns or leading linguistic change, especially sound change, because they are full of “energy and enterprise” (Maclagan et al. 1999:19). As such, I am not entirely surprised to have observed female speakers following the widespread and stable stress patterns but tending to be less consistent in following them compared to male speakers, assuming of course that patterns of variation are indicative of innovation. Overall, I have hinted at the possibility that the variation is because of potential language change driven by female speakers. However, given that this effect was found in a limited sample size of twenty speakers, I hesitate to commit to this account fully.

I also expected the first pattern beyond the context of linguistic innovation, as female speakers have been documented in variationist research to pattern differently from male speakers for a host of reasons. These include wanting to sound more interpersonal (Shin 2013), wanting to project a certain style (e.g., cool) (Eckert 1989; Eckert and McConnell-Ginet 2003), or increased exposure to languages that could encourage the use of variants (Shin 2013).

I am however intrigued by the second pattern involving male speakers, as this pattern is not as common in the sociolinguistic literature compared to the first pattern (Labov 1972; Obeidat and Hammoudi 2019). I am not entirely surprised by the pattern, as sociolinguistic research has documented such patterns, mostly in the context of stylistic variation and not innovation or (sound) change. For instance, in Algeria, Obeidat and Hammoudi (2019) found that men using Arabic tended to pattern differently from the community norm of using the glottal stop [ʔ] for the phoneme /ʔ/. They found that while men generally followed the convention, for reasons such as building community rapport and identity-stressing (Obeidat and Hammoudi 2019:375), men were found to exhibit higher rates of [g] usage than women because they wanted to avoid sounding “feminine”, a characteristic associated with the glottal stop (Obeidat and Hammoudi 2019:376).

In the context of this study, I cannot pinpoint the exact reason why men tended to pattern differently from women regarding the widespread CVT/CV pattern or why exactly women tended to be less consistent than men in following the stress pattern. The motivations for the age effect (male or female) tended to be clear in variationist studies that I have surveyed, but they are unfortunately not as clear in this study, due to lack of evidence. Perhaps the avoidance of stress and CVT/CV features encodes social meaning specific to female or male speakers (Eckert 1989; Eckert 2012), respectively, and the increased variability in the use of stress and the CVT/CV pattern is due to an increased desire to express that social meaning. However, due to lack of evidence from my interviews, I hesitate to discuss the pattern from an agentive perspective further. Future research can attempt to shed more light on the exact motivation(s) for the sex-conditioned patterns observed in this chapter.

4.6.3 Hypothesis 4: Linguistic independence

My regression results revealed a negative correlation between source language proficiency (high proficiency) and variation patterns (non-conforming tokens) in two prosodic features: lexical tone and the CVT/CV tone pattern. Many of the tokens that did not have lexical tone and tokens that did not conform to the CVT/CV tone pattern came from speakers who reported being highly proficient in languages without those features. I do not have evidence for such a trend for the lexical stress pattern or the CVR-English/Tagalog pattern. The results can be interpreted in two

ways. Depending on the interpretation, the results may or may not be useful for testing the hypothesis of linguistic independence.

If we interpret the negative correlation as a causal relationship where proficiency in the source languages affected the patterns of variation, then the results overall provide some evidence against linguistic independence for some features, and subsequently the claim that Lánnang-uè is highly language-like. Under this interpretation, the presence of a negative correlation between proficiency and variation (lexical tone pattern and the CVT/CV tone pattern) indicates that high proficiency in the source languages (or rather, knowledge of other languages' prosodic structure) partially encourages speakers to deviate from the two prosodic patterns (or transfer prosodic structure from the source languages to Lánnang-uè) – evidence for linguistic dependence and against languageness. This is consistent with what scholars have found – that high proficiency (or knowledge of structures) in a particular language contributes to linguistic transfer (Pham et al. 2018; Siegel 2012).

If we interpret the proficiency variable in the negative correlation as 'expression of proficiency' rather than actual proficiency (e.g., speakers not conforming to a pattern because they want to express high proficiency in a particular language),⁶⁴ then the results do not have anything to say about linguistic independence, but support the claim that Lánnang-uè is highly language-like, as the use of pattern-non-conforming features is interpreted as a stylistic choice (speakers' skillful manipulation of linguistic resources to express particular social meaning) (Eckert 2005; Hall-Lew et al. 2021) instead of a consequence of (subconscious) linguistic transfer. In this case, the use of pattern-non-conforming prosodic features is part of the linguistic system of Lánnang-uè. There are three possibilities.

One possibility is that speakers deliberately used non-conforming patterns (patterns derived from Lánnang-uè's source languages) because they want to indicate to others that they are proficient in those languages. But I do not have concrete evidence of this account in my post-experiment interviews. For instance, none of my participants explicitly mentioned using non-phonemic pitch (lack of lexical tone) or not following the CVT/CV pattern to stress that they are highly proficient in Hokkien, Tagalog, English, or Mandarin. Now, another possibility is that the

⁶⁴ Note that the proficiency factors I used in my analysis were self-reported. 'Proficiency' can thus be interpreted as 'expression of proficiency in a particular source language.'

agentive process is subconscious – speakers were not aware that they were using these two non-conforming prosodic features to express their command of Lánnang-uè’s source languages.

It might be worthwhile to consider a third possibility – that speakers who have high proficiency in Hokkien, Tagalog, English, or Mandarin might have patterned differently from other speakers not because they wanted to directly express proficiency in a particular language, but because they wanted to emphasize the Chinese-ness or Filipino-ness of their Lannang identity. This possibility has some merit, given what I found in my interviews. I discovered that many speakers who reported being highly proficient in Tagalog and English tended to identify as more Filipino than Chinese compared to other speakers and those who reported being highly proficient in Hokkien and Mandarin tended to identify as more Chinese than Filipino. I did not find direct evidence of a causal relationship between identity stressing and the production of innovative prosodic features in my interviews, but there is indirect evidence of this in my ethnographic work, where some Lannang speakers reported that the use of Tagalog-derived elements in Lánnang-uè stresses the Filipino aspect of their identity while the use of Hokkien-derived elements in the variety stresses Chinese-ness (Gonzales 2021a). From an ethnolinguistic repertoire perspective where linguistic resources are used to construct ethnic identities (Benor 2010:160), it is possible, for example, that speakers may have used prosodic patterns that are associated with Filipino-ness (e.g., use of non-phonemic pitch in Tagalog- and English-derived words instead of lexical tone) to construct a Lannang identity that is more Filipino-oriented (Gonzales 2021a).

Overall, in all three possible scenarios, the patterns of variation are regarded as part of Lánnang-uè’s prosodic system. They are skillfully used to express particular social meanings within Lánnang-uè. The sociolinguistic patterns, under this view, do not have anything useful to say about the linguistic independence (Hypothesis 4) of Lánnang-uè’s prosody. However, they provide support for Lánnang-uè’s high degree of languagenes nevertheless, as evidence of systematicity or structured variation (Hypothesis 3).

4.7 Conclusion

In this chapter, I conducted a formal, systematic investigation of four prosodic features that seemingly had higher rates of variation compared to other features in Lánnang-uè (described in Chapter 3 and summarized in 4.2): lexical stress, lexical tone, the CVT/CV tone distributional

pattern, and the CVR-Tagalog/CVR-English pattern. The main goal was to test whether the variation observed in the use of these features threatens my argument that Lánnang-uè is highly language-like. I focused on four properties of languagehood: feature spread within the community, stability (as measured by the levels of individual consistency and consistency of variation patterns between speakers), structured variation (as measured by the presence of conditioning sociolinguistic effects), and linguistic independence (Weinreich et al. 1968; Ghyselen and De Vogelaer 2018; Meir and Sandler 2019).

The results generally echoed my previous findings of Lánnang-uè being very language-like (Gonzales 2018; Gonzales and Starr 2020): all four prosodic features were widespread within the sample and had high levels of consistency at the individual and community level (high stability). Most speakers consistently used the four prosodic features, features that continue to be highly widespread and stable arguably due to social and linguistic factors (e.g., pattern extension, identity-stressing, articulatory effort).

The variation in the (non)use of these features was also partially structured: three of the four features investigated were systematically conditioned by age and/or sex. I found that the direction of the conditioning effect was not uniform among the three prosodic features and attempted to explain the effects in Section 4.6.2. Regardless of the asymmetry, however, the results indicate that the variation was structured.

Finally, I found some evidence against linguistic independence in Lánnang-uè prosody, assuming that there is a causal link between language proficiency and variation, where high proficiency encourages the use of pattern-non-conforming prosodic structure. Under this assumption, two prosodic features (i.e., CVT/CV tone pattern, lexical tone), were found to be influenced by proficiency in languages without these features. However, if we assume that the proficiency variable in the causal relationship represents ‘expression of proficiency’ rather than actual proficiency, then results relevant to proficiency have nothing useful to say about linguistic independence (Hypothesis 4) and instead provide evidence in support for structured variation (Hypothesis 3) – speakers use the pattern-non-conforming structure to express social meaning or to show others that they have command in Lánnang-uè’s source languages.

I identify three primary limitations in my study. First, I only had 20 participants complete the experiment. Although I mitigated the issue statistically by including a random intercept variable to normalize the data based on individual effects (licensing me to generalize)

(Konstantopoulos and Hedges 2019), having more participants will allow me to generalize with more certainty. Unfortunately, I was not able to get more participants due to the COVID-19 pandemic. Many prospective participants did not want to participate in the experiment as they wanted to attend to urgent family-, business-, and/or health-related concerns. Because of the limited number of participants, I remain cautious about all the generalizations I made in this chapter.

Second, all of my participants were less than 64 years old. I was not able to get high-quality experiment recordings from speakers 64 years old and above. Local government policies made to curb the pandemic restricted in-person data collection involving speakers above 65; online data collection was also not possible because all the old speakers (65 years old and above) that I contacted refused to participate in my (online) experiment, citing privacy issues, personal issues, and unfamiliarity with technology. This resulted in a participant distribution that is skewed towards younger speakers. Because of the skewing, I will not be able to draw any conclusions that involve this older population.

Third, two participants reported that they were unfamiliar with some of the words on the list because I used Mandarin Pinyin orthography instead of Lannang Orthography for Mandarin-origin words. Two, in particular, asked about some of Mandarin-origin words in the list. For instance, they asked about the word *shiba*, which could refer to a breed of dog in Lannang Orthography (pronounced as [ʃi³³ba⁵¹]) or the number ‘eighteen’ in Mandarin Pinyin (pronounced as [ʃe¹¹pa⁵⁵]). Because these participants were unfamiliar with the orthography of some of the words on the list (and potentially other participants who have not disclosed unfamiliarity with the words), there is a risk that unfamiliarity affected the pronunciation of some words, even after practice. This may have increased the amount of variation that my (sociolinguistic) prosodic models were not able to account for.

It is important to note, however, that while there were limitations regarding the data and analysis, the findings of this chapter should not be discounted, as they shed light on patterns in previously undocumented prosodic phenomena in Lánnang-uè. They also bring us closer to definitively answering the question of whether Lánnang-uè is highly language-like. Going back to the main question raised in the chapter, I found that the variation in the four prosodic features, anecdotally observed to be more extensive compared to variation in other features, did not pose a significant challenge to the proposition that Lánnang-uè is highly language-like: not only was the

use of the prosodic features widespread and highly stable, the little variation observed in their use was conditioned by language-external factors, just like the variation found in established contact languages such as Singlish, Baba Malay, Light Warlpiri, and Gurindji Kriol (Starr and Balasubramaniam 2019; Lee 2014; Meakins and O’Shannessy 2010). Only some prosodic features were conclusively found to be dependent on the prosodic patterns of Lánnang-uè’s source languages, under a language transfer perspective. Pending further investigation in a larger sample, the findings of this chapter support the idea that Lánnang-uè has features of a language, using prosodic evidence.

Chapter 5 : Conjunction and Preposition Lexical Patterns

5.1 Introduction

In this chapter, I conduct a more comprehensive and systematic investigation of the lexical distributional patterns of conjunctions and prepositions in Lánnang-uè – that is, the derivation of certain conjunctions and prepositions from three of Lánnang-uè’s source languages: Hokkien, Tagalog, and English. The two distributional patterns, which I have described in Chapter 3 and summarize later in Section 5.2, were anecdotally found to be more variably followed by speakers compared to other patterns or features in Lánnang-uè (e.g., pronouns, numerals), similar to the prosodic features examined in Chapter 4. Again, the potentially high rates of variation suggest that the patterns are not consistently followed by many Lánnang-uè speakers compared to other patterns/features; they suggest that the patterns have a lower degree of spread and stability than others. In addition, I was not able in my brief analysis to identify factors that could account for a significant part of the variation, suggesting that the variation I observed in the use of conjunctions and prepositions is unstructured. The possible lack of spread, stability, and structured variation⁶⁵ (Weinreich et al. 1968:187–188) in the conjunction and preposition patterns has the potential to undermine my claim that Lánnang-uè has high degrees of languageness. Another factor that can also weaken this claim involves linguistic independence: the lexical variation observed may have been a consequence of speakers’ high proficiency in Lánnang-uè source languages, via their knowledge of Hokkien, Tagalog, English, and/or Mandarin vocabulary influencing Lánnang-uè vocabulary. If such were the case, then the relationship between source language proficiency and variation could be analyzed as evidence against the hypothesis that Lánnang-uè is highly language-like, as varieties that are highly

⁶⁵ I adopt the view that structured variation is a key feature of language, following Weinreich, Labov, and Herzog (1968:187–188), who claimed that “linguistic structure includes the orderly differentiation of *speakers* and *styles* [emphasis mine] through rules which govern variation in the speech community.”

language-like have vocabularies that tend not to be influenced by surrounding languages (Lipski 2020). A formal investigation conducted on a large set of data from many speakers is needed to ascertain whether the variability observed in the adherence to the conjunction and preposition distributional patterns weakens my argument.

I attempt to address the gap by systematically examining possible spread, stability, structured variation or systematicity, and linguistic independence in the two lexical distributional patterns. Like the earlier chapter (Chapter 4), my first goal is to examine the rates of pattern adoption within the community. This will allow me to approximate the degree of spread of these two patterns. Second, I examine the degree of instability by examining consistency at the individual and group levels. Third, I examine whether the variation is sociolinguistically structured like the variation found in established contact languages such as Singlish (Starr and Balasubramaniam 2019) and Baba Malay (Lee 2014). I do this by attempting to identify language-external factors that condition the potential patterns of variation. Finally, I examine whether the conjunction and preposition patterns of Lánnang-uè are influenced by knowledge of its source languages Hokkien, Tagalog, English, and Mandarin.

The research questions I attempt to answer in this chapter are:

1. How widespread are the conjunction and preposition distributional patterns within Lánnang-uè speakers?
2. How stable are they? In other words, how consistently do individual speakers follow the distributional patterns? And how similar are their patterns of variation to each other?
3. Is the variation structured like variation in established contact languages? Will sociolinguistic factors (e.g., age, sex, attitudes, language proficiency) condition a significant part of the variation?
4. Are the distributional patterns influenced by high proficiency in Lánnang-uè's source languages (or knowledge of their vocabulary)? Will proficiency in the source languages condition the variation in Lánnang-uè conjunctions and prepositions?

The remainder of the chapter is organized as follows: In Section 5.2, I briefly describe the two conjunction and preposition lexical distributional patterns observed in Lánnang-uè. In Section 5.3, I present my hypotheses relevant to the conjunction and preposition distributional patterns.

In Section 5.4, I discuss the approach, datasets, and analyses I used to test my hypotheses. I present the results in Section 5.5. In Section 5.6, I summarize and discuss these results. I conclude this chapter in Section 5.7.

5.2 *The patterns: A summary*

My analysis in Chapter 3 revealed semi-rigid lexical distributional patterns for conjunctions and prepositions. Specifically, for conjunctions, I found that certain conjunctions ‘classes’ – conjunctions that I grouped based on common or shared functions (see Table 26, column 2) – tended to be derived from either Hokkien, Tagalog, or English (henceforth, the conjunction distributional pattern). The pattern can be observed in Table 26, where I list the 17 classes (column 1) I derived from my corpus analysis using a set of function-based criteria I developed using data from a pilot study (column 2). In column 3, I enumerate the most frequently used conjunction tokens in Lánnang-uè that met the criteria for inclusion in that class; then, I identify the source language of these tokens (column 4).

Table 26. Function-based classes of conjunctions in Lánnang-uè, criteria, most frequently used tokens, and source language

Class	Function/criteria for inclusion	Most frequently used tokens in Lánnang-uè	Source language
adversative	introduces a statement that contrasts with or seems to contradict a previous statement	<i>kasò</i> ‘but it is the case’ <i>però</i> ‘but’	Tagalog
cumulative (non-emphatic)	connects constituents that are to be taken collectively, does not emphasize the latter constituent	<i>kâp</i> ‘and’ <i>kiaū</i> ‘and’	Hokkien
cumulative (emphatic)	connects constituents that are to be taken collectively, emphasizes the latter constituent	<i>tsakà</i> ‘and/and also’	Tagalog
disjunctive	presents two or more constituents as alternatives	<i>âsī</i> ‘or’	Hokkien
conditional (general)	introduces any condition	<i>nā</i> ‘if’	Hokkien
conditional (specific)	introduces a particular condition (e.g., event triggers, condition attached to an agreement, etc.)	<i>pagkà</i> ‘as soon as’ <i>kapâg</i> ‘if’ <i>pâg</i> ‘if’ <i>kahitnà</i> ‘even if’ <i>hanggât</i> ‘so/as long as’	Tagalog

conditional (unless)	introduces ‘unless’ conditions – the only circumstances in which an event you are mentioning will not take place or in which a statement you are making is not true	<i>unlêss</i> ‘unless’	English
concession	introduce an idea that is granted in response to the main clause	<i>kahî</i> ‘although’ <i>maskî</i> ‘although’	Tagalog
consequence	introduce clauses of result or decision	<i>kayâ</i> ‘so/that is why’	Tagalog
location	indicates that following constituent contains/describes a reference to a place or situation	<i>kungsaân</i> ‘where’	Tagalog
manner	introduces clauses referring to a manner	<i>tshîntshiū</i> ‘like/as’ <i>nântshiū</i> ‘like/as’ <i>khalâng</i> ‘like/as’ <i>khânân</i> ‘like/as’	Hokkien
reason	links two clauses, where one clause contains/describes the effect and the other, the cause	<i>porkêt</i> ‘just because’ <i>kasî</i> ‘because’	Tagalog
substitution	indicates choice or something is done in place of something else	<i>kaysà</i> ‘instead of’	Tagalog
temporal (‘after’)	indicates that the clause contains/describes an event that is temporally sequenced before the event in the main clause	<i>afèr</i> ‘after’	English
temporal (general)	indicates or comments on temporal sequence, with the exception of conjunctions meaning ‘after’	<i>bagò</i> ‘before’ <i>hanggâng</i> ‘while’ <i>nûng</i> ‘when’ <i>tuwîng</i> ‘whenever’	Tagalog
relativizer/ complementizer (general)	connects the relative clause or verb phrase to the noun phrase or marks an embedded clause as functioning as a complement	<i>nà</i> ‘that’	Tagalog
relativizer (specific)	used to connect the relative constituent of either location, time, or reason to the noun phrase	<i>kungsaân</i> ‘where’ <i>kumbakî</i> ‘why’ <i>nûng</i> ‘when’	Tagalog

For prepositions, I found that the seven preposition classes – derived from clustering prepositions based on functions – tended to be derived from either Hokkien or English, as shown in Table 27. I refer to this as the preposition distributional pattern. I provide examples and discuss the function/criteria in more detail in Chapters 3.8 and 3.9.

Table 27. Function-based classes of prepositions in Lánnang-uè, criteria, most frequently used tokens, and source language

Class	Function/criteria for inclusion	Most frequently used tokens in Lánnang-uè	Source language
accompaniment	denotes ‘in the company of’	<i>wìth</i> ‘with’	English
location (general)	indicates the general area or region in a physical, conceptual, or temporal space the entity is in.	<i>tī</i> ‘at’ <i>ân</i> ‘at’ <i>tiâm</i> ‘at’ <i>tuí</i> ‘at’	Hokkien
location (specific)	indicates a specific position in a physical, conceptual, or temporal space the entity is in.	<i>abovê</i> ‘above’ <i>acrôss</i> ‘across’ <i>agâinst</i> ‘against’ <i>alông</i> ‘along’ <i>amông</i> ‘among’ <i>arouînd</i> ‘around’ <i>befôre</i> ‘before’ <i>behînd</i> ‘behind’ <i>belòw</i> ‘below’ <i>besîde</i> ‘beside’ <i>betweèn</i> ‘between’ <i>ìn</i> ‘in’ <i>insîde</i> ‘inside’ <i>òn</i> ‘on’ <i>outsîde</i> ‘outside’ <i>ovèr</i> ‘over’ <i>throug</i> h ‘through’ <i>undèr</i> ‘under’ <i>withìn</i> ‘within’	English
‘of’	expresses a correlative, meronymic, associative, or possessive relationship between the entity and another entity	<i>ôf</i> ‘of’	English
orientation	denotes ‘orienting with’, ‘with regard to’, or ‘concerning’	<i>hiông</i> ‘towards’ <i>tuí</i> ‘towards’	Hokkien
range/path	marks the point in space or time at which something starts or ends	<i>ân</i> ‘from’ <i>kaù</i> ‘to’ <i>tuí</i> ‘from’	Hokkien
temporal	indicates ‘at or during a time earlier or later than the time or event mentioned’	<i>befôre</i> ‘before’ <i>aftèr</i> ‘after’	English

5.3 *Hypotheses*

I have four hypotheses regarding the two lexical patterns described in Section 5.2, anchored on the proposition that Lánnang-uè has high degrees of languageness:

1. Spread. There will be high rates of pattern adoption within the community for both conjunction and preposition distributional patterns. Most speakers will have at least one conjunction/preposition that followed these patterns, deriving particular classes of conjunctions/prepositions from certain source languages.
2. Stability. There will be high rates of stability for both patterns. Speakers who follow the patterns at all will do so with high degrees of consistency. They will also have patterns of variation that will not differ too much from each other.
3. Structured variation/systematicity. The variation will be conditioned by at least one of the following sociolinguistic factors: age, sex, and/or attitudes.
 - a. If the variation involves change, many of the tokens that do not conform to the patterns will come from younger speakers and female speakers.
 - b. Many of the tokens that do not conform to the conjunction and preposition patterns will come from speakers who viewed Lánnang-uè as ‘broken’ and/or not reflective of their hybrid Lannang identity.
4. Independence from source languages. The patterns of variation will not be influenced by high proficiency in Lánnang-uè’s source languages (or knowledge of conjunctions and prepositions in these languages).

The first and second hypotheses (spread and stability) were motivated by previous studies on Lánnang-uè (Gonzales 2018; Gonzales and Starr 2020) as well as many features I described in Chapter 3 and Chapter 4, where I found high degrees of spread and stability for many features/patterns. In the previous chapter (Chapter 4), for example, I investigated tone and stress features in the variety and found that most speakers consistently have these features. The

findings suggested that Lánnang-uè is highly language-like. If this is true, then, I would also expect to see high degrees of spread and stability in features/patterns in other domains such as the lexicon (e.g., conjunction and preposition lexical patterns), as established languages generally have high rates of spread and stability across their features/patterns.

The third hypothesis was motivated by findings in previous studies (Gonzales 2018; Gonzales and Starr 2020) and what I discovered in Chapter 4, where I found the conditioning effects of sociolinguistic factors on variation. In Chapter 4, for example, I found that a significant portion of tokens that had non-phonemic pitch instead of (highly widespread and stable) lexical tone came from speakers who (wanted to) claim proficiency in languages that do not have lexical tone (i.e., Tagalog and English). These findings on sociolinguistically-conditioned variation – systematicity – are suggestive of high degrees of languageness in Lánnang-uè. Speakers, at least for some features (Chapter 4), use variation as a tool to express different social meanings; they systematically use (pattern-non-conforming) variants in specific social conditions. If Lánnang-uè is indeed highly language-like, then the variation in the two lexical patterns should be systematically conditioned by sociolinguistic factors, similar to the variation found in other Lánnang-uè patterns, as language-like varieties tend to exhibit systematic or structured variation across its features/patterns (Weinreich et al. 1968; Ghyselen and De Vogelaer 2018).

The hypothesized directions of the effects were motivated by sociolinguistic theories.

Age and sex. Sociolinguistic research has shown that these can be robust predictors of variation (Eckert 1989; Sankoff 2006; Maclagan et al. 1999). Particularly in the context of sound change, scholars have often reported that young and female speakers were more likely than old and male speakers to lead change. In the context of Lánnang-uè, evidence for innovation led by young and female speakers has been observed at the acoustic level (Gonzales and Starr 2020), where Starr and I found that all speakers adopted a vowel system where all vowels sound similar regardless of the language source, but young and female speakers occasionally produced certain vowels (e.g., [ʊ]) differently depending on the source language. Given documented patterns, it was reasonable to hypothesize that a sizable portion of the tokens that do not conform to the conjunction and preposition patterns will come from speakers who are young and female, assuming of course that these tokens are innovative.

However, not all variability involves change. There may be other reasons, such as engaging in age- or sex-group-specific stylistic practice, projecting particular group-specific social identities (Eckert 1989:245; Eckert and McConnell-Ginet 2003), or group-specific linguistic exposure (Shin 2013). For example, certain male groups in multilingual contexts, such as bilingual Latin American speakers in New York, have been reported to have significantly lower rates of subject pronoun use and increased sensitivity to switch-reference in Spanish compared to women, partially due to their decreased exposure to English (Shin 2013). I kept this possibility in mind when I formulated my hypotheses for age and sex. This meant that I was also open to the idea that a significant part of pattern-non-conforming tokens will not be associated with young and female speakers.

Negative attitudes towards Lánnang-uè. Apart from age and sex, I also hypothesized that a significant part of the pattern-non-conforming tokens will come from speakers who have negative attitudes towards the variety. Sociolinguistic research has shown that language attitudes can influence the (non)adoption of linguistic features (Thomason 2007:49; Borbély 1995:319). For instance, citing Kay Williamson (p.c. 1996), Thomason (2007:49) reported that Ibani speakers in southern Rivers State in Nigeria – bilingual in Ibani and Igbo – were “concerned to maintain the purity of Ibani”, indicating that these speakers have negative attitudes towards mixing Ibani with other languages. She then reported that these speakers carefully and deliberately avoided Igbo loanwords in Ibani when outside scholars elicited Ibani wordlists from the Ibani speakers, illustrating a case where negative attitudes towards language mixing influenced the use of vocabulary. It is possible that speakers’ attitudes towards the variety Lánnang-uè could affect lexical choice as well. Specifically, I hypothesized that speakers who regarded Lánnang-uè as ‘broken’ language or not reflective of their hybrid Lannang identity will be less likely to conform to the lexical patterns. Those who did not view it as ‘broken’ or ‘unemblematic of the Lannang identity’ will be more likely to adhere to the distributional patterns compared to those who did.

The final hypothesis (henceforth, Hypothesis 4) was motivated by my observation that varieties characterized as ‘languages’ or, in my terms, ‘highly language-like’ tend to have linguistic patterns that are not influenced by the patterns of other languages (e.g., Topo and Ugsha varieties of Media Lengua) (Lipski 2020). High proficiency in a source language (i.e., Spanish) did not influence the linguistic patterns of these two language-like varieties (see Chapter 4.3 for a more in-depth discussion). If Lánnang-uè has high degrees of languageness like the Topo and Ugsha varieties of Media Lengua, then I expect the patterns of variation in Lánnang-uè conjunctions and prepositions not to be influenced by proficiency in its source languages, or knowledge of conjunctions and prepositions in these languages. For example, a significant chunk of the conjunctions with adversative function derived from Hokkien – which do not conform to the distributional pattern discussed in Section 5.2 – will not be traced back to speakers who are highly proficient in Hokkien (speakers who have knowledge of adversative conjunctions in Hokkien).

5.4 Methodology

5.4.1 Approach and dataset preparation

To test my hypotheses, I adopted a corpus-based, computational approach. I analyzed the lexicon of Lánnang-uè, discarding all information about the order or structure of words, similar to what is done in “bag-of-words” representations of text (Goldberg 2017:69; Zhang et al. 2010). I first extracted all Lánnang-uè sentences from the 375,000-word Lannang Corpus (LanCorp) (Gonzales 2022a) which have all been tagged with social information about the speaker (i.e., age, sex, language proficiency, language attitudes towards Lánnang-uè).

After extraction, I tagged each word in all sentences for part-of-speech (e.g., conjunction, preposition) using a tagger program⁶⁶ I created in the Python environment. After tagging, I tokenized the sentences – I broke down tagged sentences into tagged words, ignoring the

⁶⁶ My program utilizes Conditional Random Fields (CRF). The model used is trained using 1,085 manually annotated Lánnang-uè sentences. It has a cross-validated (k-folds = 5) accuracy score of 0.83 (SD = 0.005), precision score of 0.58 (SD = 0.017), recall score of 0.56 (0.018), and an f-1 score of 0.56 (SD = 0.015). One feature of a CRF model is that it ‘learns’ the distributions from sequential data. It can identify the optimal part-of-speech of a token, given the context.

context.⁶⁷ I then tagged each word for source language by relying on a combination of rule-based and manual tagging approaches. I used publicly available English, Tagalog, and Mandarin wordlists to help me tag English-, Tagalog-, and Mandarin-origin Lánnang-uè words. Lánnang-uè words that are not found in any of the three wordlists are preliminarily tagged as Hokkien-sourced. I asked three native speakers of Lánnang-uè to go over the list and revise incorrectly tagged tokens. I also asked them to tag words that do not have a clear origin as ‘unclear’.

After tokenization, I extracted words tagged as conjunctions and prepositions from the main dataset to create two datasets – one for conjunctions and another for prepositions. To ensure that all the tokens in the dataset are indeed conjunctions and prepositions, I asked three native speakers of Lánnang-uè to go over the list and remove tokens that are incorrectly tagged as conjunctions and prepositions. After that, I manually coded each token by conjunction and preposition class using the criteria mentioned in section 5.2. I enumerate the classes below:

⁶⁷ In the process of tokenization, I disregarded word order information as well as the context in which the words are produced, which could be useful for identifying morphosyntactic and semantic factors that may or may not condition lexical choice, lexical patterns, or variants. In this chapter, I am only interested in the potential general association between certain lexical classes concerning source language (e.g., adversative conjunctions derived from Hokkien). I am not interested in the possible morphosyntactic and semantic factors that could condition the use of one Hokkien-sourced variant over another Hokkien-sourced variant, for example. Hence, the decision to tokenize the sentences into words. The primary reason for doing this is to simplify my analysis – a “parsimonious” approach (Daganzo et al. 2012:47).

Conjunction classes

- | | |
|------------------------------|--|
| 1. Adversative | 11. Manner |
| 2. Cumulative (non-emphatic) | 12. Reason |
| 3. Cumulative (emphatic) | 13. Substitution |
| 4. Disjunctive | 14. Temporal ('after') |
| 5. Conditional (general) | 15. Temporal (general) |
| 6. Conditional (specific) | 16. Relativizer/Complementizer
(general) |
| 7. Conditional ('unless') | 17. Relativizer/Complementizer
(specific) |
| 8. Concession | |
| 9. Result /Consequence | |
| 10. Location | |

Preposition classes

1. Accompaniment
2. 'Of'-type
3. Location (specific)
4. Location (general)
5. Orientation
6. Range/path
7. Temporal
8. Spatial

Then, for both datasets, I coded each token for adherence to the distributional pattern. A token was marked as '1' if the token conformed to the distributional patterns described in Section 5.2 and marked '0' if not. For example, in the conjunction dataset, a token that was coded 'Hokkien' for source language and 'adversative' for type was marked '0'. However, if that token was coded 'Tagalog', then that token was marked '1'.

Overall, each token in my datasets was coded for:

1. source language (categorical)
2. class of conjunction/preposition (categorical)
3. age (continuous)
4. sex (categorical)
5. z-scored self-reported language proficiency (Tagalog, English, Hokkien, and Mandarin) (continuous)
6. attitudes towards Lánnang-uè as a broken language (continuous)
7. attitudes towards Lánnang-uè as emblematic of the hybrid Lannang identity (continuous)
8. participant (i.e., the identification number) (categorical)
9. adherence to the distributional pattern (categorical)

My coded conjunction and preposition datasets have 15,901 and 8,134 tokens, respectively. These datasets were used for my descriptive analyses (e.g., frequency, proportion, and standard deviation) and regression analyses, which I discuss in Section 5.5.

I wanted to get a more nuanced picture of potential sociolinguistic motivations for following (or not following) the patterns, something that the two coded datasets mentioned earlier cannot provide. As such, I created another dataset that contained transcribed interviews from 77 speakers, all of whom contributed to the lexical datasets mentioned earlier. These interviews focused on questions about Lannang community, identity, and language (Appendix D). I asked questions, for example, inquiring about linguistic features that distinguish them from non-Lannang speakers, or what they think about a certain feature. The qualitative dataset complemented my quantitative examination of the relationship of sociolinguistic factors to linguistic behavior.

5.4.2 Analytical method

5.4.2.1 Descriptive analyses

For each of the coded datasets, I conducted descriptive analyses. I measured the rates of pattern adoption within the community (henceforth, ‘spread’) and degree of stability by examining the coded factor ‘adherence to the distributional pattern’ using three measures (see Section 4.4.5). Furthermore, I conducted analyses of tokens that did not conform to the patterns. I provide a breakdown of these tokens.

5.4.2.2 Regression analyses

After this, I conducted regression modeling on the lexical datasets to test for potential conditioning effects of the hypothesized factors on lexical choice and adherence to the lexical distributional patterns. Conducting regression analyses allows me to single out the (main or interaction) effects of particular factors and test for correlations between these factors and the dependent variables. Specifically, I attempted to run five generalized linear mixed-effects models with logistic link functions on each the two lexical datasets in the R environment (R Core Team 2015).

To test my hypotheses on the conjunction and preposition distributional patterns, I fitted generalized linear mixed-effects models with logistic link functions on the two datasets (i.e., the first models). The dependent variable is adherence (coded ‘1’ for adhering and ‘0’ for non-adhering). The predictors I included were age (younger vs. **older**), sex (male vs. **female**), proficiency in the source languages of Lánnang-uè,⁶⁸ and language attitudes. The reference level, or the level to which the other level is compared, is indicated in boldface. Random intercepts for participant were included in these models. Their inclusion gives me some statistical license to generalize my findings to the true population of Lánnang-uè speakers (Konstantopoulos and Hedges 2019:278).

To get a finer-grained picture of potential sociolinguistic patterns, I also fitted generalized linear mixed-effects models with logistic link functions (i.e., the rest of the five models for each dataset) where the dependent variable was source language, specifically, a binary coding of the source languages (i.e., Hokkien vs. Tagalog/English/Mandarin, English vs. Tagalog/Hokkien/Mandarin, Tagalog vs. English/Hokkien/Mandarin, and Mandarin vs. Hokkien/Tagalog/English). In models where a distributional pattern is involved (e.g., model of likelihood to derive the conjunction from Hokkien), the predictors included classes (part vs. **non-part**) (e.g., part/not a part of classes of conjunctions supposed to be from Hokkien), age (younger vs. **older**), sex (male vs. **female**), and self-reported linguistic proficiency in the language tested and the languages not tested (e.g., Hokkien proficiency, proficiency in other

⁶⁸ I created the factor ‘proficiency in the source languages’ by running Principal Components Analysis or PCA on the z-scored Tagalog, English, Hokkien, and Mandarin proficiency scores and getting the component that is positively correlated with the four scores.

source languages that are not Hokkien), with the reference levels indicated in boldface. In these models, interactions between classes and the social factors were modelled in to test whether the social factors condition the variation in adherence to the lexical distributional patterns. In models that did not involve a distributional pattern (e.g., the model of likelihood of deriving the preposition from Tagalog), only the non-interaction predictors were modelled in. In all models, I included random intercepts (participant) when possible. I did not model in language attitudes in these models to avoid overfitting.

In all my regression models, the categorical predictor variables were analyzed after (re)coding the variables using unweighted effect contrast coding conventions (i.e., 1 vs. -1) (Sonderegger 2022).

5.4.2.3 Criteria for hypothesis testing

My hypotheses on spread and stability (Hypotheses 1 and 2) will be supported if I find evidence of them in my data. Regarding my descriptive analyses, if the feature spread scores for the conjunction and preposition patterns are above average (i.e., 0.5, or more than half of the population), then my hypothesis on spread will be supported. If the lexical patterns have mean intraspeaker feature consistency scores (as measured in Section 4.4.5) that are higher than 0.5 (i.e., the patterns were followed more than 50% of the time, on average), then my hypothesis on stability will be supported. It will be further supported if I find interspeaker pattern inconsistency scores that are below 0.5 (i.e., the patterns of variation among speakers have heterogeneity levels below 50%) (Section 4.4.5).

In my regression analyses, if ‘classes’ have an effect on the dependent variables, then my hypotheses on spread and stability (Hypotheses 1 and 2) will be supported as well, as I interpret the presence of a structural effect on the dependent variable as evidence of both spread and consistency. In regression models, the effect of a specified predictor variable becomes statistically significant if there are consistent correlations between the specified predictor (e.g., structural factors) and the dependent variable for many participants.

My hypothesis on structured variation (Hypothesis 3) will be supported if I find effects of age, sex, and/or language attitudes on adherence to the distributional patterns in my general models or if I find interaction effects between the hypothesized sociolinguistic variables and ‘classes’ on the dependent variable – ‘source language’ – in my finer-grained regression models.

I did not directly model ‘variation’; instead, I interpreted the interaction effects as the potential conditioning effect of a sociolinguistic variable on the lexical patterns (e.g., the potential conditioning effect of age on the relationship between lexical classes and source language).

My hypotheses on the direction of the effect (e.g., younger speakers tending to produce pattern-non-conforming tokens) will be supported if I find the expected pattern in an examination of the marginal effects – defined as “predictions generated by a model when one holds the non-focal variables constant and varies the focal variable(s)” (Lüdecke 2018a:1; Lüdecke 2018b) or the effect the individual predictors have on the dependent variable while all other variables are held constant. I used the ‘ggeffects’ package in the R environment to compute the estimated marginal means (predicted values) for the dependent variable at the margin of specific values or levels from certain model terms (Lüdecke 2018b).

My hypotheses on linguistic independence (Hypothesis 4) will not be supported if I find evidence of negative correlations between the proficiency variables and the ‘class’ variable on the dependent variable in my regression models (i.e., high proficiency linked to use of pattern-non-conforming tokens).

5.4.3 *Measuring spread and stability*

I approximated the degree of spread by looking at rates of pattern adoption within the community. I measured the degree of stability by examining consistency rates at the individual and group level. Specifically, I relied on three measures, discussed in detail in Chapter 4.4.5.

1. *Spread* – What proportion of my speakers follows the pattern at all?
2. *Mean intraspeaker consistency* – How often/consistently do individual speakers follow the pattern?
3. *Interspeaker pattern inconsistency* – How inconsistent are the patterns of variation between speakers?

To recapitulate, the following formulas were used in this chapter:

Spread score = *number of speakers who followed the pattern at least once / number of all speakers*

Individual intraspeaker consistency score = *number of tokens where an individual speaker who followed the pattern / number of tokens where the pattern could be followed for that individual*

Mean intraspeaker consistency score = *individual intraspeaker consistency scores / number of individuals*

Interspeaker pattern inconsistency score = *standard deviation of all individual intraspeaker feature consistency scores / mean of these scores*

5.4.4 Speakers

The conjunction and preposition data came from a total of 135 speakers. All were born and raised in the Philippines, spoke Lánnang-uè, and had at least some knowledge of Tagalog, English, Mandarin, and Hokkien. All individuals were recruited via social media or word of mouth. A breakdown of all speakers by self-reported sex and age group is provided in Table 28.

Table 28. distributional pattern of speakers by self-reported sex and age group

Sex	Age Group									Total
	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99	
Female	1	13	8	6	12	8	7	14	1	70
Male	0	14	8	7	6	13	11	6	0	65
Total	1	27	16	13	18	21	18	20	1	135

The average *z*-scored proficiency levels of these speakers are as follows: Hokkien (mean = 0.27, SD = 0.80), Mandarin (mean = 0.01, SD = 0.74), Tagalog (mean = 0.10, SD = 0.68), English (mean = 0.45, SD = 0.75). A negative score indicates low proficiency, a positive score indicates high proficiency, and a score close to zero indicates average proficiency. Eight speakers were present in all three variationist studies in this dissertation.

5.5 Results

5.5.1 The conjunction lexical distributional pattern

5.5.1.1 Descriptive analyses

Most (119 out of 135) of the speakers in my sample who used conjunctions produced at least one conjunction that adhered to the distributional pattern. That is, they did one of the following at least once:

- i. derived their disjunctive, non-emphatic cumulative, general conditional, and manner conjunctions from Hokkien,
- ii. derived conditional conjunctions meaning ‘unless’ and the time conjunction meaning ‘after’ from English, or
- iii. derived the rest of the other conjunctions from Tagalog.

The spread score is 0.8815. Most of my speakers followed the conjunction pattern. Analyzing the data by class, I found that each conjunction class had varying rates of spread (Table 52 in Appendix G). The spread scores by class were consistently above 0.5 (mean = 0.8553, SD = 0.1175).

Not all the speakers who adhered to the conjunction distributional pattern at least once always were 100% consistent. Out of the 119 speakers who adhered to the distributional pattern at least once, eight adhered less than 50% of the time (mean = 29.62%, SD = 0.13), 36 adhered 50% to 74.99% of the time (mean = 64.91%, SD = 0.076), 28 adhered 75% to 89.99% of the time (mean = 83.63%, SD = 0.04), 22 adhered almost all the time (90% to 99.99% of the time) (mean = 93.8%, SD = 0.028), and 25 always followed the distributional pattern and did not vary. As the boxplot and histogram in Figure 15 indicate, the bulk of speakers either completely or occasionally adhered to the pattern.

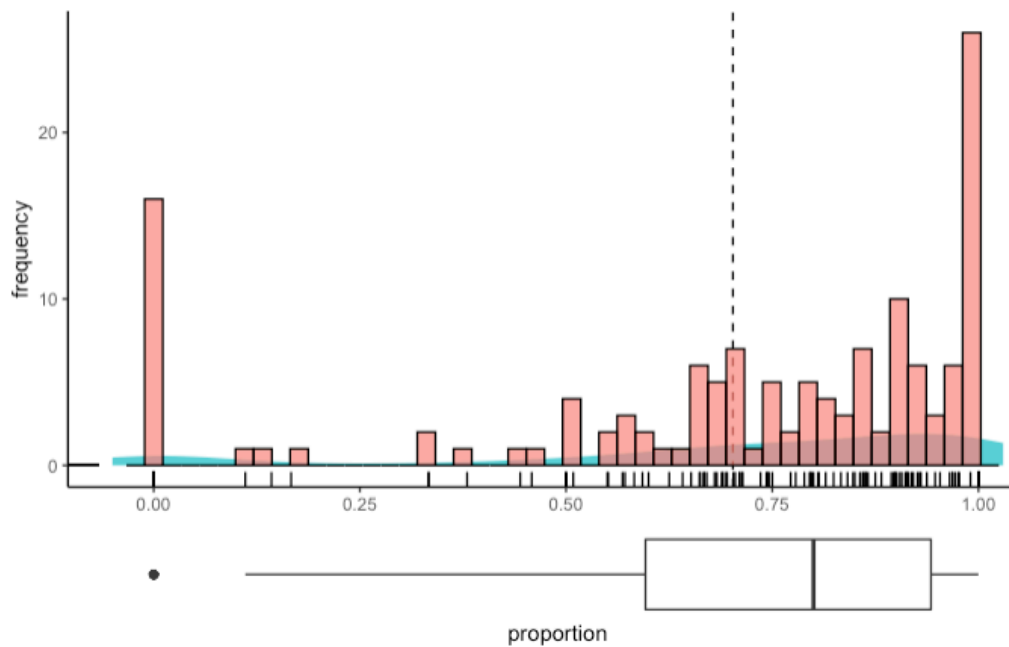


Figure 15. Histogram (frequency), density, and box plot of proportions (adherence to conjunction distributional pattern); broken line indicates mean

The mean intraspeaker consistency score for *all* participants is 0.7967 (SD = 0.1998). The interspeaker pattern inconsistency score is 0.2508. I also report these scores by class, summarized in Table 53 in Appendix G. The scores – averaged across conjunction classes – are 0.878 (SD = 0.1268) and 0.2238 (SD = 0.195), respectively. Overall, adherence to the conjunction pattern is highly stable.

Which tokens did not conform to the conjunction distributional pattern discussed in Section 5.2? I summarize the results of my analyses of pattern-non-conforming tokens for each class in Table 29. The first column of the table indicates the ‘class’, as defined in Section 5.2. Column two lists all of the pattern-non-conforming tokens that fit the criteria for that class – the conjunctions that were unexpectedly sourced from other languages (column three). The fourth column lists the total number of non-conforming tokens and the last column contains the relative frequency of these tokens in relation to the total number of tokens in that particular class: the percentage of pattern non-conformance. For instance, in the first row, I report that the non-conforming conjunctions in the adversative class are English-derived *bût* ‘but’ as well as Hokkien-derived *tānsī* ‘but’, *umkû(h)* ‘but’, and *umkô(h)* ‘but’. There are 92 English-sourced

conjunction tokens and 999 Hokkien-sourced ones in the adversative class; they comprise 8.43% and 91.5% of the non-conforming tokens in their class.

Table 29. Distribution of conjunction tokens that did not conform to the conjunction distributional pattern, by class and source language

Class	Tokens	Source language	n	Percentage of non-conformance (n/ number of total n in the class X 100)
Adversative	<i>bût</i>	English	92	8.43%
	<i>tānsī, umkû(h), umkô(h)</i>	Hokkien	999	91.57%
Cumulative (non-emphatic)	<i>ând</i>	English	226	74.35%
	<i>ât</i>	Tagalog	78	25.65%
Cumulative (emphatic)	<i>adiaû</i>	Hokkien	12	21.81%
	<i>and thèn</i>	English	42	76.37%
	<i>ranhoù</i>	Mandarin	1	0.02%
Disjunctive	<i>òr</i>	English	141	100%
Conditional (general)	<i>îf</i>	English	53	34.64%
	<i>kûng</i>	Tagalog	99	64.70%
	<i>rúguo</i>	Mandarin	1	0.66%
Conditional (specific)	<i>tâng</i>	Mandarin	1	6.67%
	<i>kîdién</i>	Hokkien	14	93.33%
Conditional ('unless')	<i>tūhuī</i>	Hokkien	10	100%
Concession	<i>even though</i>	English	6	8.11%
	<i>suidién</i>	Hokkien	68	91.89%
Result	<i>inwi âni</i>	Hokkien	1	100%
Location	NA	NA	NA	NA
Manner	<i>katulâd</i>	Tagalog	7	4.70%
	<i>lîke, similar tò</i>	English	142	95.30%
Reason	<i>becaûse</i>	English	67	9.01%
	<i>înwī, în-uī</i>	Hokkien	677	90.99%
Substitution	<i>instead ôf</i>	English	10	100%
Temporal ('after')	<i>pagkatapôs</i>	Tagalog	1	100%
Temporal (general)	<i>beforè, until, whenever, while</i>	English	31	96.87%
	<i>tâng</i>	Mandarin	1	3.13%
Relativizer/ Complementizer (general)	No deviant tokens in sample	NA	NA	NA
Relativizer/ Complementizer (specific)	NA	NA	NA	NA

5.5.1.2 Regression analysis

My model of adherence to the conjunction distributional pattern, summarized in Table 30, shows a main effect of age and ‘broken’ attitudes towards Lánnang-uè. There are no main effects of sex, proficiency in the source languages, or ‘emblematic’ attitudes. Only age and ‘broken’ attitudes towards the variety conditioned the patterns of variation – age and ‘broken’ attitudes can reliably predict a speaker’s adherence to the conjunction distributional pattern.

Table 30. Regression results – adherence to the conjunction distributional pattern (observations = 15,769, $R^2 = 0.209$, random intercepts for speaker).

Predictors	Log-Odds	SE	CI	<i>p</i>
(Intercept)	2.16	0.31	1.55 – 2.77	< 0.001
Age (younger vs. older)	-0.49	0.21	-0.91 – -0.07	0.023
Sex (male vs. female)	-0.11	0.21	-0.30 – 0.51	0.605
Proficiency (source languages)	-0.2	0.24	-0.68 – 0.28	0.416
Attitudes towards Lánnang-uè: ‘broken’	-0.11	0.06	-0.23 – -0.00	0.048
Attitudes towards Lánnang-uè: ‘emblematic of Lannang identity’	-0.02	0.09	-0.18 – 0.15	0.85

An examination of the marginal means reveals the direction of the two sociolinguistic effects. A sizable portion of tokens that did not adhere to the conjunction pattern can be traced back to speakers who are young and those who strongly believe that Lánnang-uè is ‘broken’ (Figure 16). This regression model predicts that speakers will be less likely to follow the pattern if they are young and view Lánnang-uè as ‘broken’.

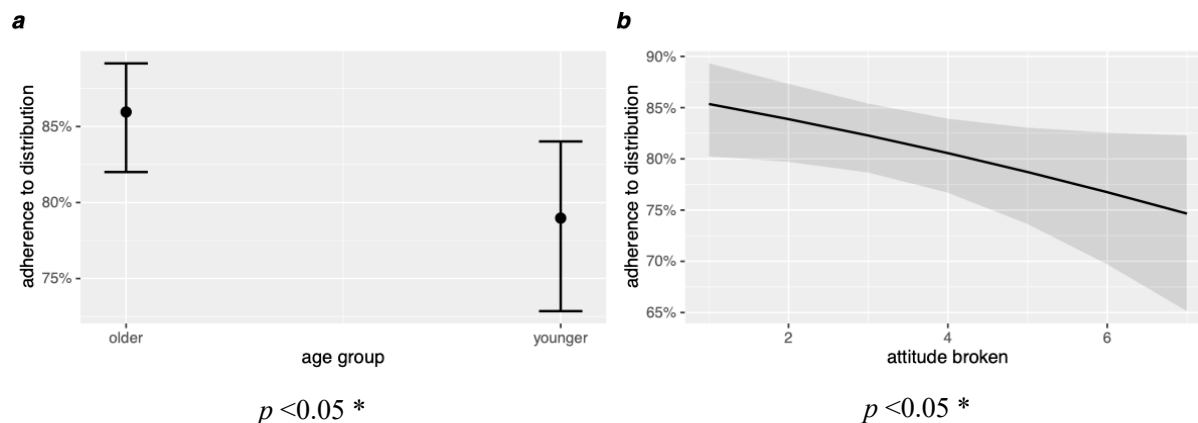


Figure 16. Marginal means/effects of sociolinguistic factors on likelihood to adhere to the distributional pattern

In Section 5.3, I hypothesized that most speakers will consistently use, or at least tend to use, Hokkien-derived conjunctions for certain types or classes (i.e., non-emphatic cumulative, disjunctive, general conditional, and manner), in accordance with the pattern described in Section 5.2. There will be high rates of spread and stability. I also hypothesized that a significant part of the non-conforming Hokkien-derived conjunctions will come from speakers who report being highly proficient in Hokkien.

After modeling the likelihood of deriving conjunctions from Hokkien (Table 54 in Appendix G), I found that ‘class’ reliably predicted the use of Hokkien-derived variants. If the conjunction is part of the classes of conjunctions that are expected to be derived from Hokkien (see Section 5.2), the Hokkien-derived variant is more likely to be used; if the conjunction is not part of the classes, the Hokkien-derived variant is less likely to be used (Figure 17a). In other words, most speakers derived certain conjunction classes from Hokkien with high levels of consistency.

Apart from spread and stability, the Hokkien model also showed that high Hokkien proficiency did not account for Hokkien-derived conjunction tokens that did not follow the distributional pattern, but age and sex did. This is supported by the lack of interaction effects between ‘class’ and ‘Hokkien proficiency’ and the presence of interaction effects between ‘class’ and ‘age’/‘sex’. Examining the marginal effects of the interaction terms, I found that younger and male speakers were more likely than older and female speakers to use Hokkien-derived variants for conjunctions that were not supposed to be derived from Hokkien. This is shown in Figure 17b and 3c, where the red bar is noticeably higher for younger and male speakers than

older and female speakers. I also found that female speakers were less likely than male speakers to use Hokkien-derived variants for conjunctions that are expected to be derived from Hokkien – in Figure 17c, the blue bar is noticeably lower for females than males.

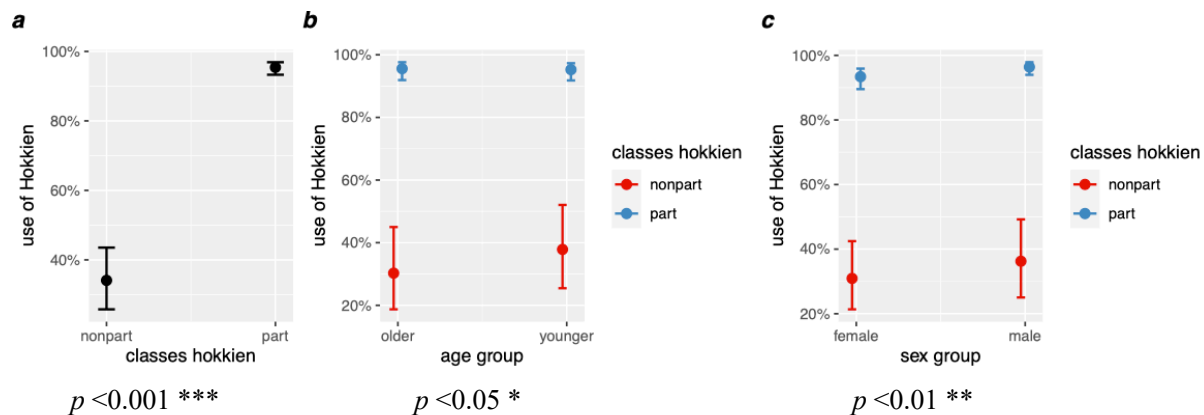


Figure 17. Marginal means/effects of sociolinguistic factors on likelihood to adhere to the distributional pattern (Hokkien-derived conjunctions)

In Section 5.3, I hypothesized that speakers will consistently use, or at least tend to use, Tagalog-derived conjunctions for certain classes (e.g., adversative, see Section 5.2). I also hypothesized that a significant portion of the Tagalog-derived conjunctions that did not conform to the pattern will come from speakers who report being highly proficient in Tagalog.

After modeling the likelihood of deriving conjunctions from Tagalog (Table 55 in Appendix G), I found that ‘class’ or conjunction type reliably predicted the use of Tagalog-derived conjunction variants. (Figure 18a). In other words, most speakers consistently derived the classes expected to be derived from Tagalog (Section 5.2) from Tagalog.

My model also indicated that Tagalog proficiency conditioned the increased use of conjunctions derived from Tagalog that were not expected to be derived from the language. This is supported the presence of interaction effects between ‘class’ and proficiency in Tagalog (Table 55 in Appendix G). An examination of the marginal effects showed that those who reported being highly proficient in Tagalog were statistically more likely than those with average or low proficiency to use Tagalog-derived variants for conjunctions that were not expected to be derived from Tagalog (Figure 18b, red).

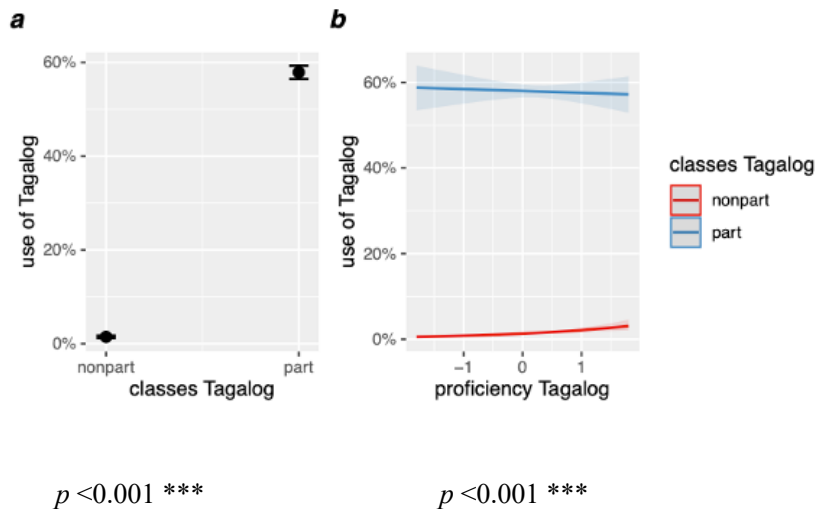


Figure 18. Marginal means/effects of proficiency on likelihood to (not) adhere to the distributional pattern (Tagalog-derived conjunctions)

In Section 5.3, I hypothesized that speakers will consistently use or tend to use English-derived conjunctions for certain classes (e.g., conjunctions meaning ‘after’, see Section 5.2). I also hypothesized that a significant portion of the English-derived conjunctions that did not conform to the pattern will come from speakers who report being highly proficient in English.

After modeling the likelihood of deriving conjunctions from English (Table 56 in Appendix G), I found that ‘class’ or conjunction type reliably predicted the use of English-derived variants. In other words, most speakers consistently derived the classes expected to be derived from English (Section 5.2) from English.

The English model also showed that high English proficiency did not account for English-derived conjunction tokens that did not follow the distributional pattern, but sex did. This is supported by the lack of interaction effects between ‘class’ and English proficiency as well as the presence of interaction effects between ‘class’ and ‘sex’. Examining the marginal effects of the interaction terms, I found that female speakers were statistically more likely to use English-derived variants for conjunctions that were not expected to be derived from English compared to male speakers.

A model for Mandarin-derived conjunction use was not created due to the small ($n = 4$) number of tokens.

5.5.2 The preposition lexical distributional pattern

5.5.2.1 Descriptive analyses

All 91 speakers in my sample who used prepositions produced at least one preposition that adhered to the pattern. They did one of the following at least once: they (1) derived the prepositions of orientation, location, and range/path from Hokkien or (2) derived the prepositions of accompaniment, spatial relations, temporal relations as well as prepositions meaning ‘of’ from English. The spread score for the pattern is 1. Analyzing the scores by class, I found that each class had varying rates of spread (Table 57 in Appendix G), and that these rates were consistently above 0.5, the average (mean = 0.904, SD = 0.14).

The 91 speakers who followed the preposition pattern at least once were not always 100% consistent. Out of the 91, only one adhered to the pattern less than 50% of the time (score = 33.33%), 11 adhered to the pattern 50% to 74.99% of the time (mean = 68.99%, SD = 0.04), 46 adhered 75% to 89.99% of the time (mean = 83.96%, SD = 0.04), 31 almost always adhered to the pattern, following it 90% and 99.99% of the time (mean = 93.56%, SD = 0.028), and two always adhered to it and did not vary at all. As shown in Figure 19, the bulk of speakers either completely or occasionally adhered to the pattern.

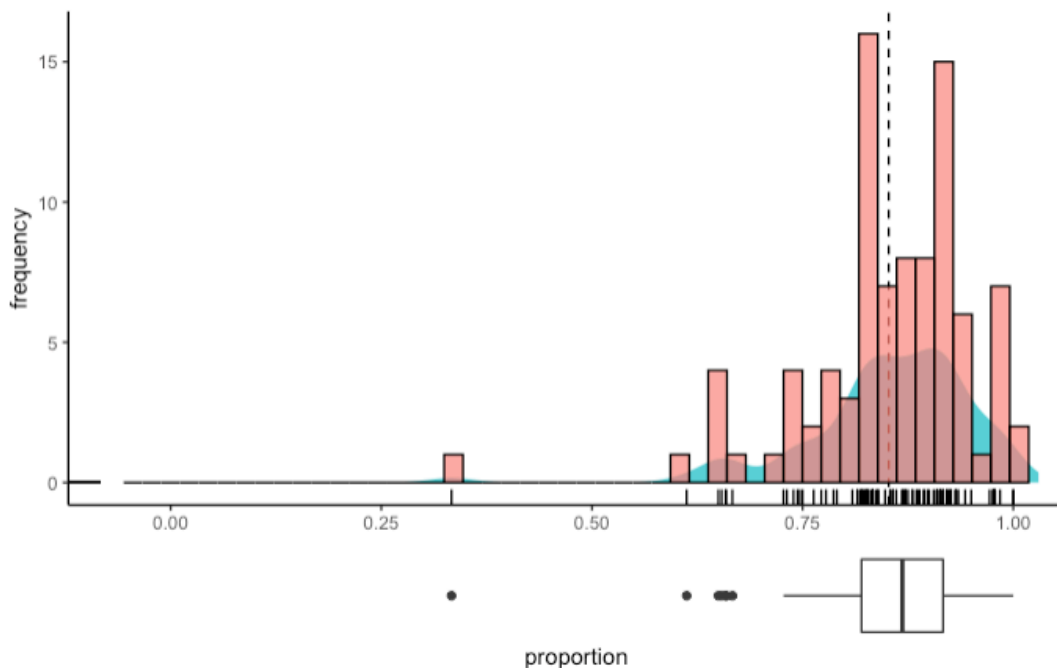


Figure 19. Histogram (frequency), density, and box plot of proportions (adherence to preposition distribution pattern); broken line indicates mean

The mean intraspeaker consistency score for *all* participants is 0.8583 (SD = 0.1017). The interspeaker pattern inconsistency score is 0.1193. I also report these scores by class, summarized in Table 58 in Appendix G. The scores – averaged across preposition classes – are 0.9122 (SD = 0.1134) and 0.1445 (SD = 0.1414), respectively. In sum, the preposition pattern is highly stable.

I provide a breakdown of the tokens that did not conform to the pattern in Table 31. The first column of the table indicates the class. Column two lists all of the non-conforming preposition tokens that fit the criteria for that ‘class’ (see Section 5.2) – the prepositions that were unexpectedly sourced from other languages (column three). The fourth column indicates the frequency of non-conforming tokens and the last column contains the relative frequency of these tokens in relation to the total number of non-conforming tokens in that particular class: the percentage of pattern non-conformance. For example, in the third row, I report that the non-conforming prepositions in the general location class are Tagalog-derived *sà* ‘at’, English-derived *ât* ‘at’, and Mandarin-derived *tsìn* ‘near’ and *tsai* ‘at’. Most of the non-conforming tokens in this class are prepositions derived from Tagalog, at 66.02%.

Table 31. Distribution of preposition tokens that did not conform to the preposition distributional pattern, by class and source language

Class	Tokens	Source language	n	Percentage of non-conformance (n/ number of total n in the class X 100)
Accompaniment	<i>kasamà</i>	Tagalog	1	100%
‘Of’	<i>ng</i>	Tagalog	52	100%
General location	<i>ât</i>	English	69	33.01%
	<i>sà</i>	Tagalog	138	66.02%
	<i>tsìn, tsai</i>	Mandarin	2	0.96%
Orientation	<i>tuei</i>	Mandarin	6	75%
	<i>towârd</i> s	English	2	25%
Range/path	<i>fròm, tò</i>	English	588	98.82%
	<i>galìng, hanggâng</i>	Tagalog	7	1.18%
Temporal	<i>bagò</i>	Tagalog	5	100%
Specific location	<i>gitná, loób,</i>	Tagalog	2	0.75%
	<i>ebìn, kèkhi, laibìn, thaukè, tiengbin, tuitioh, uìtioh</i>	Hokkien	263	99.25%

5.5.2.2 Regression analysis

None of the factors I hypothesized – age, sex, language proficiency, and attitudes – had an effect on adherence to the preposition distributional pattern (Table 32). The variation in general adherence to the preposition pattern cannot be accounted for by these factors.

Table 32. Regression results – adherence to the preposition distributional pattern (observations = 8,134, $R^2 = 0.08$, random intercepts for speaker). Reference levels are highlighted in boldface; in the *p*-values column, statistically significant values are in boldface.

Predictors	Log-Odds	SE	CI	<i>p</i>
(Intercept)	2.18	0.22	1.74 – 2.61	<0.001
Age (younger vs. older)	-0.1	0.15	-0.39 – 0.19	0.5
Sex (male vs. female)	-0.09	0.15	-0.38 – 0.20	0.557
Proficiency (source languages)	0.08	0.17	-0.26 – 0.41	0.646
Attitude towards Lánnang-uè: ‘broken’	-0.04	0.04	-0.12 – 0.04	0.368
Attitude towards Lánnang-uè: ‘reflective of hybrid identity’	-0.07	0.06	-0.19 – 0.05	0.26

In Section 5.3, I hypothesized that most speakers will consistently use, or at least tend to use, Hokkien-derived prepositions for certain classes or types (i.e., orientation, range/path/ general location), in accordance with the pattern described in Section 5.2. There will be high rates of spread and stability. I also hypothesized that a significant part of the pattern-non-conforming Hokkien-derived prepositions will come from speakers who report being highly proficient in Hokkien.

After modeling the likelihood of deriving prepositions from Hokkien (Table 59 in Appendix G), I found that ‘class’ reliably predicted the use of Hokkien-derived variants, as evidenced by the effect of ‘class’ in the model. If the preposition is part of the three classes of prepositions that are supposed to be derived from Hokkien mentioned in the previous paragraph and in Section 5.2, the Hokkien-derived variant is more likely to be used; if the conjunction is not part of the classes, the Hokkien-derived variant is less likely to be used (Figure 20a). In other words, most speakers consistently derived the enumerated preposition classes from Hokkien.

In addition to spread and stability, the results showed that the following groups were more likely than the rest to use Hokkien-derived variants for prepositions that are not expected to be derived from Hokkien: female speakers, and speakers with high Hokkien proficiency (Figure 20c, d, red bars). My examination of the marginal means also showed that older speakers and those with low Hokkien proficiency (Figure 20b and d, blue bars) were less likely than others to use Hokkien-derived variants for prepositions that are expected to be derived from Hokkien.

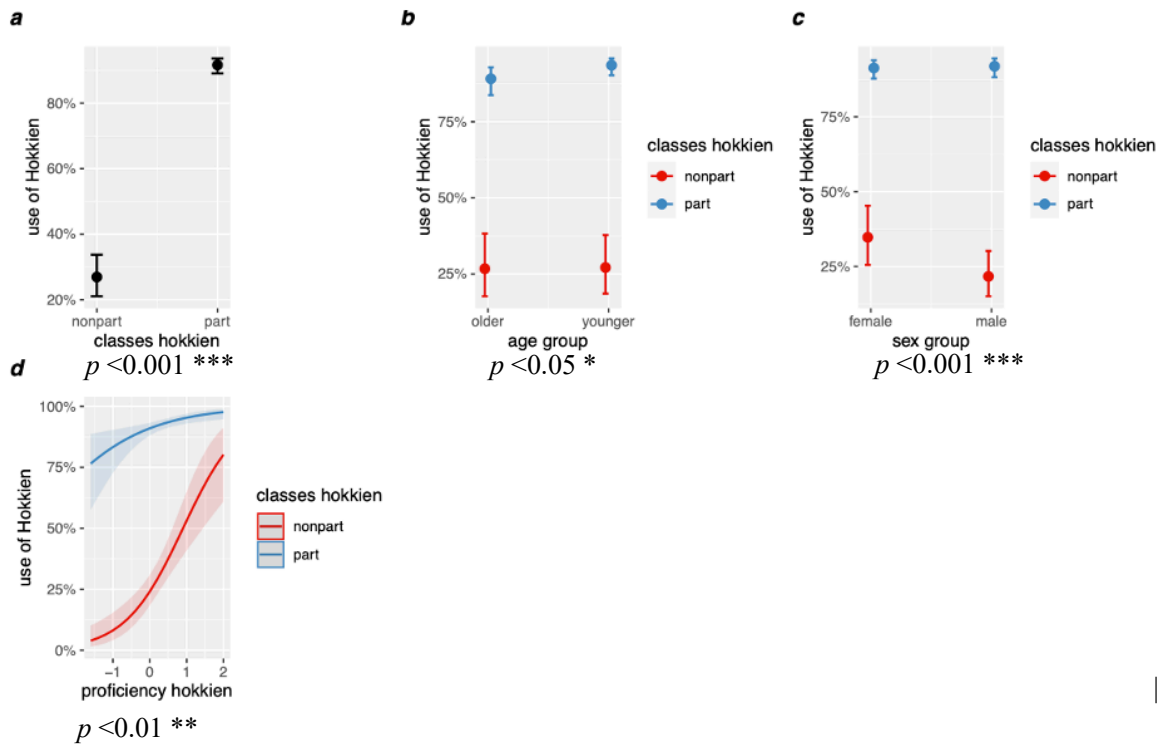
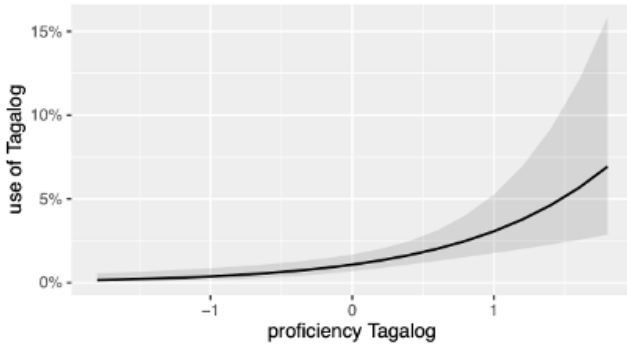


Figure 20. Marginal means/effects of sociolinguistic factors on likelihood to (not) adhere to the distributional pattern (Hokkien-derived prepositions)

In Section 5.3, I hypothesized that a significant portion of the Tagalog-derived prepositions, which are not part of the preposition distributional pattern (Section 5.2), will come from speakers who report being highly proficient in Tagalog. After modeling the likelihood of deriving prepositions from Tagalog (Table 60 in Appendix G), I found that high proficiency in Tagalog accounted for the tokens of Tagalog-derived prepositions, which were all pattern-non-conforming. This is supported by the presence of interaction effects between ‘class’ and proficiency in Tagalog. A closer examination of the data shows that speakers who have high proficiency in Tagalog use more Tagalog-derived prepositions than the rest (Figure 21).



$p < 0.001$ ***

Figure 21. Marginal means/effects of sociolinguistic factors on likelihood to use Tagalog-derived prepositions

In Section 5.3, I hypothesized that speakers will tend to use English-derived prepositions for certain classes or types (e.g., prepositions of accompaniment, see Section 5.2). I also hypothesized that a significant portion of the English-derived prepositions that did not conform to the pattern will come from speakers who reported being highly proficient in English. After modeling the likelihood of deriving prepositions from English (Table 61 in Appendix G), I found that ‘class’ or preposition type reliably predicted the use of English-derived variants. In other words, most speakers consistently derived the classes expected to be derived from English (Section 5.2) from English.

The model fitted on English-derived preposition data also showed that English proficiency conditioned the variation in the preposition pattern, supported by the presence of interaction effects between ‘class’ and English proficiency. Sex also did. Examining the marginal effects closely, I found that speakers with low proficiency in English and female speakers were less likely than the rest to use English-derived variants for prepositions that were expected to be derived from English (Figure 22, blue). However, the use of English-derived variants for prepositions that were not expected to be derived from English was not conditioned by sex or English proficiency (Figure 22, red). The model only predicts that speakers with low proficiency in English will be less likely than others to produce English-derived prepositions that conform to the pattern. It does not predict that speakers with high proficiency in English will be more likely than others to produce English-derived prepositions that do not conform to the preposition pattern.

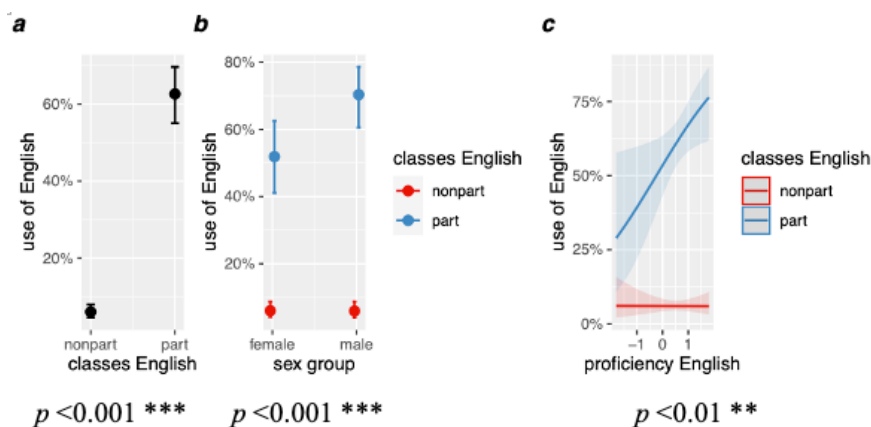


Figure 22. Marginal means/effects of sociolinguistic factors on likelihood to (not) adhere to the distributional pattern (English-derived prepositions)

The results of my model fitted on Mandarin-derived preposition data indicated no evidence of main effects of age, sex, and Mandarin language proficiency. None of the sociolinguistic factors were reliable predictors of the likelihood of using Mandarin-derived prepositions (Table 62 in Appendix G).

5.6 Discussion

5.6.1 Hypotheses 1 and 2: Spread and stability

The results in my descriptive and regression analyses support my hypothesis that the conjunction and preposition distributional patterns are highly widespread and stable. I have high rates of spread for both patterns: most speakers derived non-emphatic cumulative, disjunctive, general conditional, and manner conjunctions from Hokkien, conjunctions meaning ‘unless’ and ‘after’ from English, and the rest of the conjunction types from Tagalog at least once; they also derived prepositions of accompaniment, specific location, and time as well as prepositions meaning ‘of’ from English and derived prepositions of general location, orientation, and range/path from Hokkien at least once. Most did not derive conjunctions and prepositions from Mandarin and did not derive prepositions from Tagalog. In addition, I also found high individual and group consistency in the adherence to the patterns: speakers, on average, followed the distributional patterns with high levels of consistency and had patterns of variation similar to each other.

The widespread and stable distributional patterns observed are not surprising if Lánnang-uè has a high degree of languageness, as I have argued in previous studies such as Chapter 3 and Chapter 4 and other published work (Gonzales 2018; Gonzales and Starr 2020).

If we assume that Lánnang-uè is highly language-like, it is very likely that Lánnang-uè's conjunction and preposition patterns are not fully immune to language-external influences. For example, the patterns may be influenced by knowledge of vocabularies of Lánnang-uè's source languages, as Lánnang-uè is used in a high-contact environment where its source languages Hokkien, Tagalog, English, and Mandarin are used alongside it. Given the possibility of external influence, what factors may have reinforced the spread and stability of the two lexical patterns?

5.6.1.1 Possible reinforcers of the distributional patterns (general)

One possible reason why the two lexical patterns remain widespread and stable is that speakers wanted to minimize the perceived difficulty of aurally comprehending their speech. In my interviews, some speakers claimed that if linguistic elements from Tagalog, English, Hokkien, and Mandarin were randomly used by the speaker, it would confuse the listeners. These speakers believed that consistent sourcing of linguistic elements aids in listening comprehension and that random mixing of elements increases the risk of communication breakdown within the community (487 and 488). For instance, speaker PC0072, responding to Lánnang-uè speakers' mock random admixture of Mandarin-, Hokkien-, Tagalog-, and English-derived elements, claimed that they could not understand what the mixed utterances meant (489). Pending a more systematic investigation, speakers' attempt to minimize the perceived difficulty in the comprehension of randomly mixed utterances within the Lánnang-uè-speaking community may have contributed to the continued use of the highly widespread and stable conjunction and preposition distributional patterns.

(487) *Khânân pag dî na tshîntshai kay-tshâm lâng kô ó, khânân ke khâ kangkhô intindî.*
'It's like people find it harder to understand you when you randomly mix.'
<CLIN-19-9:3882>

(488) *Dî na bo halô, ín a buehiaú thiānn la.*
'If you don't mix (properly), they won't understand you.'
<CLIN-19-68:41493>

- (489) *So, guâ bue bingpiák î lê kông shà.*
 ‘So, I can’t understand what they were saying.’
 <CLIN-18-72:11138>

Social factors may have also been reinforcing factors. One possible social reason for the continued widespread and consistent use of the patterns is negative attitudes towards random mixing. Speakers generally viewed unsystematic mixing as undesirable and uneducated (490). When made aware of random ‘mixes’, they would correct themselves, or upon noticing another speaker using an unconventional term, they would occasionally correct the speaker either directly or indirectly by emphasizing the correct form through their own speech. This is supported by my anecdotal observations during my interviews, where I noticed some speakers correcting themselves by using, for instance, Hokkien-derived conjunctions for conjunction classes expected to be expressed using the Hokkien-derived variant just after we discussed the topic of their attitudes towards ‘random mixing’. For example, they used the Hokkien-derived locative preposition *tī* ‘at’ instead of Tagalog-derived *sā* ‘at’ after talking about their concerns about random mixing. Pending further investigation, the evidence suggests that negative attitudes towards random admixtures could have been one of the factors that prompted speakers to continue using the widespread and stable lexical distributional patterns.

- (490) *Khânân pag dī na tshîntshai kay-tshām lâng kô ó ...khâ phaíthiānn ânithê.*
 ‘It’s like you sound bad when you randomly mix.’
 <CLIN-19-9:3882>

Another social factor is stylistic practice. Speakers manipulate the resources that they have (e.g., linguistic resources) and combine them in certain way to construct and present certain personae. This is well-documented in the sociolinguistic literature (Eckert 2012) for established languages. Researchers, for example, have found that yuppies in Beijing – fashionable young middle-class women with well-paid jobs – consistently use full tone and avoid rhotacization, yielding a staccato sound matching “the crisp image required of women in the gendered cosmopolitan marketplace”; they use these features to distinguish themselves from other speakers, who use “smooth tone” (Eckert 2012:95; Zhang 2005). In the context of Lánnang-uè, speakers also engage in a similar process of stylistic “bricolage” (Eckert 2012:94; Hebdige 1979), where they use their systematic admixture (including the conjunction and preposition distributional pattern described) as an ethnolinguistic resource (Benor 2010) for constructing the hybrid, in-group

Lannang identity. For these speakers, the use of mixed features is emblematic of being a Lannang – a mixed identity that has Chinese, Filipino, and uniquely Lannang aspects (Gonzales 2021a:5). Specifically, it was characterized as ‘in-group’, ‘familiar’, ‘unique’, and ‘natural’ (Chuaunsu 1989; Gonzales 2021a). Based on ethnographic observations, I argue that the speakers maintained the distributional patterns because these patterns indexed these attributes, and speakers wanted to present themselves as coming from a unique Lannang culture. Evidence of this can be found in select interview excerpts below:

- (491) *Guâ gotsapkuí hè lo pero guâ si aî halo-halô rîn.*
 ‘I am fifty plus years old but **I love mixing the vocabulary** too.’
 <CLIN-18-2:711>
- (492) *Tsap tshai lomì a. So hîge si guâ e generatiòn, guâ e pég lo kô. Guâ na kâp in kong uè guâ e thia hosè la. Guâ bue left ôt. Kasî kông bue tuì. Hîge wavelêngth khâ sáng â.*
 ‘Everything is mixed. **This came from my generation [young generation]**. If I talk with them, I can understand them. I won’t feel left out because [if I speak Hokkien], there would be miscommunication. The wavelength would not be the same.’
 <CLIN-19-41:7286>
- (493) *Hîge mixed e languâge ... î tsigé sî tsîge community lo rîn ê. Na dân lê kong-uè piêntsue yá phóthong lo kaya hîgé na dân uhuâtthang comunicâte... khâ kín comunicâte kaysa hîge yá pùre e bá.*
 ‘The mixed language is part of our community. **It is a common thing that we use to communicate better** compared to the pure varieties.’
 <CLIN-18-71:10714>
- (494) *Kiaū fellow Fil-Chì, parang it’s like you have a unique secret language.*
 With fellow Filipino-Chinese (Lannangs), it is like you have a **unique secret language**.’
 <CLIN-19-10:4076>
- (495) *Only the Chinese Filipino can understand the trilingual Chinese Filipino. Guâ na kóng tampo Tagalôg, Iengbún, kap Lánláng-uè dî e bingpiak guâ. Filipino bue understand guâ, American bue understand guâ, Taidiokláng ma thia bó dân diba?*
 ‘**Only the Chinese Filipinos [Lannang] can understand the trilingual Chinese Filipinos [Lannang]**. If I speak a little bit of Tagalog, English, and Lánlang-uè, you [a Lannang] can understand me. Filipinos, Mainlanders, and Americans won’t understand me, right?’
 <CLIN-18-5:1896>

There is also a potential cognitive explanation. Another possible reason for the spread and stability of the distributional patterns is optimization of the conjunction and preposition lexicon to reduce potential cognitive load and processing costs. Research has shown that multilinguals

exert cognitive effort when accessing (lexical) resources from multiple languages: they are more likely to encounter “interference” arising from co-activated lexical concepts from their multilingual repertoire(s) (Verreyt et al. 2016:183; Plass et al. 2003; Oberauer 2009:346; Monaghan and Roberts 2019). A representation for a concept frequently gets “overwritten” (Oberauer 2009:355) by similar competing representations, leading to more tasks involving working memory. The overlap of representations for concepts also incurs “shift costs,” such as when speakers alternate between lexical repertoires from different languages, as in the case of Mandarin-English and Spanish-English bilinguals (Verreyt et al. 2016:184; Prior and MacWhinney 2009). Research has also shown that multilinguals deal with this by exercising “cognitive control” (Bosma and Blom 2019:1432), continuously inhibiting co-activated lexical items that are not accessed. The recruitment of control resources to inhibit conflicting activation of competing items leads to increased cognitive load and effort (Bosma and Blom 2019; Verreyt et al. 2016; Green 1998; Plass et al. 2003:221). This was observed in the case of Dutch-Frisian bilinguals, who were found to exercise cognitive control during code-switching from Dutch to Frisian and from Frisian to Dutch (Bosma and Blom 2019:1431). It was also observed in other research involving speaking (Hermans et al. 2003) and listening (Lagrou et al. 2011).

In the context of Lánang-uè, speakers might have (unconsciously) decided to continue using a single representation (e.g., Tagalog-derived variant) for a particular concept (e.g., adversative conjunctions) instead of having multiple alternatives (e.g., Hokkien-derived, Mandarin-derived conjunctions) to minimize the cognitive costs incurred by the overlap or interference among co-activated conjunction and preposition items derived from Hokkien, Mandarin, Tagalog, and English. This could be one of the factors that explain the maintenance of the distributional patterns.

While a cognitive account of the distributional patterns is appealing, I hesitate to commit to it in the absence of concrete evidence in the context of Lánang-uè. Future (experimental) research – testing whether a causal relationship can be found between cost reduction and the maintenance of the distributional patterns – is needed to definitively conclude that cognitive factors led to the reinforcement of the widespread and stable adherence to the distributional patterns.

In summary, I proposed four factors that reinforce the spread and stability of the distributional patterns: the minimization of perceived difficulties in listening comprehension,

negative attitudes towards random mixing, stylistic practice, and reduction of cognitive load. However, I have yet to comment on the nature of the distributional patterns. For instance, why weren't Mandarin-derived elements included in the distributional patterns? What factors reinforced the exclusion of Mandarin in the patterns? And why do many speakers continue to derive certain word classes from certain languages? I attempt to answer these questions in the next section.

5.6.1.2 Possible explanations for the (specific) nature of the distributional patterns

One possible reason for the continued exclusion of Mandarin elements from the distributional patterns is the Lannang community's lack of continuous exposure to Mandarin. As discussed in Chapter 2.2.4, community members are generally exposed to Mandarin only in the context of formal education – there are some members who frequently encounter Mandarin in other communicative contexts (e.g., dealing with Mandarin-proficient Mainland Chinese immigrants, frequent travel to Mandarin-speaking areas in China, etc.), but they are in the minority. The community's limited exposure to Mandarin creates a linguistic environment that is not conducive to the importation of closed-class lexical items or grammatical morphemes to Lánnang-uè and would explain the continued lack of (closed-class) Mandarin-derived conjunctions and prepositions in the variety.

Another possible factor that reinforced the general exclusion of Mandarin-sourced elements in the patterns today despite the trend in Lannang schools towards introducing Mandarin subjects and promoting Mandarin use (Tan 1993; Poa 2004; Gonzales 2017c) has to do with attitudes towards the language. Many speakers viewed Mandarin positively, claiming that it is a global language that connects them to the Chinese-speaking world and gives them an advantage over non-Mandarin speakers.

(496) *So, Kogî yá tiong-iaù dîn là, kasî Kogî, feeling ko sī khânân English lò la. All over the world Kogi i ehiaû.*

‘So Mandarin is very important too. Because Mandarin, I feel, has become English. All over the world, Mandarin is known.’

<CLIN-19-126:26958-26959>

However, they claim that Mandarin is not a language that directly indexes Lannang-ness or ‘localness’. They have no strong motivation to use Mandarin (or incorporate Mandarin-derived

elements) when communicating with other Lannang speakers. Speakers' views towards Mandarin arguably reinforced the general exclusion of Mandarin-sourced elements in the lexical patterns.

(497)

A: *So dī ū kamkâk Kogî ko tloh khâlâng tsiūsī kay-replâce ... dân ē community-e languâge bò?*

'So do you mean that Mandarin should replace our community language?'

B: *Bo su-iaù. Kasî dân dī si tsia*

'Not necessary. Because we – you – are locals [Lannang].'

<CLIN-19-68:28325, CLIN-19-129:28326, 28328 >

(498)

A: *Kogî ū tiông-iaù bo tī dân-e community asi bo à?*

'Is Mandarin important in our community or is it not?'

B: *Community-wise, not much.*

'Community-wise, not much.'

<CLIN-19-118:20790, CLIN-19-68:20789>

(499) *no, parang I'm okay with teaching Mandarin, pero it's not the language that we will be communicating with...*

'No, I'm okay with schools teaching Mandarin, but it is not the language that [our Lannang community] communicates in...'

<CLIN-19-10:4091>

What about the use of pattern-conforming Tagalog-, Hokkien-, and English-derived elements?

What factors could explain why Hokkien-, Tagalog- and English-derived elements continue to be used in the distributional patterns? There is no single explanation.

One factor is congruence and perceptual salience. Speakers of Lánnang-uè may have continued to subconsciously identify perceptually salient "congruent" features (e.g., form, meaning) in the conjunctions and prepositions of Hokkien, Tagalog, and English (languages that my speakers all know), increasing the likelihood of these features being retained in Lánnang-uè (Baptista 2020:162; Matras and Sakel 2007). For instance, in the case of conditional conjunctions, speakers may have subconsciously noted that Hokkien and Tagalog have *na* [na] as

conjunctions – in Hokkien, *na* means ‘if’, while in Tagalog, it can mean ‘that’ or ‘if’. The perceptual salience of the *na*-form across the source languages may have reinforced the place of Hokkien-derived *na* in the distribution, explaining why speakers still tend to use Hokkien-derived *nā* over English derived *if* and Tagalog-derived *kûng* for the class of conditional conjunctions.

So far, I have identified congruence and perceptual salience as possible cognitive-linguistic factors that may have reinforced the unique distributional pattern of Lánngang-uè prepositions and conjunctions. However, it is important to note that linguistic factors are often not good predictors of (lack of) linguistic innovation (Thomason 2008). Even if linguistic factors (do not) favor the use of one variant over the other, speakers have been found to defy expectations. For example, attempting to predict lexical innovations by only using linguistic clues derived from language data produced poor results (Miller et al. 2020). In the context of Lánngang-uè, I have found cases where one can find both potential examples and counterexamples for the effect of linguistic factors on lexical preference. Research has, for example, shown that speakers generally tend to prefer shorter words over longer words due to factors involving memory (Calude et al. 2020; Monaghan and Roberts 2019). In Lánngang-uè we do find potential examples of this (e.g., the continued preference of monosyllabic Hokkien-derived *kiaū* ‘with’ over trisyllabic Tagalog-derived *kasamà* ‘with’); however, there are also potential counterexamples (e.g., the continued preference of the disyllabic Hokkien-derived *āsī* ‘or’ over monosyllabic *òr* ‘or’ and *ò* ‘or’), suggesting that memory does not always play a role in the maintenance of lexical patterns. Research has also shown that speakers tend to prefer words that have fewer functions and meanings over those that are multifunctional or polysemous (Calude et al. 2020). Again, there are some possible examples of this in Lánngang-uè (e.g., the continued preference of monofunctional English-derived *ôf* ‘of’ over multifunctional Tagalog-derived *ng* ‘of/iterative marker’), but possible counterexamples also abound (e.g., the continued preference of Tagalog-derived multifunctional *parà* ‘so that/stop’ over English-derived *so thât* ‘so that’ and Hokkien-derived *uitioh* ‘so that’, both of which are monofunctional). This suggests that the number of linguistic functions also does not always condition the maintenance of lexical patterns.

In sum, I do not have a set of factors that can explain the speakers’ continued use of particular source language variants with certainty. I do, however, have a list of linguistic factors

that can contribute to the reinforcement of some of the lexical patterns, but not all of them: pattern/form congruence, word length, and number of word meanings/functions.

Social factors can also influence the use and maintenance of patterns, and can even trump linguistic factors as predictors of (contact-induced) innovations (Thomason 2008:52), so I am not discounting social explanations. But while I have evidence of general innovations (i.e., general mixing) resulting from deliberate decisions (see Section 5.6.1.1), I have not observed speakers providing a reason for why they derived a particular class or type from a particular language. In the absence of finer-grained sociolinguistic data, I am reluctant to commit to social accounts for the maintenance of the conjunction and preposition distributional patterns. But this is not to say that social factors play no role in maintenance. In fact, it is likely that the lexical patterns remained widespread and stable due to multiple causation (Thomason 2008:47) – a combination of at least some the factors I hinted at or proposed.

5.6.2 *Hypothesis 3: Structured variation (Systematicity)*

5.6.2.1 General patterns

To recapitulate, the results in my regression analyses at the macro level partially supported my hypothesis (Hypothesis 3) that the variation in the adherence to the conjunction and preposition distributional patterns is structured. My findings for variation in general (not source-language-specific variation, which will be discussed in 5.6.2.2) showed that a significant part of the variation in the adherence to the conjunction pattern can be traced back to speakers who had negative attitudes towards Lánng-uè (i.e., viewing the variety as *barôk* ‘broken’) and those who are young, and not female speakers or speakers who did not view Lánng-uè as emblematic of the Lannang identity. The results also failed to provide evidence that the variation in the adherence to the preposition distributional pattern was conditioned by any of the sociolinguistic factors hypothesized.

Two questions naturally emerge from these findings: (1) why were most of the pattern-non-conforming tokens associated with speakers who are young and those who viewed Lánng-uè as ‘broken’? And (2) why do we only have evidence of age and attitudes conditioning the variation in conjunction patterns, but not evidence of them conditioning preposition patterns? I attempt to answer these questions, keeping in mind that, Lánng-uè and its highly widespread and stable lexical patterns may be partially influenced by language-external factors (and be

viewed as variety that is not fully crystallized) – factors that could trigger change, and that Lánnang-uè could have variation systematically embedded into its system (i.e., variation is a crucial part of Lánnang-uè; command of Lánnang-uè requires skillful manipulation of variants).

One possible reason why most of the non-conforming tokens came from younger speakers is because of a potential change-in-progress led by the youth. This claim is not far-fetched, as research in sociolinguistics has often found younger speakers – “people with energy and enterprise” or initiative (Maclagan et al. 1999:19) – leading language change. In previous variationist work on the morphology and phonology of Lánnang-uè (Gonzales 2018; Gonzales and Starr 2020), for example, I have found evidence that younger speakers introduced innovations to already conventionalized phonological and morphological systems. Assuming that the patterns observed in previous work are like the patterns observed here (that the non-conforming conjunction tokens are innovative or products of a change-in-progress), youth-led innovations can explain why many of the non-conforming tokens tended to come from the younger speakers.

Another possible sociolinguistic explanation of the link between young age and non-conformity involves attitudes and stylistic practice. Many of my young speakers collectively viewed Lánnang-uè as Hokkien *konyò* (Filipino: *conyo/konyo*) (Figure 23) – a negatively stigmatized linguistic style that is associated with young individuals who are perceived to be status-conscious, fussy, empty-headed, privileged, and effeminate (Reyes 2017:213). This select subset of Lánnang-uè-speaking youth seem to be associating distinctive linguistic features linked to this style (e.g., the hyper-articulation and lengthening of certain segments, altered prosody, innovative syntax, occasional redundant use of linguistic elements that have the same function or meaning, and frequent sourcing of elements from particular languages) (Reyes 2017:214) to Lánnang-uè, whose linguistic feature pool coincidentally overlaps with varieties traditionally viewed as *konyò* (e.g., Conyo English, Davao Conyo). As such, they also view Lánnang-uè as a product of “unholy mix[ing]”, “mangled mish-mash[ing]”, and “bastardiz[ing] languages” by *konyò* people (Reyes 2017:225). These younger speakers may have attempted to use more Hokkien-derived conjunctions to avoid sounding like they are using ‘mixed’ *konyò* Hokkien, increasing rates of non-conformity.

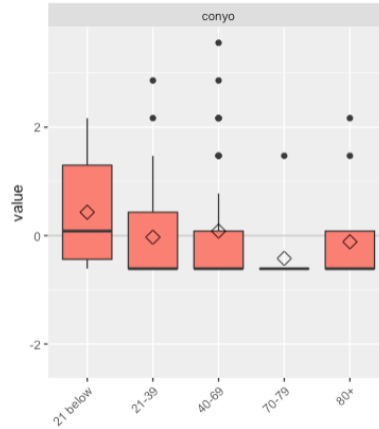


Figure 23. Speakers' z-scored ratings of Lánnang-uè being *konyò* (n= 117) Diamonds represent the mean; the heavy line indicates the median; the dots represent outliers.

(500) *Oh, kasî hîge.. khâ siaûdien láng ulê kông yá konyò dâw lân Lánnang-ue.*
 'Oh, because the... younger people have been saying that our Lánnang-uè is very conyo.'
 <CLIN-19-68:13387>

Another potential reason why many non-conforming conjunctions come from younger speakers is that they want to avoid sounding 'old'. One Lánnang-uè speaker, in their 20s, noted that the Lánnang-uè vocabulary I was using sounded 'old'. Another speaker, also in their 20s, said that the variety is *mejo khâ tiongpuè* 'somewhat more elderly'. Pending a more systematic investigation, it is possible that the (conscious?) avoidance of 'elderly' features could have also motivated the use of pattern-non-conforming conjunctions observed in many younger speakers.

How about the link between 'broken' attitudes towards Lánnang-uè and pattern-non-conforming conjunctions? A possible reason is that speakers who viewed Lánnang-uè as 'broken' were deliberately embodying these attitudes in their speech. It is well-established that attitudes towards language play a major role in the maintenance, removal, or innovation of particular linguistic features and patterns (Clarke and Erskine 2010; Thomason 2007). If speakers have positive attitudes towards a particular language or linguistic feature, they are more inclined to use it; however, if they attach a negative stigma to it, then the likelihood of them suppressing its use is amplified, even if (linguistic) factors favor its use (Thomason 2007). In the case of Lánnang-uè, I discovered that some speakers viewed Lánnang-uè as *barôk* 'broken' Hokkien in my interviews:

- (501) *Khâlang tsin phai thiānn lo âsi khâlang, khâlân bo suî kò hó...*
 ‘It’s like it [mixed Hokkien] sounds bad or like not pleasant, okay...’
 <CLIN-18-5:1837>
- (502) *most probably it's broken* [Hokkien].
 <CLIN-19-98:14808>
- (503) *gua khatsa e khuahuat kap huai kha tiongpue e lang..., dinna kong tit e ... ia gau ko.*
Tansi dinna halo-halo ay! Khalang tsinchiu, khalang tsinchiu puatsinn siak asi khalang
bo ho thia ane ko, khalang kangkho tsiapsiunn ba.
 ‘Earlier, my view was similar to that of my elders, if you speak Hokkien, you’re very smart. But if you mix, oops! It’s like ... It’s like being half-baked, and unpleasant-sounding...It’s like hard to accept for the elders.’
 <CLIN-18-5:1862>

The speakers’ perception of Lánnang-uè being ‘broken’ Hokkien influenced how they followed the conjunction pattern. Speakers who viewed the variety as such know the lexical pattern in Lánnang-ue (as evidenced in the unconscious use of the pattern-conforming conjunctions in speech) but manipulated it to reflect their views on the variety. Specifically, they perceived that their Lánnang-uè is ‘broken’ and as such, attempted to avoid using Lánnang-uè features (e.g., the conjunction pattern). This is reflected in their rates of pattern-non-conforming token use, which are higher than the rest of the speakers. There is evidence of deliberate manipulation in my interviews: some of these speakers said that they were embarrassed about their ‘broken’ Lánnang-uè and said that they hoped to use a more *tít* ‘straight’ Lánnang-uè by introducing more Hokkien-derived elements – and they did for a while, but eventually reverted subconsciously to what they claimed was ‘broken’ (i.e., the variety with a stable and widespread conjunction pattern). This negative perception of the variety and the action they took (i.e., “embodying” the ‘broken-ness’ of Lánnang-uè) (Esposito and Gratton 2020:10) could explain why these speakers had significantly higher rates of pattern-non-conforming conjunctions compared to the rest of the speakers.

If the (attempted) avoidance of sounding ‘broken’, ‘*konyò*’, and ‘elderly’ was responsible for most of the variation in conjunction use, then why wasn’t a significant part of the variation in preposition use associated with these speakers as well? I argue that the asymmetry between prepositions and conjunctions is because of differences in awareness or levels of “sociological consciousness” (Dodsworth 2005:99) of the conjunction and preposition word classes.

Research has shown that speakers who are aware of the semiotic resources that they can use to embody a certain set of attributes are more likely to (deliberately) manipulate these resources, and those who are unaware are less likely to (Dodsworth 2005). In other words, the differences in levels of consciousness can account for differences in variation patterns, which is exactly what I found for the variation in patterns in conjunction and preposition use. In my interviews, many speakers constantly brought up conjunctions when asked what elements they thought of as being distinctively Lánnang-uè, but did not mention anything about prepositions at all:

(504) *O, chiunn guâ lê kông là, 'kasi' dibá?*
 'Here, just like what I said earlier, *kasi* ['because'], right?'
 <CLIN-19-117:20213>

O, dī khuà, 'pero' nanamân.
 "Here, look, *pero* 'but' again."
 <CLIN-19-117:20240>

Dī buetsuê kong sêh "KASÍ guâ ti tsia".
 "You can't just say *KASI gua ti tsia* 'BECAUSE I am here' (to the Mainlanders)."
 <CLIN-19-117:20325>

(505) *Dân lê kâp Taidiokláng kong-uè e sītsùn kóng 'però' ... dī kuakîn ua lo 'tansī'.*
 'The time you speak with the Mainlanders and use *pero*, you quickly change it to *tansi*.'
 <CLIN-19-119:21435>

I interpret this as speakers having high levels of sociological consciousness of conjunctions and low levels of consciousness of prepositions. The varying levels of awareness can (partially) explain why I found different patterns of sociolinguistic variation for conjunctions compared to prepositions.

5.6.2.2 Source-language-specific patterns

In my finer-grained regression models, I found that sex and age conditioned the use of pattern-non-conforming conjunction and preposition tokens derived from Hokkien and/or English. By 'pattern-non-conforming tokens', I refer to conjunctions or prepositions that were derived from an unexpected source language (Section 5.2). For example, as Table 33 shows, a large portion of the non-conforming conjunction tokens that are Hokkien-derived were associated with younger

speakers and male speakers, whereas a significant part of the non-conforming preposition variants that were derived from the same language came from female speakers.

Table 33. Associations between tokens that did not conform to distributional pattern and age/sex groups, by word class and source language

Tokens that did not conform to distribution	Conjunctions	Prepositions
Hokkien-derived	<ul style="list-style-type: none"> • Younger • Male 	<ul style="list-style-type: none"> • Female
English-derived	<ul style="list-style-type: none"> • Female 	No evidence of associations
Tagalog-derived	No evidence of associations	No evidence of associations
Mandarin-derived	No evidence of associations	No evidence of associations

I identified two patterns in my finer-grained regression analyses (Table 33):

1. Only variation involving Hokkien-derived conjunction tokens (pattern-non-conforming tokens) was conditioned by age.
2. The direction of the sex effect is not uniform. Pattern-non-conforming tokens were associated with either male or female speakers.

The finding that Hokkien-derived non-conforming conjunctions can be traced back to younger speakers is unsurprising if we consider the discussion involving young speakers in Section 5.6.2.1 where I pointed out that a sizable number of young speakers consider Lánnang-uè as a *konyò*-style Hokkien. These young speakers, trying to avoid being viewed as *konyò* when they speak Lánnang-uè, introduced more Hokkien-derived conjunctions to conjunction classes that are not expected to be derived from Hokkien (e.g., adversative conjunctions, conjunctions of reason), increasing their use of non-conforming Hokkien-derived conjunctions.

What about the associations between non-conforming tokens and particular sex groups (pattern 3)? Sociolinguistic research shows that sex-conditioned language variation can be motivated by complex stylistic practice or the desire to project particular social identities (Eckert 1989:245; Eckert and McConnell-Ginet 2003). In other words, the use of certain variants can be associated with either females or males (or other sex groups) depending on the context. For

example, in one of the first variationist studies (a study of diphthongs in Martha's Vineyard in 1972), Labov (1972) showed a link between the use of particular variants and men. He found that a tightly-knit group of native Vineyarder fisherfolk – young men – tended to centralize the /aʊ/ and /aɪ/ diphthongs compared to the rest of the population, because diphthong centralization had social meaning for the young fishermen, i.e., they used centralization to stress their identity as a Vineyarder. I have yet to find evidence of male-specific and female-specific stylistic practices or social identities in the Lannang context, but there could be styles or personae that some female and male Lánnang-uè speakers are trying to avoid or present that would encourage speakers to occasionally influence the lexical distributional patterns for conjunctions and prepositions. Male and female speakers may increase the usage of Hokkien- and English-derived linguistic elements beyond their expected word classes because they (do not) want to be perceived a certain way. Overall, I do not yet have a detailed or definitive explanation for the pattern involving sex, but it is worth noting that this finding is consistent with what I have previously found in Lánnang-uè, where I also found links between variation and both male and female sex (Gonzales and Starr 2020) (see also Chapter 4).

5.6.3 *Hypothesis 4: Linguistic independence*

The results of my general regression models yielded no evidence of correlations between proficiency in Lánnang-uè's four source languages (high proficiency) and variation patterns (non-conforming conjunction and preposition tokens). However, a finer-grained analysis – analyzing the patterns by source language – revealed proficiency effects. The results revealed negative correlations between proficiency and conformance to patterns for some conjunctions and prepositions (Table 34). Many of the pattern-non-conforming Hokkien-derived preposition tokens, Tagalog-derived preposition tokens, and Tagalog-derived conjunction tokens came from speakers who reported being highly proficient in the source language. For example, the bulk of the Hokkien-derived prepositions that did not conform to the Lánnang-uè preposition pattern came from speakers highly proficient in Hokkien. I failed to find evidence of such a trend for the remainder of the conjunctions and prepositions (tokens labeled 'no evidence of associations' in Table 34).

Table 34. Associations between tokens that did not conform to distributional pattern and proficiency, by word class and source language

Tokens that did not conform to distribution	Conjunctions	Prepositions
Hokkien-derived	No evidence of associations	<ul style="list-style-type: none"> • Highly proficient in Hokkien
English-derived	No evidence of associations	No evidence of associations
Tagalog-derived	<ul style="list-style-type: none"> • Highly proficient in Tagalog 	<ul style="list-style-type: none"> • Highly proficient in Tagalog
Mandarin-derived	No evidence of associations	No evidence of associations

The results can be interpreted in two ways, and depending on the interpretation, the results may or may not be useful for evaluating the linguistic independence hypothesis. If one interprets the negative correlation as a causal relationship where proficiency in the source languages affected the patterns of variation, then the results overall provide some evidence against linguistic independence. Under this interpretation, the presence of a negative correlation indicates that high proficiency in the source languages (or rather, knowledge of other languages’ vocabulary) partially encourages speakers to deviate from some of the lexical patterns in Lánnang-uè (or “transfer” lexicon from the source languages to Lánnang-uè) (Thomason 2001; Hermans et al. 2003; Siegel 2012:187; Klaus et al. 2018; Pham et al. 2018). This could be viewed as evidence of linguistic dependence and evidence against languageness.

If, on the other hand, the proficiency variable in the negative correlation is interpreted as ‘expression of proficiency’ rather than actual proficiency (e.g., speakers not conforming to the conjunction pattern because they want to express high proficiency in a particular language),⁶⁹ then the results have little to say about linguistic independence, but they support the claim that Lánnang-uè is highly language-like, as the use of pattern-non-conforming features is interpreted as a stylistic choice that is embedded in Lánnang-uè’s linguistic system (speakers’ skillful manipulation of linguistic resources to express particular social meaning) (Eckert 2005; Hall-Lew et al. 2021) instead of a consequence of (subconscious) linguistic transfer or borrowing. It is possible that speakers (sub)consciously used non-conforming conjunction and preposition tokens

⁶⁹ Note that the proficiency factors I used in my analysis were self-reported. ‘Proficiency’ can thus be interpreted as ‘expression of proficiency in a particular source language.’

(tokens derived from Lánnang-uè's source languages) because they want to show others that they have proficiency in those languages. The use of non-conforming tokens is thus interpreted as being part of Lánnang-uè linguistic system, as an element that denotes a particular social meaning. For example, the increased use of Hokkien-derived prepositions that do not conform to the preposition pattern and the increased use of non-conforming conjunctions and prepositions derived from Tagalog signal high proficiency in Hokkien and Tagalog. Evidence of this can be found in my ethnographic observations conducted in Manila (summer of 2019) as well as in my interviews with the participants of this study. I discovered that community members who reported being proud of their command of a certain language were more likely to use vocabulary from that language. For example, a speaker who wished to show me that they are proficient in Hokkien used more Hokkien-derived Lánnang-uè vocabulary than other speakers who did not explicitly wish to do so.

Overall, the patterns of sociolinguistic variation, in this interpretation, are regarded as a component of Lánnang-uè's linguistic system. Some (pattern-non-conforming) conjunctions and prepositions are skillfully used to express particular social meanings within Lánnang-uè. The sociolinguistic patterns, under this view, do not have anything useful to say about the linguistic independence (Hypothesis 4) of Lánnang-uè's lexicon but they provide support for Lánnang-uè's high degree of languageness nevertheless, as evidence of systematicity or structured variation (Hypothesis 3).

5.6.4 *Synthesis*

Overall, I was able to identify some factors that might have contributed to the increased use of pattern-non-conforming tokens in my data. I have analyzed the variation in adherence to the conjunction and preposition lexical patterns holistically as well as by source language. I proposed that a combination of factors might have led to increased non-conformity. Some of these include negative language attitudes, sociological consciousness of certain linguistic features, stylistic practice, and linguistic transfer. But although I have accounted for a significant part of the variation in the adherence to the conjunctions and preposition patterns, there remains some variation that I could not account for. What could account for this? Perhaps coder and machine error, discussed in Section 5.7, might account for some of this variation.

Morphosyntactic and semantic factors could condition this, as they were found to do in the

lexicon of other contact varieties (Fisher et al. 1994; Defior and Alegria 2005; Vejdemo and Hörberg 2016). Alternatively, other factors that I have not mentioned might account for the variation. I leave the explanation of variation that has not yet been accounted for to future research.

5.7 Conclusion

In this chapter, I conducted a more thorough examination of the conjunction and preposition distributional patterns described in Chapter 3 and summarized in Section 5.2. I tested whether the conjunction and preposition patterns have high rates of spread by looking at rates of pattern adoption in the community. Second, I examined how stable the patterns were by measuring consistency at the individual and group levels. Third, I investigated the degree to which the variation observed is systematic and structured – specifically, to test the hypothesis that certain conjunction and preposition variants will be used to express social meaning, as part of Lánnang-uè’s linguistic system, similar to the variation found in established contact languages such as Singlish (Starr and Balasubramaniam 2019) and Baba Malay (Lee 2014). Finally, I examined Lánnang-uè’s degree of lexical independence from its source languages Hokkien, Tagalog, English, and Mandarin. With this investigation, I hoped to ascertain whether the variability observed in the conjunction and preposition patterns weakens my argument that Lánnang-uè has high degrees of languageness or not.

The results generally supported earlier observations (Gonzales 2018; Gonzales and Starr 2020) of Lánnang-uè being very language-like. I found high rates of pattern spread within the sample and high rates of pattern consistency within the individual, and low rates of variation pattern heterogeneity between my speakers (high degrees of stability). Most speakers of Lánnang-uè consistently used conjunctions and prepositions that adhered to the distributional patterns – characteristics of languageness.

My results also revealed that although for some of the variation in the adherence to the lexical patterns there is no evidence of conditioning by sociolinguistic factors, there is evidence of sociolinguistic conditioning for a sizable portion of the variation. A significant amount of the variation in conjunctions, for example, was conditioned by age (younger speakers), sex (female and male speakers), and attitudes (‘broken’ attitudes towards Lánnang-uè). The lexical variation is socially meaningful to (at least some of) these social groups. The systematic use of variation to

express particular social meaning (e.g., *konyò*) can be regarded as evidence for Lánnang-uè's languageness (Weinreich et al. 1968).

Finally, I found some evidence against linguistic independence in Lánnang-uè's conjunction and preposition lexicon, assuming that there is a causal link between source language proficiency and variation, where high proficiency (knowledge of source language vocabulary) partially encourages the use of pattern-non-conforming vocabulary in Lánnang-uè. Under this assumption, I found that some of the preposition and conjunction patterns (i.e., patterns involving Tagalog-derived conjunctions and prepositions, Hokkien-derived prepositions) were influenced by high proficiency in Tagalog and Hokkien (or knowledge of Tagalog and Hokkien conjunctions and prepositions). However, assuming that the proficiency variable in the causal relationship does not reflect actual proficiency but 'expression of proficiency', then results relevant to proficiency have nothing useful to say about linguistic independence (Hypothesis 4). They instead provide evidence in support for structured variation (Hypothesis 3) – speakers use the pattern-non-conforming conjunctions and prepositions to express social meaning or to show others that they have command of Lánnang-uè's source languages.

I see three limitations in my analyses. First, because I relied on human transcribers to annotate the data that the machine used to create a model to automatically tag the data, there is the risk of coder error. The average accuracy rate of the data coders based on a coding/tagging assessment given after training was 92%. It is thus very likely that my coders made some errors in annotating the model data for part-of-speech, even after peer review corrections. This error, included in the machine input, can cause the machine to predict part-of-speech classes with less accuracy and more variability, which may skew the results of this study. Assuming that the coders did not make any errors (which is very unlikely), the machine will also not be able to tag words for part-of-speech with 100% accuracy – the second limitation. This is because the machine relies on a “probabilistic” model (i.e., Conditional Random Fields) learned from a small text sample assumed to be representative (Lafferty et al. 2001:282). The machine encounters new cases that the model cannot handle due to limited observations in the training data among other concerns. As such, it can potentially inaccurately tag some words, thus also possibly skewing the results and increasing the amount of variation. Although I attempted to mitigate these issues by

asking three native speakers who have linguistics training to correct the data after the automatic tagging, my data correctors might have missed some tokens.

The third limitation involves my approach for representing the lexicon. Research has shown that semantics and morphosyntactic structure can condition lexical distributional patterns (Fisher et al. 1994; Defior and Alegria 2005; Vejdemo and Hörberg 2016). However, in this chapter, I did not consider these factors when I analyzed my data. I took a “parsimonious” approach (Daganzo et al. 2012:47) by focusing only on the correlations between sociolinguistic factors, source language, and classes. As such, some of the potential variability in the data may have not been accounted for, and I was not able to comment on possible linguistic factors (e.g., morphosyntactic structure) that could condition the lexical distributional patterns of conjunctions and prepositions in Lánnang-uè.

Despite the limitations, the results shed much-needed light on the conjunction and preposition lexical distributional patterns in Lánnang-uè, as well as the variation found in them. The anecdotally observed high rates of variation in the two distributional patterns – the impetus for this study – do not pose a significant challenge to the argument that Lánnang-uè is highly language-like, as I found high degrees of pattern adoption within the community (high spread), low rates of inter- and intra-speaker variation (high stability), and sociolinguistic constraints governing variation. There is also some indication that the variation patterns in Lánnang-uè conjunctions and prepositions are not totally influenced by or dependent on the prosodic patterns of its source languages. Altogether, my findings support my claim that Lánnang-uè has high degrees of languageness, using evidence from its lexicon.

Chapter 6 : The *Wh*-phrase Position Distributional Pattern

6.1 Introduction

In this chapter, I pursue the same line of inquiry explored in Chapter 4 and Chapter 5. However, this time, I focus on a syntactic pattern that I have identified as exhibiting greater rates of variation than other features in Lánnang-uè: the *wh*-phrase position distributional pattern, specifically the pattern involving the position (i.e., sentence-initial vs. sentence-medial/final) of adjunct (i.e., *why*-, *how*-, *when*-, *where*-phrases) and object argument phrases (i.e., *what*-, *who*-phrases) in Lánnang-uè matrix *wh*-questions (described in Chapter 3 and summarized in Section 6.3).

The overarching goal of this investigation is similar to Chapter 4 and Chapter 5: to test whether the seemingly high rates of variation in the *wh*-phrase position distributional pattern challenge my argument that Lánnang-uè exhibits a high degree of languageness. Like Chapter 4 and Chapter 5, I once more focus on spread, stability, structured variation, and linguistic independence – four hallmarks of languageness. I utilize the same methods: I examine rates of pattern adoption and consistency at the individual and group levels; I also analyze the variation using sociolinguistic and contact-linguistic lenses. To my knowledge, no work has attempted to test for Lánnang-uè's languageness by analyzing the position distributional pattern using these three variables yet, so this investigation fills that gap.

In this chapter, I address the following questions hoping to fill that gap:

1. How widespread is the *wh*-phrase position distributional pattern within Lánnang-uè speakers?
2. How stable is it? In other words, how consistently do individual speakers follow the pattern? And how similar are their patterns of variation to each other?
3. Is the variation structured like the variation in established contact languages? Will sociolinguistic factors (e.g., age, sex) condition a significant part of the variation?
4. Is the distributional pattern influenced by high proficiency in Lánnang-uè's source languages (or knowledge of their *wh*-phrase patterns)? Will proficiency in the source languages condition the variation in adherence to the *wh*-phrase distributional pattern?

In Section 6.2, I contextualize the study by describing the *wh*-question systems of the source languages of Lánnang-uè with respect to the position of the *wh*-phrase. In Section 6.3, I briefly describe the *wh*-phrase position distributional pattern in Lánnang-uè *wh*-questions. This is followed by Section 6.4, where I state my hypotheses about the pattern and motivate them. In Section 6.5, I discuss the methodology, detailing the experiments I designed as well as the analyses I did on the data. I present the results and discuss them in Sections 6.6 and 6.7, respectively. I conclude the chapter in Section 6.8.

6.2 *Wh*-phrase position in the source languages of Lánnang-uè

In Hokkien *wh*-questions, *wh*-phrases⁷⁰ do not undergo overt *wh*-movement (Sato 2013:311).

Subject argument *wh*-phrases do not move from sentence-medial/final position to the beginning of the sentence because they are already in the sentence-initial position by default (i.e., *wh-in-situ*) (Bodman 1987).

- (506) *Tsītsuī tsō lánġ =ē tshuì ni?*
 Who create person =GEN mouth PRT
 ‘Who made a person’s mouth?’
 (Pioneer Generation Philippine Hokkien, The Amoy Audio Bible Project, Biblical Seminary of the Philippines, Exodus 4:11a)

⁷⁰ In accordance with syntactic literature, I use the term ‘*wh*-phrase’ to refer to *wh*-words (simple *wh*-phrases) and phrases containing *wh*-words (complex *wh*-phrases) (Sato 2013).

Object argument *wh*-phrases (i.e., *who*-, argument *what*-phrases) are located in the verb phrase complement position, which is sentence-final except in certain cases (e.g., constructions with sentence-final discourse particles) as in (507). They do not move to the beginning of the sentence.

- (507) *Li beq ciao simmiq?*
 2SG will eat what
 ‘What are you going to eat?’
 (Amoy Hokkien, Bodman 1987:49)

Adjunct *wh*-phrases (i.e., *why*-, *how*-, *when*-, *where*-, adjunct *what*-phrases)⁷¹ are located in an adverbial position (508 and 509), which is sentence-medial except in certain cases (e.g., subject-less constructions). *Where*-phrases are additionally embedded in a prepositional phrase, as in (510).

- (508) *Din uisiammîh tsuê tsî hang taitsi?*
 2PL why do DEM CLF affair
 ‘Why have you done this?’
 (Pioneer Generation Philippine Hokkien,⁷² The Amoy Audio Bible Project, Biblical Seminary of the Philippines, Exodus 1:18b)

- (509) *Gùn siammîh sitsun khuâkhîdî iaū âsi tshuîtānn...?*
 1.PL what moment see 2.SG hungry or thirsty
 ‘When did we see you hungry or thirsty?’
 (Pioneer Generation Philippine Hokkien, The Amoy Audio Bible Project, Biblical Seminary of the Philippines, Matthew 25:44b)

- (510) *Tsîde lāng si tui tolóh ū tsîkhuân=ē tîhuī ... ?*
 This person COP from where have this=MOD wisdom
 ‘Where then did this man get all these things?’
 (Pioneer Generation Philippine Hokkien, The Amoy Audio Bible Project, Biblical Seminary of the Philippines, Matthew 13:56)

Overall, Hokkien is a *wh-in-situ* language – object argument *wh*-phrases and adjunct *wh*-phrases are not in the sentence-initial position by default. However, if a speaker wishes to topicalize these phrases, they may front them (Tang 1988; Wu 1999; Sato 2013:315), as in the example below. There are no restrictions on what type of *wh*-phrases can be topicalized.

⁷¹ *Wh*-phrases in Hokkien *wh*-questions may be expressed using a *what*-phrase (e.g., *what time* for *when*).

⁷² The audio data were collected from two female Lannangs who were around 80 years of age at the time of recording and who are proficient in Hokkien (Biblical Seminary of the Philippines 2011).

- (511) *Dixi* *yingang* *ke* *Pakia?*
 when 3.PL go Beijing
 ‘(Lit. When, they go to Beijing?)’
 (Sato 2013:316)

Like Hokkien, Mandarin is a *wh-in-situ* language (Cheung 2014). Subject argument *wh*-phrases do not move from sentence-medial/final position to the beginning of the sentence because they are already, by default, in the sentence-initial position. Object argument phrases are placed in the VP complement position by default. This position is sentence-final except in certain cases (e.g., constructions with sentence-final discourse particles). In (512), for instance, the object argument *wh*-phrase *shenme dongxi* ‘what thing’ is placed after the verb, at the end of the question.

- (512) *Ni* *mai* *-le* *shenme* *dongxi?*
 2.SG buy -PFV what thing
 ‘What thing did you buy?’
 (Cheung 2014:398)

Adjunct *wh*-phrases are placed in an adverbial position – typically sentence-medial except in certain cases (e.g., subject-less constructions). In the following example, the adjunct *wh*-phrase *shenme shihou* ‘what time/when’ is situated sentence-medially between the subject and the verb. It is not fronted by default.

- (513) *Tamen shenme* *shihou* *qu* *Beijing ?*
 3.PL what time go Beijing
 ‘What thing did you buy?’
 (Sato 2013:315)

Mandarin is reported to exhibit two special cases where the sentence-initial construction is licensed. The first is topicalization (Tang 1988; Wu 1999; Yuan and Dugarova 2012; Sato 2013:315; Cheung 2014), as in (514), where *shenme cai* ‘what dish’ is at the beginning of the clause to emphasize the topic. This is like Hokkien.

- (514) *Shenme* *cai* *ni* *mei* *you* *chi?*
 what dish 2.SG not have eat
 ‘What dish(es) did you not eat?’
 (Yuan and Dugarova 2012:534)

The positioning of *wh*-phrases sentence-initially in Mandarin also occurs when a speaker wants to mark questions with a wide-scope *why*. If the speaker wants to indicate that the *why* phrase applies

to the whole clause rather than the verb phrase, they can place the phrase at the beginning. In (515), for example, the person is asking for the reason why nobody resigned, not the reasons nobody had for resigning. A Mandarin speaker would tend to use (516) for the latter interpretation, where an analogous complex *wh*-phrase, *yinwei shenme*, is located after the subject *no one*.

(515) *Weishenme meiyou ren cizhi?*
 Why no person resign
 ‘Why didn’t anyone resign?’
 (Jin 2014:5)

(516) *Meiyou ren yinwei shenme cizhi?*
 No person because-of what resign
 ‘What reasons_i did nobody have for resigning _i?’
 (Jin 2014:5)

In contrast with the two Sinitic languages, Tagalog is a language that, by default, has the *wh*-phrase in the sentence-initial position (Schachter and Otnes 1972:51). Argument *wh*-phrases are fronted (517, 518).

(517) *Ano ang g<in>a-gawa =mo?*
 What ABS RED-PFV-do =2.SG.ERG
 ‘What are you doing?’
 (Aldridge 2002:414)

(518) *Sino ang na-matay sa ilog?*
 Who NOM PFV-die LOC river
 ‘Who died in the river?’
 (native speaker elicitation data 2020)

Adjunct *wh*-phrases in Tagalog *wh*-questions are also fronted:

(519) *Saan =ka b<um>ili ng libro?*
 Where 2.SG.ABS PFV-buy OBL book
 ‘Where did you buy your books?’
 (Aldridge 2002:416)

(520) *Bakit =mo ako p<in>atay?*
 why =2.SG.ERG 1.SG.ABS PFV-kill
 ‘Why did you kill me?’
 (native speaker elicitation data 2019)

The positioning of *wh*-phrases in English is identical to that of Tagalog. Object argument and adjunct *wh*-phrases are, by default, located at the sentence-initial position:

- (521) *What* did John kill? (object argument *wh*-phrase)
 (522) *Why* did John kill? (adjunct *wh*-phrase)
 (523) *Why* did John kill Mary? (adjunct *wh*-phrase)

Overall, then, Hokkien, Mandarin, Tagalog, and English differ in the way their *wh*-questions are constructed. In Hokkien and Mandarin, object argument and adjunct *wh*-phrases are, by default, in the VP complement (typically sentence-final) or adverbial (typically sentence-medial) position. The *wh*-phrase is only placed in the sentence-initial position in special cases or conditions (e.g., topicalization and/or wide scope interpretations). In Tagalog and English, the default position of the *wh*-phrase is the sentence-initial position.

6.3 *The wh*-phrase position distributional pattern in Lánnang-uè

Lánnang-uè has a *wh*-phrase position pattern that is distinct from its source languages, based on my exploratory analyses of 727 Lánnang-uè *wh*-questions elicited from ten speakers as well as supplementary data from the Lannang Corpus (Chapter 3) (Gonzales 2022a).

Subject argument *wh*-phrases in Lánnang-uè questions are sentence-initial, similar to English, Hokkien, Tagalog, and Mandarin.

- (524) *Shangá* *phâhsi* *i?*
 who kill 3.SG
 ‘Who killed them (singular)?’
 (male speaker, 26)

Object argument *wh*-phrases tend to be placed in the VP complement position, just like in Hokkien and Mandarin (525). They are typically placed in sentence-final position. An exception is when a sentence-final discourse particle is used, in which case the phrases are placed sentence-medially.

- (525) *În* *lê* *sûng* *tsuê* *shammîh?*
 3.PL PROG count as what
 ‘What did they count that as?’
 (male speaker, 24)

Adjunct *how*-, *when*-, or *where*-phrases (i.e., *tsiûwâ/chûngâ* ‘how’, *tîsí* ‘when’, or *tôlôh* ‘where’) tend to be placed in an adverbial position, also just like Hokkien and Mandarin. They are

generally placed in sentence-medial position. One exception to this placement pattern is in subject-less constructions, where the adjunct *wh*-phrase is placed in the sentence-initial position.

- (526) *Í tōlōh bēh explode a?*
 3.SG where will explode PRT
 ‘Where will they (SG) explode?’
 (male speaker, 23)

Unlike Hokkien and Mandarin, however, Lánnang-ue tends to have its adjunct *why*-phrases in the sentence-initial position or fronted by default ($\chi^2 = 47.53, p < 0.0001, n = 727$), patterning after default *wh*-fronting languages Tagalog and English instead of Hokkien and Mandarin (527).

- (527) *Kânâ dîn bēh tsaû a?*
 why 2.PL want run PRT
 ‘Why do you want to run?’
 (female speaker, 39)

The position of most *wh*-phrases in Lánnang-ue *wh*-questions is conditioned by phrase type. *Why*-phrases and subject argument *wh*-phrases are placed in the sentence-initial position while the rest of the *wh*-phrases are situated in medial or final position. I refer to this phenomenon as the *wh*-phrase position distributional pattern.

There is some variation in the speakers’ adherence to the pattern. *Why*-phrases are sometimes placed in the adverbial (sentence-medial) position, as in the following example:

- (528) *Í kâna bēh ho guâ candy a?*
 3.SG why want give 1.SG candy PRT
 ‘Why will they give me candy?’
 (female speaker, 39)

Object argument *wh*-phrases and adjunct *how*-, *when*-, or *where*-phrases are occasionally placed in the sentence-initial position:

- (529) *Tisí à tsíge snâke bēh tsiah rabbit?*
 When PRT DEM snake want eat rabbit
 ‘When will this snake eat the rabbit?’
 (female speaker, 39)

6.4 Hypotheses

I had four hypotheses regarding the position pattern described in Section 6.3 (sentence-initiality of *why*-phrases and sentence-mediality/finality for *how-/when-/where-/object who-/object what*-phrases), anchored on the premise that Lánnang-uè has high degrees of languageness:

1. Spread. The pattern will be highly widespread. Most speakers will follow the pattern at least once.
2. Stability. The pattern will be highly stable. Speakers who follow the pattern at all will do so at the individual level with high degrees of consistency. They will also have patterns of variation that will not differ too much from each other.
3. Structured variation (systematicity). A significant part of the variation in the pattern will be conditioned by at least one sociolinguistic factor, such as age and/or sex. If the variation involves innovation/change, a large part of non-conformance to the pattern will most likely come from younger speakers (specifically, younger females).
4. Independence from source languages. The patterns of variation will not be influenced by high proficiency in Lánnang-uè's source languages (or knowledge of *wh*-question patterns in these languages). Pattern-non-conforming behavior will not be traced back to speakers who report having high proficiency in default *wh*-fronting languages (i.e., Tagalog and English) and/or high proficiency in default *wh*-in-situ languages (i.e., Hokkien and Mandarin) (see subsection pertaining to Hypothesis 4 in this section for more specifics).

Hypotheses 1 and 2 had the same impetus as the hypotheses on spread and stability in Chapter 4 and Chapter 5. My previous work on Lánnang-uè features had consistently shown evidence of high degrees of spread and stability across different levels of language – the prosodic level (Chapter 4), the morphological level (Chapter 3) (Gonzales 2018), and the lexical level (Chapter 5). The presence of highly stable and widespread features and patterns across different linguistic levels of Lánnang-uè – indicative of high levels of languageness in Lánnang-uè – suggests that high rates of spread and stability will also be present in the position pattern of *wh*-phrases, as

established languages generally have these characteristics across their features/patterns. In other words, assuming that Lánnang-uè is highly language-like, I expect to observe a widespread and highly consistent position pattern with minimal variation.

Like Chapter 4 and Chapter 5, Hypothesis 3 was motivated by sociolinguistic research in Lánnang-uè where I found evidence of structured variation in many of its features (Gonzales 2018; Gonzales and Starr 2020), or specifically, evidence that variation is (systematically) constrained by social contexts – evidence that Lánnang-uè is highly language-like (Weinreich et al. 1968; Ghyselen and De Vogelaer 2018). If this is true, then the variation in adherence to the syntactic pattern should be structured like the variation found in other Lánnang-uè features/patterns, as languages tend to have structured variation across their features/patterns (Weinreich et al. 1968).

The hypothesized directions of the age and sex effects were motivated by sociolinguistic and language contact theories. I followed the same line of reasoning as I did in Chapter 4 and Chapter 5. In the context of language change, I expected pattern-non-conforming behavior to be associated with younger speakers (specifically, younger female speakers), as they are often regarded as “people with energy and enterprise” or initiative (Maclagan et al. 1999:19). The pattern of young women tending to innovate established conventions is also widely documented in the sociolinguistic literature (Eckert 1989; Maclagan et al. 1999; Starr and Balasubramaniam 2019). In other (non-innovation) contexts, I do not expect the effect to be in a particular direction. For example, in the context of stylistic practice, males instead of females may tend to use non-conforming features to express particular social meanings (Labov 1972; Obeidat and Hammoudi 2019) (see also Chapter 4.3 for a more in-depth discussion).

Hypothesis 4 was motivated by the observation that varieties that are ‘highly language-like’ tend to have linguistic patterns that are not influenced by the patterns of other languages (e.g., Topo and Ugsha varieties of Media Lengua) (Lipski 2020). High proficiency in a source language (i.e., Spanish) did not influence the structural patterns of these two language-like varieties (see Chapter 4.3 for a more in-depth discussion). If Lánnang-uè has high degrees of languageness like the two Media Lengua varieties, then I expect the patterns of variation in Lánnang-uè *wh*-questions not to be influenced by proficiency in its source languages (high proficiency), or knowledge of *wh*-question patterns in these languages. Specifically, I hypothesized that speakers with high proficiency in default *wh*-fronting languages (Tagalog and English) will not be more likely to:

1. use (pattern-non-conforming) sentence-initial *how-*, *when-*, *where-*, object *who-*, object *what-* phrases,
2. find constructions with these favorable, and
3. find constructions with sentence-medial/final *how-*, *when-*, *where-*, object *who-*, object *what-* phrases (pattern-conforming constructions) unfavorable.

Speakers proficient in default *wh*-in-situ languages (Hokkien and Mandarin) will not be more likely to:

1. use (pattern-non-conforming) sentence-medial *why*-phrases,
2. find constructions with these favorable, and
3. find constructions with sentence-initial *why*-phrases (pattern-conforming constructions) unfavorable.

6.5 Methodology

6.5.1 Approach

To test my hypotheses, I opted for an experimental approach. My initial plan was to analyze the structure of *wh*-questions from my corpus (Lannang Corpus) (Gonzales 2022a), but it did not have enough *wh*-questions per individual speaker to analyze. I also wanted to control the environment or context in which the *wh*-questions are produced, to minimize (or control for) possible semantic conditioning effects (e.g., topicalization, wide-scope interpretation). In addition, I wanted to get speakers from different age and sex groups with varying linguistic proficiency, to test my hypotheses on variation. I thus conducted production and acceptability experiments on 72 participants. Specifically, I conducted elicitation and scale-rating tasks that each lasted around 20 minutes.

After eliciting the questions and responses, I immediately conducted a 20-minute interview with each participant that focused on questions about Lannang community, identity, language, and education (Appendix D). At the end of the interview, I asked for age and sex information. I also asked the participants to rate their proficiency in Tagalog, English, Hokkien,

and Mandarin using a 7-point Likert scale. These were used for my analysis of the relationship of social factors on linguistic behavior.

6.5.2 *Production experiment*

6.5.2.1 Design and stimuli

This experiment was designed to elicit six sets of *wh*-questions containing object position inanimate *what* (i.e., instrument *what*), object position animate *what* (i.e., animal *what*), object position *who*, *when*, *where*, *why*, and *how* *wh*-phrases from Lánnang-uè speakers as naturally as possible. It was conducted in the guise of a crime investigation task, which I invented.

In the task, participants assisted a chief investigator (me) in collecting murder-related information (shown below) from six witnesses. Unlike a murder game where the goal is to identify the murderer, this task only asked the participants to come up with a description of the murder scene. Participants took the role of an assistant investigator, directly interrogating the witnesses (using *wh*-questions) to acquire information that would help them describe what happened in the crime scene. Here the character witnesses were printed on character cards, which were shown to the participants one by one, so the participants asked the character cards (in lieu of six actual persons). Participants were told that the character witnesses respond via response cards, arranged in eight decks (corresponding to the type of *wh*-phrase, such as *who*, *what*, *when*, etc.), shuffled and stacked according to the type of information (Figure 24). So, for instance, when a participant asked, “*Shāngá híge láng phâhsì?*” ‘Who did the person kill?’, the task facilitator (me) flipped a response card from the *who* deck and showed it to the participant.



Figure 24. Set-up of criminal investigation task

The participant then attempted to describe the picture shown. For instance, the participant could describe the picture as ‘*Tshiēng āng siâk nā dûwe*’ ‘the girl that is wearing red’. To facilitate their notetaking and to mask the task’s motive, participants were also asked to put their descriptions on a grid (Figure 25) so that they could keep track of the information they have already acquired.

	时间	WHO	如何使用		PLACE		为什么	
	七十月	牧师	X	青蛙	教堂	剪刀	为了男人	听人叙述
	XMAS DAY	PASTOR	ANGRY	CAT	HOSPITAL	KNIFE	X MONEY	HIDE IN THE CABINET
	圣诞节	COOK	X	青蛙	厨房	蝎	MONEY	BEHIND THE DOOR
	NEW YEAR	LADY	生气	想办法	SCHOOL	POISON	CRASH	WALA
	EPSA REVOLUTION	BAR OWNER	ENVIY	TURTLE	BEDROOM	HAMMER	CAR CRASH	偷看
	春节	房东太太	很快	蛇	BEDROOM	HAMMER	她弄坏了车	HIDING UNDER THE TABLE

Figure 25. Sample note-taking grid (filled)

The grid's first (left-most) column shows the six witnesses' portraits (the character cards), while the grid's upper-most row consists of icons that represent the crucial information that the participant must collect by asking specific *wh*-questions (the response cards): the time or date (*when*, column 2 in Figure 25), the victim (*who*, object position, column 3), the manner of using the weapon (*how*, column 4), the kind of pet (*what* - animate) that belonged to the victim, which was also killed by the murderer (column 5), the location (*where*, column 6), the weapon (*what* - inanimate, column 7), and the reason for killing (*why*, column 8). At the end, participants also needed to note how the witnesses acquired the information (*how*) (column 9). After the questioning and the filling out of the grid, participants were then asked to report their findings.

For some elderly participants, a simplified version⁷³ of the task had to be employed due to their inability to comprehend the task despite repeated explanations, or their unwillingness to

⁷³ I acknowledge that the simplified version of the task, done by 4 out of the 10 participants in the 80-89 age range, poses complications for the analyses. But given the difficulty of finding Lánnang-uè-speaking participants around

participate in a repetitive and ‘childish’ task. Instead of having them assume the role of an investigator and asking them to take down notes, I presented them with a scenario that requires a certain piece of information. Specifically, I gave them an unfinished sentence (containing the scenario) with the expectation that the participant would fill in the gap with a *wh*-question that helps them acquire that piece of information. For instance, for the information that corresponds to the inanimate *what*, participants were presented with the following: “The murderer killed a man. We want to know the identity of the person she killed. Your question for the murderer should be _____.” The participants filled the gap by asking the appropriate question. For instance, a participant might say “*Dî phâhsí shāngá*” ‘Who did you kill?’. This was then recorded by the facilitator on the grid.

Overall, in both versions of the task, a minimum of 48 clauses (8 *wh*-phrases × 6 stimuli) were elicited and audio-recorded per participant. The recordings were transcribed and coded by me (e.g., for position, for type).

The stimuli for this experiment are the black cards. I used iconic illustrations on a black card to induce the participants to produce specific *wh*-questions (the condition). For example, a picture of a house with a question mark meant the task was asking the participant to produce a *where* question. I used a picture of a knife so that participants could ask questions that would have a weapon as the direct object. Before my fieldwork, I asked three people to participate in the game to make sure that card stimuli were able to elicit the question structures I was interested in.

I summarize the stimuli by condition in Table 35. There was a practice block before the actual task to familiarize the participants with the task.

this age range, I decided to include data from the simplified version of the task in the analyses. The results, I hope, can still be compared to some degree, given that 6 out of the 10 old participants still did the original version, and their data is analyzed together with the data from the simplified version of the task. However, since the older subjects played different versions of the task, I am cautious in generalizing.

Table 35. Blocks, condition, and stimuli for the *wh*-question production experiment

Block	Condition <i>wh</i> -phrase type	# of stimuli
1 (practice)	<i>what</i> (inanimate)	3
2	<i>what</i> (inanimate)	6
	<i>what</i> (animate)	6
	<i>who</i>	6
	<i>why</i>	6
	<i>when</i>	6
	<i>where</i>	6
	<i>how</i>	12

6.5.2.2 Dataset and preparation

I first transcribed the recordings from the speakers. I extracted all matrix *wh*-questions produced by the participants, excluding questions that did not have a subject as well as questions that only had subject argument *wh*-phrases. Then, I coded each question for position. If the *wh*-phrase in the question was at the sentence-initial position, it was coded as ‘1’, if it was at the sentence-medial or sentence-final position, it was coded as ‘0’. I also coded each question for type of *wh*-phrase (i.e., *why*, *how*, *when*, *where*, *who*, *what*). I coded *what*-phrases that function as adjunct *wh*-phrases (e.g., *siammîh sîkân* ‘what time’ for ‘when’), as ‘what’. I linked the demographic information (i.e., age, sex, language proficiency) collected after the interviews to each utterance.

I then coded each utterance for adherence to the distributional pattern. A question was marked as ‘1’ if the question’s structure conforms to the pattern described in Section 6.3 and marked ‘0’ if not. For example, a question that was coded ‘why’ for type and ‘1’ for position was marked ‘1’ because sentence-initial *why*-phrases adhere to the observed distributional pattern. However, if that token was coded ‘how’ instead of ‘why’, then that token was marked ‘0’, as *how*-phrases in sentence-initial position do not adhere to the pattern observed.

In summary, each question in my dataset (forming rows in my spreadsheet) was coded for:

1. position of *wh*-phrase (categorical)
2. type of *wh*-phrase (categorical)
3. age (continuous)
4. sex (categorical)
5. *z*-scored self-reported language proficiency (Tagalog, English, Hokkien, and Mandarin)
6. *z*-scored self-reported language proficiency in default *wh-in-situ* languages (Hokkien and Mandarin)
7. *z*-scored self-reported language proficiency in default *wh*-fronting languages (Tagalog and English)
8. adherence to the pattern (categorical)

My dataset has a total of 3,163 questions and was used for descriptive and regression analyses, which I discuss in Section 6.5.4.

6.5.3 *Acceptability experiment*

6.5.3.1 Design and stimuli

In this experiment, I instructed participants to rate *wh*-questions with sentence-initial, sentence-medial, or sentence-final *wh*-phrases (i.e., the stimuli) using a 7-point Likert scale – 1 (not acceptable) to 7 (very acceptable) using a portable laptop (MacBook Pro 13, 2017) running PsychoPy 3.0.

For each stimulus, participants – expected to be unaware of the research questions – first heard an audio recording⁷⁴ of the stimulus twice. The stimuli recordings were created and produced by me in a silent environment; I made multiple recordings and only selected those that were clear and noise-free. After being presented the audio recordings via headphones (Audio-Technica ATH-M20x), my participants were shown a visual stimulus⁷⁵ related to the audio and a picture of a red stoplight (Figure 26a) on the laptop screen, which prohibited them from

⁷⁴ To ensure that literate participants had no advantage over those that cannot read the questions using Lánnang-uè orthography, participants were not given textual stimuli and were instead given audio ones.

⁷⁵ The visual stimuli were primarily intended as cues for *wh*-phrase animacy. For instance, the audio stimulus *Siammih híge yayá pháh?* ‘What did the maid hit?’ was presented to participants with a picture of a dog to indicate that the *what* in the sentence refers to an animate entity, not an inanimate one. I tried my best to choose pictures that represent the *wh*-phrase of interest.

responding to the stimulus too early. If the audio stimulus, for example, was referring to a dog (e.g., ‘Why did the maid hit the dog?’), a picture of a dog and the stoplight was shown after the audio stimulus was played twice, as in Figure 26a. After a one second delay, they were presented a Likert scale and a picture of a green stoplight (Figure 26b). The stoplight indicates that they are permitted to respond to the audio and visual stimuli. Participants responded by clicking the appropriate button on a special keyboard – the number keys (1-7) were superimposed with emoticons that correspond to the acceptability judgments of the participant (see Figure 26b). A judgment of ‘1’ means that the participant deems the construction as highly unacceptable for them while a judgment of ‘7’ indicates the opposite – that the construction is highly acceptable.

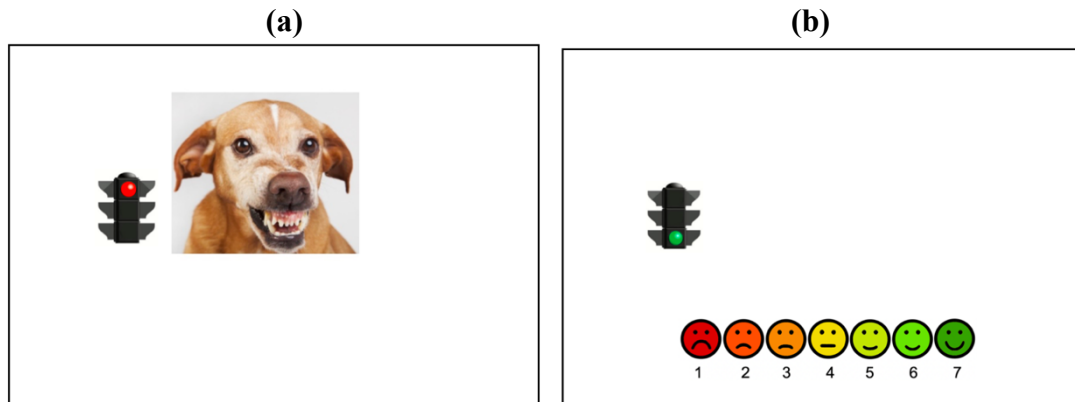


Figure 26. Screenshots of acceptability experiment

The experiment had two blocks. The first was a practice block with five trials, where participants judged stimuli that were not included in the main experiment. In this block, participants were allowed to ask the facilitator (me) questions about the experiment. The second block was the experiment proper. Participants rated the stimuli in a manner outlined in the previous paragraph without supervision.

The stimuli were stratified by *wh*-phrase type (i.e., who, animate what, inanimate what, when, where, how, why) and position (i.e., sentence-initial, sentence-medial/final), forming a total of 14 conditions (7 types X 2 position). The total number of items was 146 (Table 36). Examples of stimuli with argument *wh*-phrases (i.e., who, what) and adjunct *wh*-phrases (i.e., why, how, when, where) can be found in

Table 37 and Table 38, respectively.

Table 36. Conditions, and items for the *wh*-question acceptability experiment

Condition	Type	Position	Items
1	what (animate)	sentence-initial	11
2		sentence-final	11
3	what (inanimate)	sentence-initial	11
4		sentence-final	11
5	who	sentence-initial	11
6		sentence-final	11
7	why	sentence-initial	10
8		sentence-medial	10
9	how	sentence-initial	10
10		sentence-medial	10
11	when	sentence-initial	10
12		sentence-medial	10
13	where	sentence-initial	10
14		sentence-medial	10
Total			146

Table 37. Sample item set for argument *wh*-questions. Translation: ‘What/who will the visitor stab?’

Type	Animacy	Position	Example
what	inanimate	sentence-final	<i>Hígē lāngkhêh bēh tūsók siāmmih*?</i> ART visitor want stab what
		sentence-initial	<i>Siāmmih* hígē lāngkhêh bēh tūsók?</i> what ART visitor want stab
	animate	sentence-final	<i>Hígē lāngkhêh bēh tūsók siāmmih**?</i> ART visitor want stab what
		sentence-initial	<i>Siāmmih** hígē lāngkhêh bēh tūsók?</i> what ART visitor want stab
who	animate	sentence-final	<i>Hígē lāngkhêh bēh tūsók siāngá?</i> ART visitor want stab who
		sentence-initial	<i>Siāngá hígē lāngkhêh bēh tūsók?</i> who ART visitor want stab

* object, ** animal

Table 38. Sample item set for adjunct *wh*-questions. Translation: ‘{Why/how/when/where} will the maid carry the bed?’

Position	Example (frame)
sentence-medial	<i>Hîgē yáyá { } bêh káh hîgē tshúhng?</i> ART maid why want carry ART bed
sentence-initial	{ } <i>hîgē yáyá bêh káh hîgē tshúhng?</i> why ART maid want carry ART bed

{ } = *kânâ/uīsiāmmîh* ‘why’, *tsiûwâ/tsâi-iùnn* ‘how’, *tīsí* ‘when’, *tolóh* ‘where’

Not all participants saw all items for all conditions, in the interest of time. For conditions with 11 items, participants were only exposed to two to three items per condition; for conditions with 10 items, participants were only exposed to five items per condition. Items were presented to the participants in a random order. No fillers were used.

6.5.3.2 Dataset and preparation

The 6,048 ratings from PsychoPy were transferred to a spreadsheet. Each rating (forming rows in my spreadsheet) was coded for:

1. position of *wh*-phrase (categorical)
2. type of *wh*-phrase (categorical)
3. age (continuous)
4. sex (categorical)
5. *z*-scored self-reported language proficiency in the source languages of Lánnang-uè (Hokkien, Tagalog, English, Mandarin)
6. *z*-scored self-reported language proficiency in default *wh-in-situ* languages (Hokkien and Mandarin)
7. *z*-scored self-reported language proficiency in default *wh-fronting* languages (Tagalog and English)
8. adherence to the pattern (categorical)

‘Adherence to the pattern’ is estimated by coding the z -scored ratings binarily based on position and phrase type. A rating or trial was marked as ‘adhering’ or ‘1’ if any of the four criteria are passed:

1. the position is ‘sentence-initial’, the type is ‘why’, and if the rating is above the mean (i.e., 0)
2. the position is ‘sentence-medial/final’, the type is ‘how/when/where/who/what’, and if the rating is above the mean
3. the position is ‘sentence-initial’, the type is ‘how/when/where/who/what’, and if the rating is below the mean
4. the position is ‘sentence-medial/final’, the type is ‘why’, and if the rating is below the mean

If the trial did not fulfill any of the four criteria, it was marked as ‘non-adhering’ or ‘0’.

The coded dataset was used for mean distribution and variability analyses as well as several regression analyses.

6.5.4 *Analytical method*

6.5.4.1 Descriptive analyses

To test for spread and stability, I first measured the degree of spread and stability of the position pattern by examining the factor ‘adherence to pattern’ (Sections 6.5.2.2 and 6.5.3.2) using three measures discussed in Section 4.4.5. I also conducted analyses of (ratings of) tokens that did not conform to the pattern. I provide a breakdown of these tokens.

6.5.4.2 Regression analyses

I conducted several regression analyses on the dataset and subsets of the dataset to test for the potential effects of the hypothesized factors on *wh*-phrase position and adherence to the position pattern. Regression allows me to single out the (main or interaction) effects of the hypothesized factors and test for correlations between these factors and the dependent variables. I fitted three

generalized linear mixed-effects models on my production data and five linear mixed-effects models on my acceptability data.

To test my hypotheses on variation in general adherence to the pattern, I fitted a generalized linear mixed-effects model with logistic link function on the entire production dataset and a similar model on the entire acceptability dataset. The response or dependent variable for the production data was position, binarily coded. I included *wh*-phrase type (*why* vs. ***how/when/where/object who/object what***), sex (**male** vs. female), age (younger vs. **older**), and self-reported linguistic proficiency in the source languages⁷⁶ (i.e., Tagalog, English, Hokkien, Mandarin) as predictors in the model. The reference level, or the level to which the other level is compared, is indicated in boldface. Interactions between type and the sociolinguistic factors were modeled in to test whether the sociolinguistic factors condition the variation in adherence to the pattern. I included random intercepts⁷⁷ for participant. In the acceptability model, ‘adherence to the pattern’, detailed in 6.5.3.2) is my dependent variable. In addition to random intercepts for participant and item, I included age (continuous), sex (male vs. **female**), and proficiency in the source languages. No interaction effects were included.

To test my hypotheses about whether proficiency in particular languages affects likelihood of using a sentence-initial construction (production) or affects acceptability ratings of certain constructions (Hypothesis 4), I fitted mixed-effect models on subsets of the production and acceptability data. For production, two subsets (i.e., *why*-questions, *how-/when-/where-/object who-/object what*-questions) were made; for acceptability, four were made (i.e., constructions with sentence-initial *why*, constructions with sentence-initial *how/when/where/object who/object what*, constructions with sentence-medial *why*, constructions with sentence-medial or final *how/when/where/object who/object what*).

In the two production models, the dependent variable is the same as the first model – position (initial vs. non-initial). However, the predictors are slightly different. Both models have sex (**male** vs. female) and age as common predictors, with the reference level in bold. However, in the model involving *why*-questions, I added the predictor ‘language proficiency in default *wh*-

⁷⁶ I created the factor ‘proficiency in the source languages’ by running Principal Components Analysis or PCA on the z-scored Tagalog, English, Hokkien, and Mandarin proficiency scores and getting the component that is positively correlated with the four scores.

⁷⁷ Doing so gives me some statistical license to generalize my findings to the true population of Lánnang-uè speakers (Konstantopoulos and Hedges 2019:278).

in-situ languages’, as I hypothesized this variable would not condition the variation in the use of sentence-initial structure. In the other model involving *how-/when-/where-/object who-/object what*-questions, I added proficiency in default *wh*-fronting languages, as I hypothesized that this would not account for a significant portion of the *how-/when-/where-/ object who-/object what*-questions that have non-conforming sentence-initial *wh*-phrases. In both models, I included random intercepts for participant and did not include interaction variables.

In the four acceptability models, the dependent variable is *z*-scored ratings. All models have age (continuous) and sex (**male** vs. female) included as covariates (i.e., variables that can condition linguistic behavior, but which is not of direct interest). These models also all have random intercepts for participant and items and have no interaction terms. Two of these models (i.e., model of ratings of constructions with a sentence-initial *how/when/where/who/what* phrase, model of ratings of constructions with a sentence-medial or final *how/when/where/object who/object what* phrase) have proficiency in default *wh*-fronting languages as a predictor, as high proficiency in these languages was hypothesized not to condition the variation in the ratings for these constructions (i.e., increase ratings for sentence-initial *how/when/where/object who/object what* constructions and decrease ratings for sentence-medial or final *how/when/where/object who/object what* constructions). The remaining two acceptability models (i.e., ratings of constructions with a sentence-initial *why*-phrase, ratings of constructions with a sentence-medial *why* phrase) have proficiency in default *wh-in-situ* languages as a predictor, as high proficiency in these languages was hypothesized not to decrease ratings for sentence-initial *why* constructions or increase ratings for sentence-medial *why* constructions.

All statistical models were fitted in the R environment (R Core Team 2015). The `lme4` and `lmerTest` packages were used to estimate *p*-values (Bates et al. 2015; Kuznetsova et al. 2019). I used the `ggeffects` package (Lüdtke 2018b) to compute the estimated marginal means (predicted values) for the dependent variable at the margin of specific values or levels from certain model terms.

In all my regression models, the categorical predictor variables were analyzed after (re)coding the variables using unweighted effect contrast coding conventions (i.e., 1 vs. -1) (Sonderegger 2022).

6.5.4.3 Criteria for hypothesis testing

My hypotheses on spread and stability will be supported if I find evidence of them in my data. Regarding my descriptive analyses, if the spread scores for pattern adherence in production and acceptability data are above 0.5 (average) and the instability scores are above average (i.e., 0.5, or more than half of the population), then my hypotheses on spread will be supported. If the pattern has mean intraspeaker feature consistency scores (as measured in Section 4.4.5) higher than 0.5 (i.e., the pattern was followed more than 50% of the time, on average), then my hypothesis on stability will be supported. It will be further supported if I find interspeaker pattern inconsistency scores that are below 0.5 (i.e., the patterns of variation among speakers have heterogeneity levels below 50%) (Section 4.4.5).

On my general regression analysis of production data, if ‘type’ affects the dependent variable in my first production model, then my hypotheses on spread and stability will be supported as well, as I interpret the presence of a structural effect on the dependent variable as evidence of both spread and consistency. Regression model effects emerge if there are consistent correlations between the hypothesized factor (predictor) and the dependent variable for many speakers.

My general hypotheses on socially-conditioned variation in pattern adherence (i.e., age and sex conditioning much of the use of ‘non-adhering’ question constructions, high ratings for ‘non-adhering’ constructions, and low ratings for ‘adhering’ constructions) will be supported if I find effects of age and sex on the likelihood of adhering to the pattern in both production and acceptability datasets.

My hypotheses on linguistic autonomy (Hypothesis 4) will not be supported if I find evidence of negative correlations involving language proficiency in default *wh-in-situ* languages in the second production model and evidence of negative correlations involving language proficiency in default *wh*-fronting languages in the third. They will also not be supported if I find negative correlations involving language proficiency on ratings in the four subset acceptability models.

My hypotheses on the direction of the effect (e.g., female speakers varying more) will be supported if I find the expected pattern in an examination of the marginal effects or “predictions generated by a model when one holds the non-focal variables constant and varies the focal

variable(s)” (Lüdecke 2018a:1; Lüdecke 2018b) – the effects the individual predictors have on the dependent variable while all other variables are held constant.

6.5.5 *Measuring spread and stability*

I approximated the degree of spread by looking at rates of pattern adoption within the community. I measured the degree of stability by examining consistency rates at the individual and group level. Specifically, I relied on three measures, discussed in Chapter 4.4.5.

1. *Spread* – What proportion of my speakers follows the pattern at all?
2. *Mean intraspeaker consistency* – How often/consistently do individual speakers follow the pattern?
3. *Interspeaker pattern inconsistency* – How inconsistent are the patterns of variation between speakers?

To recapitulate, the following formulas were used in this chapter:

Spread score = *number of speakers who followed the pattern at least once / number of all speakers*

Individual intraspeaker consistency score = *number of tokens where an individual speaker who followed the pattern involving the wh-phrase / number of tokens where the pattern could be followed for that individual*

Mean intraspeaker consistency score = *individual intraspeaker consistency scores / number of individuals*

Interspeaker pattern inconsistency score = *standard deviation of all individual intraspeaker feature consistency scores / mean of these scores*

6.5.6 Participants

The individuals who participated in the experiments were recruited online or by word-of-mouth. I posted an announcement via social media and shared it with the University of the Philippines Department of Linguistics. To ensure that I had a balanced sample of participants according to age and linguistic proficiency in the source languages of Lánnang-uè, I did not add all interested participants to my participant pool. They were selected based on age and self-reported linguistic proficiency in the fronting languages (Tagalog and English) and proficiency in the *in-situ* languages (Hokkien, Mandarin). In addition, I only invited participants who were born and raised in the Philippines and identify as Lannang. I had a total of 72 participants (Table 39).

Table 39. Participant matrix (*wh*-questions)

Gender	Age Group							Total
	21-29	30-39	40-49	50-59	60-69	70-79	80-89	
Female	5	5	5	5	5	5	8	38
Male	6	5	5	5	6	5	2	34
Total	11	10	10	10	11	10	10	72

The average *z*-scored proficiency levels of these speakers are the following: Hokkien (mean = 0.2889, SD = 0.811), Mandarin (mean = -0.01, SD = 0.76), Tagalog (mean = 0.13, SD = 0.62), English (mean = 0.408, SD = 0.74). A negative score indicates low proficiency, a positive score indicates high proficiency, whereas a score close to zero indicates average proficiency. Eight of these participants participated in the prosody and lexicon studies of this dissertation.

6.6 Results

6.6.1 Production

6.6.1.1 Descriptive analyses

All my 72 speakers did one of the following at least once: they (1) placed the *wh*-phrase of a *why*-question sentence-initially or (2) placed the *wh*-phrase of *how-/when-/where-/* object *who-/*object *what*-questions either sentence-medially or sentence-finally. The spread score for adherence to the pattern is 1.

Not all the speakers who adhered to the distributional pattern at least once always were 100% consistent. Out of the 72 speakers who adhered to the pattern at all, one speaker followed

the pattern 50% to 74.99% of the time (i.e., 70.2%); 43 (59.72%) followed the pattern 75% to 89.99% to the time (mean = 85.24%, SD = 0.03); 24 (33.3%) followed the pattern almost all the time (90% to 99.99% of the time) (mean = 93.56%, SD = 0.02), and four never varied, following the pattern all the time. Figure 24 shows that the bulk of these speakers either completely or occasionally adhered to the pattern.

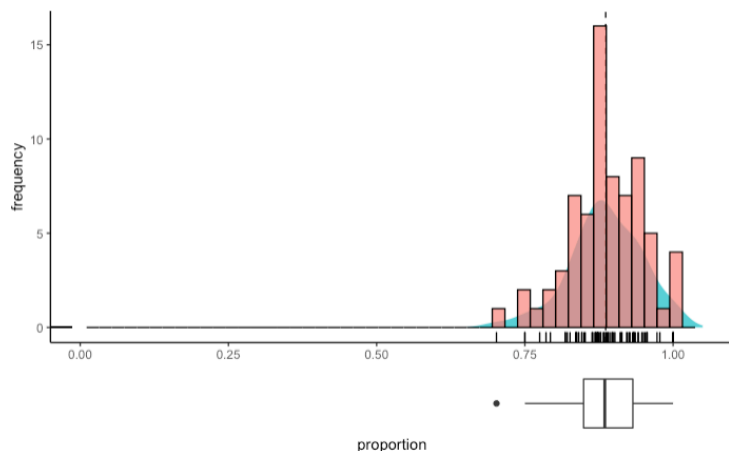


Figure 27. Histogram (frequency), density, and box plot of proportions (adherence to *wh*-phrase distributional pattern, production); broken line indicates mean

The mean intraspeaker consistency score for *all* participants is 0.8863 (SD = 0.0609). The interspeaker pattern inconsistency score is 0.0687. Overall, adherence to the *wh*-phrase position pattern is highly stable.

What exactly are the *wh*-phrases that did not adhere to the pattern? I provide a breakdown of these pattern-non-conforming tokens by *wh*-phrase type in Table 40. The first column of the table indicates the *wh*-phrase type. Column two lists the actual *wh*-phrase tokens whose position I coded. The third column lists the expected position for the type and the fourth lists the actual position observed. Column five has the number of occurrences in the data. The last column indicates the proportion of those occurrences to the total number of occurrences across all *wh*-phrase types observed. For instance, in the first row, I report that I found that 244 *why*-phrases in the data were in the sentence-medial or final position instead of the sentence-initial position, and that these phrases form 68.16% of the 358 non-conforming phrases found in the data.

Table 40. Distribution of *wh*-phrases (in the questions) that did not conform to the position distributional pattern, by type

<i>wh</i> -phrase type	Tokens	Expected position	Actual position	n	Percentage of non-conformance (n/ number of total n across all <i>wh</i> -phrase types X 100)
why	<i>kàna(tsi), uisiammîh</i>	sentence-initial	sentence-medial or final	244	68.16%
object what	<i>siammîh, shammîh, siammîh/ shammîh + NP</i>	sentence-final	sentence-initial	56	15.64%
object who	<i>siangá, shangá</i>			3	0.84%
when	<i>tisí</i>	sentence-medial	sentence-initial	19	5.31%
where	<i>tolóh, PP + tolóh</i>			21	5.87%
how	<i>tsaî-iùnn, tsiûwâ</i>			15	4.19%
Total				358	100%

The breakdown of distribution-non-adhering constructions in Table 40 indicates a possible asymmetry in variation patterns. There appears to be more variability in the construction of *why* questions (Table 40, shaded cell) compared to *how-/when-/where-/who-/what*-questions. On the surface, the raw frequencies alone (without individual participant information) (Table 40) suggest that most of my participants were highly consistent in adhering to the *how/when/where/who/what* portion of the position distribution but were not as consistent in adhering to the *why* portion. They suggest that *how/when/where/who/what* sentence-mediality or finality is diffused and consistent but that *why* sentence-initiality is not.

After examining the frequencies and variability of each participant by type (i.e., *why*-questions and *how-/when-/where-/who-/what*-questions), I found that most of my speakers who produced a *why*-question (47/68) placed the *wh*-phrase of a *why*-question sentence-initially at least once (spread score = 0.6912) and that all of my speakers placed the *wh*-phrase of *how-/when-/where-/who-/what*-questions either sentence-medially or sentence-finally at least once (spread score = 1). Both constructions are highly widespread within my speakers. Regarding

consistency, I found higher rates of variability in the use of *why*-questions (Figure 28a) but not for *how-/when-/where-/who-/what*-questions (Figure 28b). This is shown in Figure 28, where the values are more spread out in (a) compared to (b). Speakers who adhered to the pattern varied more in *why*-questions compared to *how-/when-/where-/who-/what*-questions.

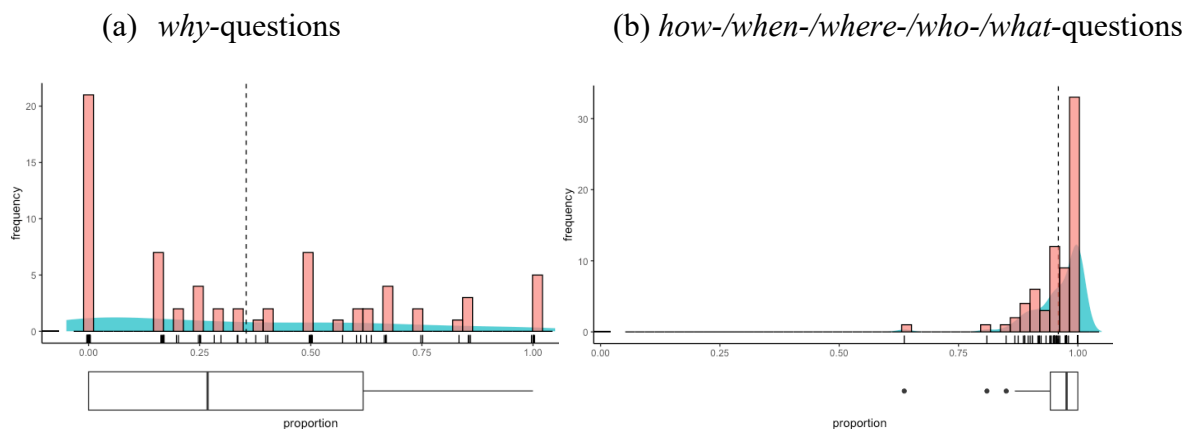


Figure 28. Histogram (frequency), density, and box plot of proportions (adherence to *wh*-phrase position distributional pattern, stratified by *wh*-type; left, *why* questions only, right, *how-/when-/where-/who-/what*-questions only); broken line indicates mean

Examining the variation quantitatively using the consistency measures in Section 4.4.5, I found that my participants were borderline consistent in producing sentence-initial *why* individually (mean intraspeaker consistency score = 0.5131, SD = 0.2726). Their patterns of variation for sentence-initial *why* are also somewhat heterogeneous: the interspeaker pattern inconsistency score is 0.5314, indicating borderline inconsistency.

Regarding the production of sentence-medial *how/when/where* and sentence-final *who/what*, I found high levels of consistency and stability (mean intraspeaker consistency score = 0.9591, SD = 0.2726, interspeaker pattern inconsistency score = 0.0617).

6.6.1.2 Regression analyses

My model of likelihood of putting the *wh*-phrase sentence-initially, summarized in Table 41, indicates a main effect of *wh*-phrase type (i.e., *why* vs. *how/when/where/who/what*). There are no main effects of sex or proficiency in the source languages. There is an interaction effect between type and age on this likelihood but none for type and the other two sociolinguistic factors. Phrase

type can predict the position of the *wh*-phrase. For certain age groups, it is a less reliable predictor.

Table 41. Linear regression results for likelihood to put the *wh*-phrase sentence-initially (observations = 3,163, conditional $R^2 = 0.427$, random intercepts for participant)

Predictors	Log-Odds	SE	CI	<i>p</i>
(Intercept)	-3.66	0.32	-4.29 – -3.04	<0.001
Type (why vs. how/when/where/ object who/ object what)	3.5	0.29	2.92 – 4.07	<0.001
Age (younger vs. older)	0.34	0.37	-0.39 – 1.07	0.364
Proficiency (Source languages)	0.09	0.23	-0.35 – 0.53	0.678
Sex (male vs. female)	-0.61	0.38	-1.35 – 0.13	0.108
Type : Age	-1.18	0.35	-1.86 – -0.50	0.001
Type : Proficiency	-0.09	0.2	-0.48 – 0.30	0.655
Type : Sex	0.33	0.35	-0.35 – 1.02	0.34

The marginal means of the interaction terms in this model indicate an asymmetry in variability between the production of sentence-initial *why*-phrases and the production of sentence-initial *how/when/where/who/what*-phrases. The error bars for the blue values are wider than the error bars for the red values, indicating that there is more variability in *why* constructions compared to *how/when/where/who/what* constructions (Figure 29).

Analyzing the means by social group, I found that a significant portion of constructions that do not conform to the distributional pattern – indicated by narrower differences between blue and red values in Figure 29 – is associated only with younger speakers. These speakers tended to produce significantly fewer sentence-initial *why*-phrases compared to older speakers (Figure 29a, blue). They also produced slightly more sentence-initial *how/when/where/who/what*-phrases (Figure 29a, red) compared to older speakers, who have patterns that resemble the expected pattern more (Figure 29a, left). A significant portion of the non-conforming constructions is not conditioned by sex (Figure 29b) or proficiency in the source languages (i.e., Tagalog, English, Mandarin, English) (Figure 29c). This model predicts that speakers are likely to not conform to the position pattern if they are younger.

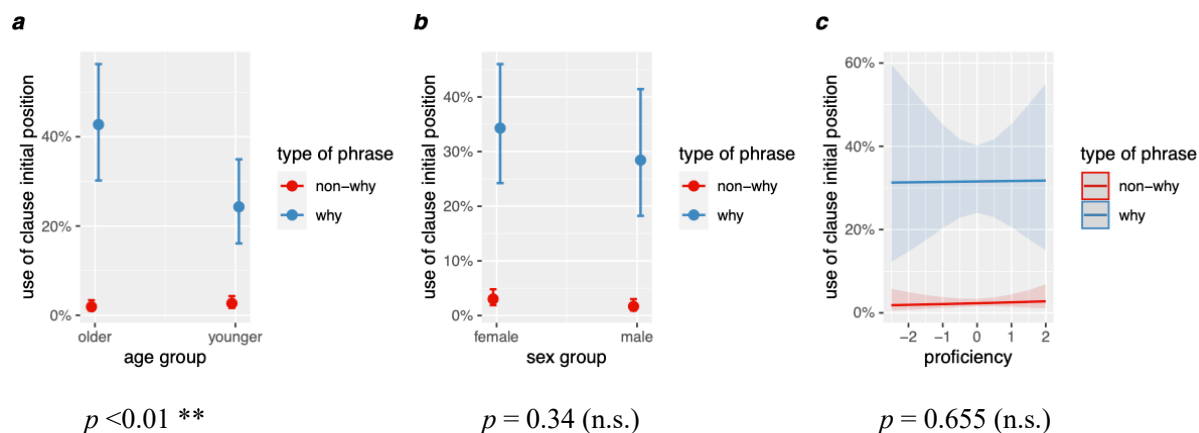


Figure 29. Marginal means/effects of *wh*-phrase type, age, sex, and proficiency on likelihood to put *wh*-phrase at sentence-initial position (production)

My model predicting likelihood of putting the *why*-phrase sentence-initially, shown in Table 42, indicates a main effect of age. There are no main effects of sex or proficiency in the source languages with default *wh-in-situ*. Only age can reliably predict this likelihood.

Table 42. Linear regression results for likelihood to put the *why*-phrase sentence-initially (observations = 390, conditional $R^2 = 0.469$, random intercepts for participant); in the Predictors column, reference levels are highlighted in bold; in the *p*-values column, statistically significant values are in bold.

Predictors	Log-Odds	SE	CI	<i>p</i>
(Intercept)	-2.42	0.9	-4.19 – -0.66	0.007
Age	0.03	0.02	0.00 – 0.06	0.044
Sex (male vs. female)	-0.21	0.5	-1.20 – 0.77	0.672
Proficiency (Source languages with default <i>wh-in-situ</i>)	0.02	0.22	-0.41 – 0.45	0.924

An examination of the marginal means of predictors (Figure 30) shows that a sizable portion of the *why*-phrases that are not in sentence-initial position comes from young speakers (Figure 30a). This model predicts that as age increases, the likelihood of putting the *why*-phrase in the sentence-initial position increases.

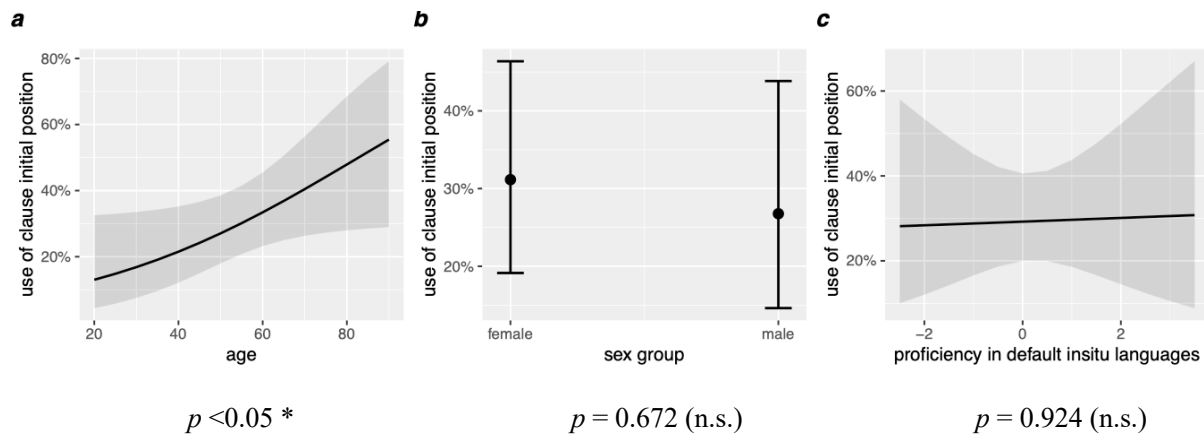


Figure 30. Marginal means/effects of age, sex, and proficiency on likelihood to put *why*-phrases at sentence-initial position

My model predicting likelihood of putting the *how/when/where/who/what*-phrase sentence-initially, shown in Table 43, shows no main effects of age, sex, or proficiency in the source languages with default *wh*-fronting. None of the factors are reliable predictors of position.

Table 43. Linear regression results for likelihood to put the *how/when/where/who/what*-phrase sentence-initially (observations = 2,770, conditional $R^2 = 0.302$, random intercepts for participant)

Predictors	Log-Odds	SE	CI	<i>p</i>
(Intercept)	3.96	0.66	2.67 – 5.25	<0.001
Age	0.00	0.01	-0.02 – 0.02	0.988
Sex (male vs. female)	-0.46	0.38	-1.21 – 0.28	0.223
Proficiency (Source languages with default <i>wh</i> -fronting)	-0.34	0.21	-0.76 – 0.08	0.109

6.6.2 Acceptability

6.6.2.1 Descriptive analyses

All 72 speakers adhered to the pattern at least once. Specifically, they did one of the following at least once:

- rated constructions with sentence-initial *why* high
- rated constructions with sentence-medial *how/when/where* or sentence-final *who/what* high
- rated constructions with sentence-initial *how/when/where/who/what* low
- rated constructions with sentence-medial or final *why* low

The spread score for adherence to the pattern is 1. None of my speakers were 100% consistent. Two speakers (0.03%) adhered to the pattern less than 50% of the time (mean = 43.45%, SD = 0.075). Thirty-five speakers (48.6%) adhered to the pattern 50% to 74.99% of the time (mean = 65.9%, SD = 0.06); 31 (43.05%) adhered to the pattern 75% and 89.99% to the time (mean = 82.29%, SD = 0.03); 2 (2.77%) adhered to the pattern almost all the time (90% to 99.99% of the time) (mean = 91.67%, SD = 0). Figure 31 shows that the bulk of these speakers adhered to the pattern but occasionally varied.

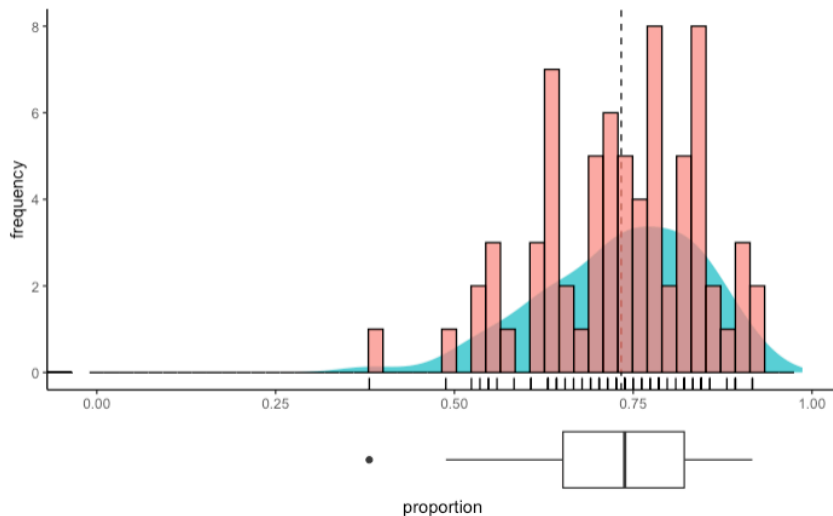


Figure 31. Histogram (frequency), density, and box plot of proportions (adherence to *wh*-phrase position distribution, acceptability); broken line indicates mean

The mean intraspeaker consistency score for *all* participants is 0.7333 (SD = 0.117). The interspeaker pattern inconsistency score is 0.1524. The scores suggest high levels of stability for the position pattern.

A breakdown of the pattern non-conforming ratings by *wh*-phrase type and position is provided in Table 44. The first column of the table indicates the *wh*-phrase type. Column two lists the actual *wh*-phrase used in the stimuli. The third column lists the position of the phrase in the stimuli. The fourth column lists the expected rating and the fifth lists the actual (non-conforming) rating. Column five lists the number of ratings. The last column indicates the proportion of those ratings to the total number of pattern non-conforming ratings across all *wh*-phrase types observed. For instance, in the first row, I report that 64 (3.64%) of the 1,756 non-conforming ratings in the acceptability dataset were (low) ratings of questions with sentence-initial *why*.

Table 44. Distribution of ratings that did not conform to the position distributional pattern, by type (cell with highest percentage highlighted)

Wh-phrase type	Wh-phrase in stimuli	Position	Expected rating	Actual rating	n	Percentage of non-conformance (n/ number of total n across all <i>wh</i> -phrase types X 100)
why	<i>kàna, uisiammih</i>	sentence-initial	high	low	64	3.64%
		sentence-medial/final	low	high	301	17.14%
object what	<i>siammih, shammih</i>	sentence-initial	low	high	213	12.13%
		sentence-final	high	low	177	10.07%
object who	<i>siangá, shangá</i>	sentence-initial	low	high	120	6.83%
		sentence-final	high	low	107	6.09%
when	<i>tisi</i>	sentence-initial	low	high	173	3.13%
		sentence-medial	high	low	55	9.85%
where	<i>tolóh</i>	sentence-initial	low	high	95	5.41%
		sentence-medial	high	low	242	13.78%
how	<i>tsai-iunn, tsiúwá</i>	sentence-initial	low	high	128	7.28%
		sentence-medial	high	low	81	4.61%
Total					1,756	100%

The breakdown in Table 44 indicates possible differences in variation patterns. There appears to be more variability in the ratings of certain constructions (see highlighted cell in Table 44). On the surface, the raw frequencies alone (without individual participant information) (Table 44) suggest that most of my participants were consistent in adhering to the *how/when/where/who/what* condition of the position distribution but may be less consistent in adhering to the *why* condition. They suggest that the pattern in the context of acceptability judgments of *why* questions may not be highly widespread and consistent.

An examination of the individual speakers' acceptability ratings (binarily coded as high and low, see 6.5.3.2) by type (i.e., *why*-questions and *how-/when-/where-/who-/what*-questions) and position (i.e., sentence-initial, sentence-medial or final) revealed low rates of spread for constructions with sentence-medial *why* but high rates of spread for the rest of the constructions (Table 45). Most speakers did not rate constructions with sentence-medial *why* low at all. Only 43.06% of my 72 speakers rated these constructions low in line with the distributional pattern.

Table 45. Spread scores by position and *wh*-phrase type (highlighted cells indicate low degrees of spread)

<i>wh</i> -phrase type	Position	Expected rating	n of speakers with expected rating	Spread score
<i>why</i>	sentence-initial	high	69	0.9583
	sentence-medial/final	low	31	0.4306
<i>how, when, where, object what, object who</i>	sentence-initial	low	72	1
	sentence-medial/final	high	72	1

For speakers who had rated constructions with sentence-medial or final *why* low at least once, I found high levels of variability in the judgments. This is reflected in Figure 32, where, out of the four subfigures, only the frequency distribution in the subfigure pertaining to sentence-medial *why* (Figure 32c) is skewed towards the left or zero, indicating lower adherence to the pattern (lower intraspeaker consistency). This subfigure also has the flattest distribution, as indicated by its rug plot (Figure 32c, directly above the x-axis), which shows the absence of data clustering. It indicates a high degree of interspeaker variation. I examined the variation quantitatively using

the consistency formulas in 4.4.5 and found that the ratings of sentence-medial *why* constructions were, indeed, highly inconsistent (Table 53).

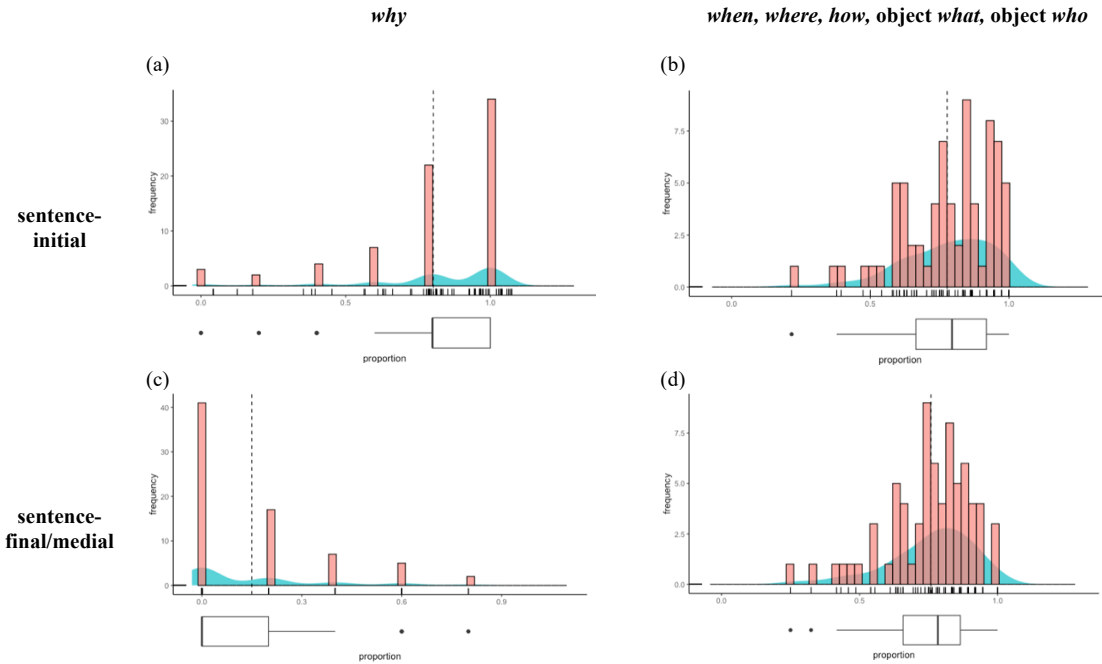


Figure 32. Histogram (frequency), density, and box plots of proportions (adherence to *wh*-phrase distributional pattern, acceptability), stratified by position and *wh*-phrase type; broken line indicates mean

Table 46. Intra- and interspeaker variation scores by position and *wh*-phrase type (highlighted cells indicate high degrees of inconsistency)

<i>wh</i> -phrase type	Position	Mean intraspeaker consistency score	SD	Interspeaker pattern inconsistency score
<i>why</i>	sentence-initial	0.8377	0.2066	0.2466
	sentence-medial	0.3484 (below 0.5)	0.1930	0.554 (above 0.5)
<i>how/when/where</i> /object <i>who</i> / object <i>what</i>	sentence-initial	0.7771	0.1657	0.2133
	sentence-medial or final	0.7587	0.1572	0.2071

6.6.2.2 Regression analysis

My model of (general) adherence to *wh*-question distributional pattern, based on acceptability judgments, indicates a main effect of age and proficiency in the source languages of Lánng-uè (Table 47). There are no main effects of sex. Age and proficiency can be used to reliably predict speakers' adherence to the pattern. Specifically, an examination of the marginal effects in Figure 33 shows that a significant portion of pattern-non-conforming judgments/ratings can be traced back to old speakers and those not proficient in the source languages of Lánng-uè. The older (Figure 33a) and less proficient a speaker is in the source languages (Figure 33c), the less likely they are to follow the pattern.

Table 47. Linear regression results – likelihood to adhere to *wh*-question distributional pattern based on acceptability judgments (observations = 6,048, conditional $R^2 = 0.05$, random intercepts for participant)

Predictors	Estimates	SE	CI	<i>p</i>
(Intercept)	0.83	0.04	0.76 – 0.91	<0.001
Age	0	0	-0.00 – -0.00	0.002
Sex (male vs. female)	0.01	0.02	-0.04 – 0.06	0.706
Proficiency (Source languages)	0.03	0.01	0.01 – 0.05	0.006

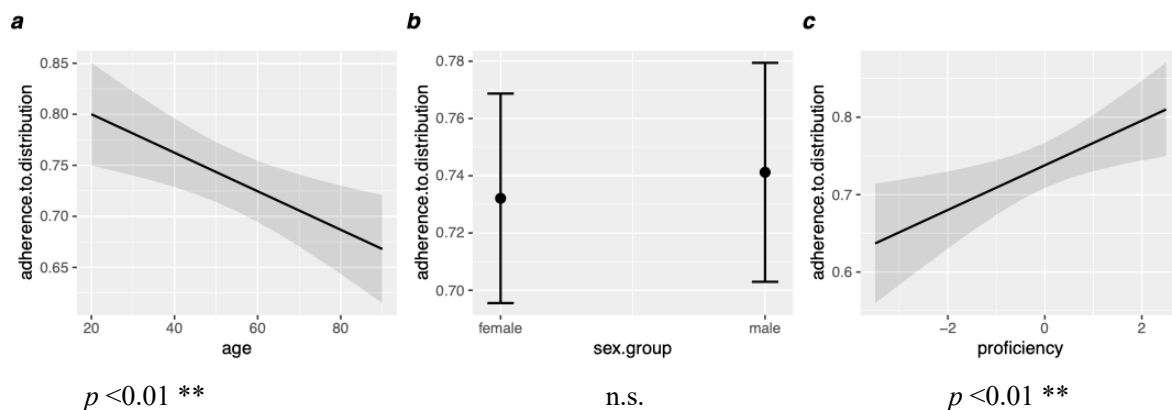


Figure 33. Marginal effects of age, sex, and proficiency on likelihood to adhere to *wh*-question pattern based on acceptability judgments

In what follows, I report the results of regression analyses conducted on four subsets of the acceptability rating data: (1) ratings of constructions with sentence-initial *why*, (2) ratings of constructions with sentence-medial *why*, (3) ratings of constructions with sentence-initial *how/when/where/object who/object what*, and (4) ratings of constructions with sentence-medial or final *how/when/where/object who/object what*, in that order.

The results of the first subset regression analysis indicate no evidence of main effects of age, sex, or proficiency in the default *wh-in-situ* languages on acceptability ratings for constructions with sentence-initial *why*. The variation in the ratings (e.g., lower ratings for sentence-initial *why*-questions) cannot be accounted for by any of these sociolinguistic factors (Figure 34).

Table 48. Linear regression results – ratings of sentence-initial constructions featuring *why*-phrases (observations = 360, conditional $R^2 = 0.274$, random intercepts for participant and item)

Predictors	Estimates	SE	CI	<i>p</i>
(Intercept)	0.6	0.2	0.20 – 0.99	0.003
Age	-0.002	0.003	-0.01 – 0.00	0.543
Sex (male vs. female)	0.09	0.11	-0.13 – 0.31	0.434
Proficiency (Source languages with default <i>wh-in-situ</i>)	0.02	0.05	-0.08 – 0.13	0.651

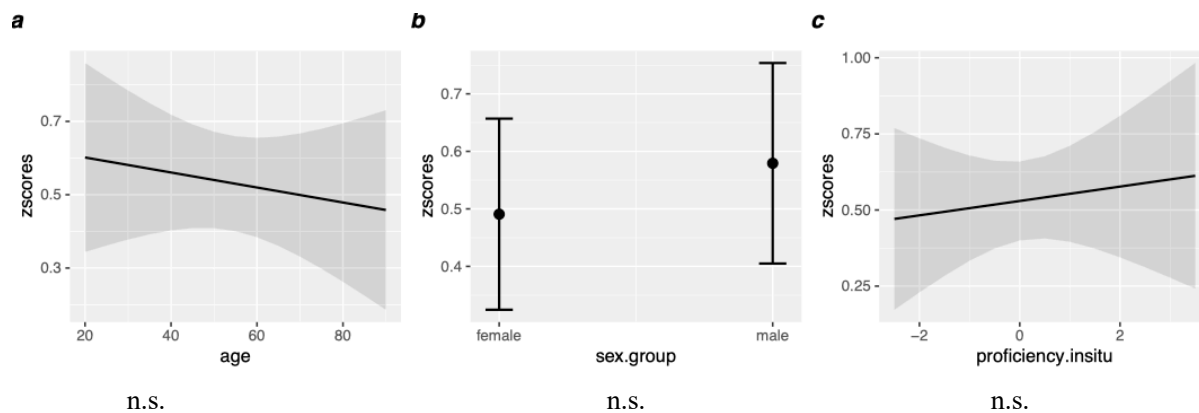


Figure 34. Marginal effects of age, sex, and proficiency in *insitu* languages on ratings of sentence-initial constructions featuring *why*-phrases

The results of the second subset analysis indicate main effects of age and proficiency in languages with default *wh-in-situ* on the participants' ratings of constructions featuring sentence-medial *why*-phrases; no evidence of such effects was found for sex (Table 49). A significant portion of the variation in ratings (i.e., higher than usual ratings for constructions with sentence-medial *why*) is associated with younger speakers (Figure 35a) and those highly proficient in languages with default *wh-in-situ* (Figure 35c).

Table 49. Linear regression results – ratings of sentence-medial constructions featuring *why*-phrases (observations = 360, conditional $R^2 = 0.191$, random intercepts for participant and item)

Predictors	Estimates	SE	CI	<i>p</i>
(Intercept)	0.94	0.16	0.62 – 1.26	<0.001
Age	-0.01	0	-0.01 – -0.00	0.022
Sex (male vs. female)	0.13	0.09	-0.05 – 0.31	0.159
Proficiency (Source languages with default <i>wh-in-situ</i>)	0.1	0.04	0.02 – 0.18	0.019

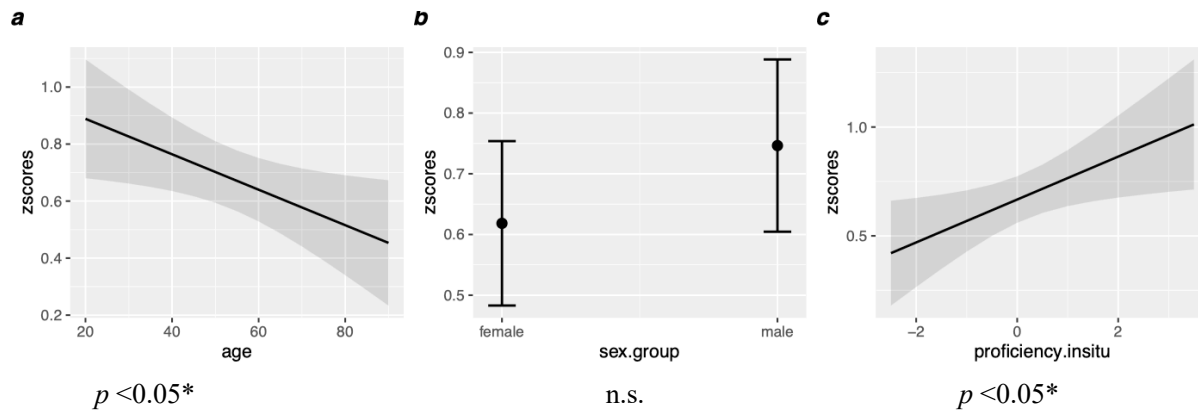


Figure 35. Marginal effects of age, sex, and proficiency in *insitu* languages on ratings of sentence-medial or final constructions featuring *why*-phrases

My model of ratings of sentence-initial constructions featuring *how/when/where/object who/object what*-phrases (i.e., third subset analysis) indicates a main effect of proficiency in languages with default *wh*-fronting, but none for age or sex (Table 50). The bulk of the variation in ratings (i.e., higher than usual ratings for constructions with sentence-initial *how/when/where/object who/object what*) is only associated with speakers who have low proficiency in languages with default *wh*-fronting (Figure 36c).

Table 50. Linear regression results – ratings of sentence-initial constructions featuring *how/when/where/object who/object what*-phrases (observations = 2,667, conditional $R^2 = 0.106$, random intercepts for participant and item)

Predictors	Estimates	SE	CI	<i>p</i>
(Intercept)	-0.73	0.11	-0.95 – -0.52	<0.001
Age	0	0	-0.00 – 0.01	0.411
Sex (male vs. female)	0	0.06	-0.12 – 0.12	0.998
Proficiency (Source languages with default <i>wh</i> -fronting)	-0.11	0.03	-0.18 – -0.05	<0.001

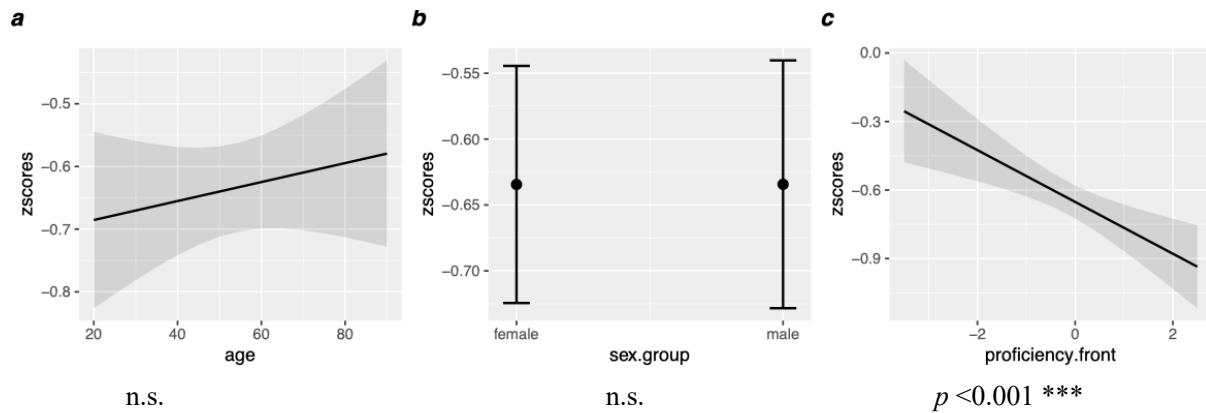


Figure 36. Marginal effects of age, sex, and proficiency in *wh*-fronting languages on ratings of sentence-initial constructions featuring *how/when/where/who/what*-phrases

My model of ratings of sentence-medial or final constructions featuring *how/when/where/object who/object what*-phrases (i.e., fourth subset analysis) indicates a main effect of proficiency in languages with default *wh*-fronting, but no evidence of main effects for age or sex (Table 51). A significant portion of the variation in ratings (i.e., lower than usual ratings for constructions with sentence-medial *how/when/where* or sentence-final object *who/what*) is only associated with speakers who have low proficiency in languages with default *wh*-fronting (Figure 37c).

Table 51. Linear regression results – ratings of constructions featuring sentence-medial *how-/when-/where*-phrases and sentence-final object *who-/what*-phrases (observations = 2,661, conditional $R^2 = 0.07$, random intercepts for participant and item)

Predictors	Estimates	SE	CI	<i>p</i>
(Intercept)	0.57	0.1	0.39 – 0.76	<0.001
Age	0	0	-0.00 – 0.00	0.313
Sex (male vs. female)	-0.01	0.05	-0.11 – 0.09	0.861
Proficiency (Source languages with default <i>wh</i> -fronting)	0.09	0.03	0.04 – 0.14	0.001

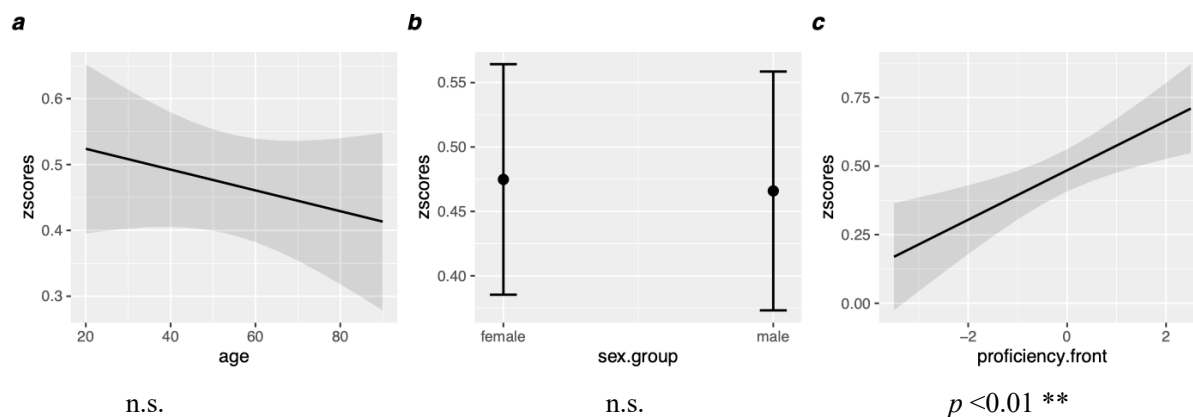


Figure 37. Marginal effects of age, sex, and proficiency in *wh*-fronting languages on ratings of sentence-medial and final constructions featuring *how/when/where/object who/object what*-phrases

6.7 Discussion

6.7.1 Hypotheses 1 & 2: Spread and stability

The results supported two of my hypotheses: that the position pattern will be highly widespread and stable. To recapitulate, I found high levels of spread for general adherence to the position distributional pattern. Most of my speakers placed *why*-phrases sentence-initially and placed *how/when/where-phrases* in the sentence-medial position or object *who/what*-phrases in the sentence-final position at least once. They also rated constructions that adhered to the pattern (i.e., constructions with sentence-initial *why* phrases, constructions with sentence-medial or final *how/when/where/object who/object what* phrases) high or rated constructions that did not adhere to the pattern (i.e., constructions with sentence-medial *why* phrases, constructions with sentence-initial *how/when/where/object who/object what* phrases) low at least once. In addition to spread, I have observed high rates of consistency at the individual and group level for general adherence to the pattern in both production and acceptability data. Overall, I found evidence that the position pattern exhibits high degrees of spread and stability within the community.

The patterns of spread and stability are expected if Lánnang-uè has a high degree of languageness. In other words, a possible reason why these patterns exist is because Lánnang-uè is highly language-like – an argument I made in previous work (Gonzales 2018; Gonzales and Starr 2020) and in Chapter 3 to Chapter 5.

Despite evidence of spread and stability (two hallmarks of languageness), I acknowledge the possibility that Lánang-uè is susceptible to the influence of external factors (i.e., Lánang-uè may be viewed as not being fully crystallized), given that it is situated in a complex, multilingual setting. With this in mind, what factors reinforced the spread and stability of the position pattern?

A possible factor involves “congruence” or cross-linguistic availability/similarity of features (Baptista 2020:162; Matras and Sakel 2007). I argue that most of my speakers continued to adhere to the pattern because components of this distributional pattern (i.e., sentence-initial, sentence-medial, or sentence-final structure) can be found in multiple languages that my speakers reported having at least some proficiency in: Tagalog, English, Hokkien, and Mandarin. As described in Section 6.2, Mandarin and Hokkien can have *why*-phrases in the sentence-initial position, in very specific contexts (e.g., topicalization). Mandarin, Hokkien, Tagalog, and English all have *how-/when-/where*-phrases in the sentence-medial position and object *who-/what*-phrases in sentence-final position. It is very likely that speakers (subconsciously) noticed the unique sentence-initial structure of *why*-questions in Mandarin⁷⁸ and Hokkien. They may have also noticed that Mandarin, Hokkien, Tagalog, and English all have sentence-medial or final phrases for *how-/when-/where-/object who-/object what*-questions. The salience of these features in the linguistic repertoire of my speakers might have encouraged the continued spread and stability of the position pattern in Lánang-uè, due to its reinforcing effect.

Social factors may have also played a role in popularizing and stabilizing the distributional pattern. However, I hesitate to discuss them in absence of concrete evidence.

6.7.2 Hypothesis 3: Structured variation (Systematicity)

6.7.2.1 General patterns

The results supported my hypothesis that variation in the position pattern is structured – that it will be conditioned by at least one sociolinguistic factor, such as age or sex.

I found that age conditioned the variation in the pattern. This, again, is suggestive of high degrees of languageness in Lánang-uè, as established contact languages also have

⁷⁸ I should note again that many Lannangs are not continuously exposed to Mandarin, so Mandarin here and throughout should not be regarded as an equal partner with Hokkien or Tagalog and English.

sociolinguistically conditioned variation (e.g., Singlish, Baba Malay, Light Warlpiri, Gurindji Kriol) (Starr and Balasubramaniam 2019; Lee 2014; Meakins and O’Shannessy 2010).

6.7.2.2 Type-specific patterns

The results partially supported my hypotheses. I found the following:

1. **Asymmetric variation.** A significant part of the pattern-non-adhering behavior in both production and acceptability data is related to *why* questions, particularly constructions with sentence-medial *why*.
2. **Age.** In the context of production, younger speakers tended not to adhere to the position pattern compared to older speakers. In acceptability judgments, the situation is reversed: older speakers were – in general – less consistent in making judgments that adhered to the pattern compared to younger speakers. This is surprising, as I expected uniform effects in production and acceptability.
3. **Sex.** There was no evidence of sex conditioning the bulk of pattern non-conforming behavior in acceptability and production.

Some questions naturally emerged from my findings:

1. Why is a significant portion of the variation in production and a subset of the variation patterns in acceptability – particularly variation involving constructions with sentence-medial *why* – associated with younger speakers?
2. What can explain the asymmetry in effects? Specifically, why do we have evidence of the effect of age on variation in general adherence to the pattern in production but no such evidence in acceptability?
3. Why don’t we have evidence of sex conditioning the variation at all?

I currently do not have a satisfying explanation for the lack of evidence for sex-conditioned patterns of variation (question 3). I can, however, provide some answers for the rest of the questions.

6.7.2.3 Younger speakers and pattern-non-conforming sentence-medial *why* constructions

One possible explanation for why many younger speakers tended to use constructions with sentence-medial *why* frequently and rated these constructions as highly acceptable is social. Youth-specific attitudes towards Lánnang-uè (and its features) may have contributed to the increased variation observed in many younger speakers.

In my interviews and ethnographic work, I have found that younger speakers tended to view Lánnang-uè as *konyò*-style Hokkien (see also, Chapter 6.7.2). That is, they perceived Lánnang-uè to be a variety of Hokkien used by elite, status-conscious, demanding, and privileged speakers (Reyes 2017:213):

(530) *Hahaha. Di naman ako ganyan magsalita* <laughing and crying emoji> *Chinese conyo*
<laughing emoji>

‘Hahaha. I don’t really speak like that! Chinese [Hokkien] konyo.’

(Facebook post written in Filipino in response to a job advertisement by *The Lannang Archives* written in Lánnang-uè, Lannang female speaker, 20s, posted May 2019)

(531) *Oh, kasí híge.. khâ siaûdien láng ulê kông yá konyò dâw ...*

‘Oh, because the... younger people have been saying that it is very conyo...’

<CLIN-19-68:13387>

(532) *Oo, sobrang konyo yon.*

‘Yes, that [Lánnang-uè] is so conyo.’

<CLIN-19-93:13925>

Older speakers either do not know what *konyò* is or do not view Lánnang-uè as *konyò* Hokkien:

(533) *Hige si shammih ah?*
‘What is that [*konyò*]?’
<CLIN-19-91:13386>

(534) *Hm, so ako personally, guâ buē khuatloh as-as konyò at àll, î sī piêntsuê necessity if-if you want... Hokkien in your life.*
‘Hm, so I personally don’t view it [Lánnang-uè] as *konyò* at all. It becomes a necessity if you want Hokkien in your life.’
<CLIN-19-101:15872>

Given that Hokkien is *wh-in-situ* by default, and given that innovative syntactic structure is a classical *konyò* feature (Reyes 2017:214), it is likely that many younger speakers interpreted (or perceived) the default sentence-initial *why* feature in Lánnang-uè as a *konyò* Hokkien innovation. In other words, sentence-medial or sentence-final is the ‘default’ or ‘proper’ structure for these younger speakers whereas the sentence-initial position of *wh*-phrases has acquired the social meaning *konyò*. As such, many younger speakers may have tried to be ‘proper’ by attempting to increase their use of sentence-medial *why* and rate constructions with sentence-medial *why* as highly acceptable. In effect, these practices ‘Hokkienify’ their Lánnang-uè or make Lánnang-uè appear ‘proper’. They can be described as practices that counter the *konyò*-ness perceived in Lánnang-uè (and the older speakers). There is preliminary evidence for this. In a series of judgment tasks involving Lánnang-uè constructions with sentence-initial and non-sentence-initial (sentence-medial or final) *wh*-phrases (conducted online, January 2022), most of my young informants judged sentence-initial constructions such as *Kânâ papá bêh phâh shótī à?* ‘Why did father hit little brother?’ and *Tisí papá bêh phâh shótī à?* ‘When did father hit little brother?’ as appearing ‘*konyò*’. There were informants who did not associate sentence-initiality with *konyò*-ness, associating them instead with awkwardness; however, these informants formed the minority. Pending a more systematic investigation, youth-specific stylistic practice (i.e., increased use and higher ratings for ‘proper’-style constructions) can partially explain a significant amount of age-conditioned variation in production and acceptability data.

Another possible explanation for the link between non-conformity (e.g., sentence-medial *why* constructions) and younger speakers is a potential change-in-progress. The profusion of

sentence-medial *why* constructions among younger speakers could indicate that the *wh*-position pattern in Lánnang-uè is starting to change, specifically towards the direction of Hokkien, which has sentence-medial *why* constructions by default. If the non-conforming pattern observed is indeed indicative of change and innovation, then it would not be surprising to have younger speakers associated with this non-conformity, as young speakers have often been reported to introduce innovations to conventionalized systems (Milroy and Milroy 1985; Palacios-Martínez 2018). I have also found evidence of younger speakers innovating in the morphological and phonological domains (Gonzales 2018; Gonzales and Starr 2020). Given these patterns, my claim that the pattern-non-conformance observed reflects a change-in-progress led by younger speakers has some merit. However, because I do not have concrete evidence of a ‘before’ period for Lánnang-uè – (oral) Lánnang-uè data before the 2010s – I am much more reluctant to commit to this explanation compared to the previous one regarding *konyò*-ness. In the absence of evidence, this change-in-progress explanation should be examined with great caution.

6.7.2.4 Age effect asymmetry

As reported earlier, older speakers in my sample generally produced constructions that followed the pattern more frequently than younger speakers, who tended to use certain constructions such as sentence-initial *why* more restrictively due to the *konyò*-ness attached to it (see 6.7.2.3). However, older speakers were more inconsistent than younger speakers in rating constructions in accordance with the distributional pattern. This asymmetry may be due to generational differences in perceived ownership of particular distributional features. While there is evidence that the distributional pattern originated from older speakers (or speakers even older than those in my sample, as suggested by the age effect using apparent time lenses), older speakers claimed that certain features of the distributional pattern did not come from them. Directly after the acceptability experiment, three old speakers in their 80s commented that the *tôtiaù* ‘inverted’ structures (i.e., sentence-initiality) came from the young speakers and that these speakers popularized it. They admitted giving higher-than-usual (i.e., average) ratings for these constructions, saying that they are somewhat acceptable because younger speakers are doing it, leading to greater variation or inconsistency in acceptability ratings. I did not encounter the same sentiments in my interviews with younger speakers. Although my younger speakers did not explicitly tell me the sentence-initial *why* feature belonged to them, their acceptability judgments,

which showed the higher levels of adherence to the pattern, suggests that they have claimed ownership of the sentence-initial *why* feature and the distributional pattern. Linguistic ownership can partially explain why younger speakers have significantly higher acceptability ratings for pattern-conforming structures compared to older speakers. This account is not incompatible with my earlier finding (Section 6.7.2.3), where I found high rates of acceptability for sentence-medial *why* among younger speakers. At first glance, this finding seems to contradict my account of the youth taking ownership of the distribution, as high acceptability for sentence-medial *why* constructions does not conform to the pattern. However, if we take the discussion about different linguistic styles in Lánnang-uè (i.e., *konyò* and ‘proper’) into consideration, the conflict is resolved. The younger speakers may have very well normalized (and accepted) both sentence-medial *why* and sentence-initial *why* in their Lánnang-uè. They may have expanded the stylistic repertoire of Lánnang-uè by using constructions with sentence-medial *why* to stylize their utterances as ‘proper’ and using constructions with sentence-initial *why* to stylize their utterance as *konyò*. Overall, I argue that the production-acceptability asymmetry in the context of age effects is partially motivated by the participants’ perceptions of ownership regarding particular distributional features.

6.7.3 Hypothesis 4: Linguistic independence

My results yielded no evidence of correlations between proficiency in Lánnang-uè’s four source languages (high proficiency) and variation patterns (pattern-non-conforming linguistic behavior) with respect to production, specifically. They, however, showed evidence of some correlations between proficiency and conformance to patterns in the context of acceptability. Specifically, I found the following:

1. Positive correlations. A large portion of the higher ratings for sentence-initial *how/when/where/object who/object what* constructions and lower ratings for sentence-medial or final *how/when/where/object who/object what* constructions (ratings that are not consistent with the pattern) came from speakers who had relatively low proficiency in the *wh*-fronting languages (Tagalog and English).

2. Negative correlation. A sizable portion of the higher ratings for sentence-medial *why* (ratings that are not also consistent with the pattern) can be traced back to speakers who had relatively high proficiency in default *wh-in-situ* languages (Hokkien and Mandarin).
3. No evidence of correlation. There is no evidence that the bulk of the lower ratings for sentence-initial *why* (ratings that are not also consistent with the pattern) is conditioned by proficiency in default *wh-in-situ* languages (Hokkien and Mandarin).

There are two interpretations for this set of findings, and depending on the interpretation, the findings may or may not be useful for evaluating my hypothesis regarding linguistic independence.

6.7.3.1 Testing the hypothesis using results from a linguistic transfer perspective

If one interprets the negative correlation as a causal relationship where proficiency (high proficiency) in the source languages affected (or contributed significantly to) the patterns of variation, then the results provide mixed evidence for linguistic independence. Under this interpretation, the lack of a negative correlation indicates that the speakers' possession of knowledge of the source language *wh*-question constructions did not affect their knowledge of Lánnang-uè *wh*-question constructions. It indicates that high proficiency in the source languages is not responsible for the occasional non-conformance to some of the *wh*-question distributional patterns. It can be interpreted as evidence for Lánnang-uè's linguistic independence. The presence of a negative correlation, under this interpretation, indicates that high proficiency in the source languages (or rather, knowledge of other languages' *wh*-question constructions) partially motivates speakers to deviate from some of the *wh*-question patterns in Lánnang-uè (or "transfer" structure from the source languages to Lánnang-uè) (Thomason 2001; Hermans et al. 2003; Siegel 2012:187; Klaus et al. 2018; Pham et al. 2018). This would be evidence of linguistic dependence and, thus, evidence against languageness.

6.7.3.2 Testing the hypothesis using results from an agentive perspective

If one interprets the proficiency variable in the negative correlation as 'expression of proficiency' rather than actual proficiency (e.g., speakers not conforming to the *wh*-phrase

pattern because they want to express high proficiency in Chinese),⁷⁹ then the results do not say anything useful about the linguistic independence of Lánng-uè's *wh*-phrase pattern. They, however, support the hypothesis relevant to systematicity or structured variation; they unequivocally support the hypothesis that Lánng-uè is highly language-like. This is because the acceptability of pattern-non-conforming constructions can be interpreted as a stylistic choice anchored in Lánng-uè's linguistic system (speakers' skillful manipulation of linguistic resources to express particular social meaning) instead of a consequence of (subconscious) linguistic transfer or borrowing (Eckert 2005; Hall-Lew et al. 2021). For example, the participants' higher acceptability for Hokkien-like sentence-medial *why* constructions mean/signal being highly proficient in Hokkien.

If the second interpretation is taken as fact, then why is variation only used to index high proficiency in the context of acceptability but not production? One potential reason is awareness. The production and acceptability experiments appear to involve varying levels of awareness. In the production experiment, speakers were not aware that I was interested in the position of their *wh*-phrases. However, some speakers became aware of this in the acceptability experiment, where the position of the *wh*-phrase was explicitly manipulated. Some evidence of awareness was found post-experiment, where these speakers asked if the variable that I am interested in was position. In addition, one old speaker in their 80s asked if they could withdraw from the experiment mid-way, saying that I should just do the experiment for her – by rating all sentence-initial constructions low and sentence-medial or final constructions high. If the scale-rating experiment did indeed involve a higher degree of awareness than the production experiment, varying levels of awareness would explain the production-acceptability asymmetry. It seems that speakers do not use variation in *wh*-question constructions to index command in a particular language when attention is not given to the patterns of variation (as evidenced by absence of proficiency effects in the production or 'lack of awareness' context). However, when the participants were made aware of what constructions or features were in the spotlight (the acceptability or 'awareness' context), they would actively use these linguistic resources to index command in a particular language. This is reminiscent to what was observed by Dodsworth

⁷⁹ Note that the proficiency factors I used in my analysis were self-reported. 'Proficiency' can thus be interpreted as 'expression of proficiency in a particular source language.'

(2005) and in the conjunction and preposition chapter of this study (Chapter 5.6), where awareness was found to condition patterns of linguistic behavior.

It should be noted that awareness is not a necessary condition for variation (Dodsworth 2005). Speaker groups (e.g., younger speakers) do not have to be actively aware of a particular linguistic resource and its relevant social meaning to employ (knowledge of) that resource. This would explain the asymmetric effects of awareness (differences in production and acceptability experiments) on the age-results and proficiency-related results.

How about the results involving the positive correlations? Under the second interpretation, the findings suggest that some speakers preferred pattern-non-conforming syntactic constructions because they want to index ‘low proficiency’ in a particular source language. Upon closer examination of the results, I found that the high ratings were still relatively low and the low ratings were relatively high – most participants did not find constructions with sentence-initial *how/when/where/object who/object what* acceptable (Figure 36c, *z*-scores below zero); they also found constructions with sentence-medial *how/when/where* or sentence-final object *who/what* acceptable (Figure 37c, *z*-scores above zero). The ‘pattern-non-conforming’ behavior observed was actually pattern-conforming. This finding renders the suggestion that speakers prefer pattern-non-conforming syntactic constructions to index ‘low proficiency’ moot.

What, then, can explain the positive correlations? I argue that the correlations are due to some dissimilation process instead of stylistic practice: speakers who had more knowledge of *wh*-fronting languages (i.e., Tagalog and English) have most likely used this knowledge to make Lánnang-uè distinct from the languages they are proficient in, as they actively and consciously compare structures from languages they know with structures in Lánnang-uè. They relied on their knowledge of *wh*-fronting Tagalog and English and rated constructions with sentence-initial *how/when/where/object who/object what* more unacceptable and constructions with sentence-medial or final *how/when/where/object who/object what* more acceptable to solidify the contrast between the source languages with *wh*-fronting and Lánnang-uè. Evidence of active comparison between Lánnang-uè and the *wh*-fronting languages can be found post-experiment, where three speakers highly proficient in Tagalog and English explicitly commented that constructions with sentence-initial *wh*-phrases looked similar to Tagalog and English and are, thus, not well-formed.

Overall, in the second interpretation, the patterns of sociolinguistic variation are regarded as a component of Lánnang-uè's linguistic system. Some (pattern-non-conforming) *wh*-question constructions appear to be used to express social meaning – specifically, great command of Lánnang-uè's source languages. The sociolinguistic patterns, under this view, are irrelevant to the hypothesis of linguistic independence (Hypothesis 4) but provide further support for the hypothesis involving systematicity or structured variation (Hypothesis 3).

6.8 Conclusion

Do the seemingly high rates of variation in the *wh*-phrase position distributional pattern challenge my argument that Lánnang-uè exhibits a high degree of languageness? The answer heavily leans toward no, based on an examination of four properties of languageness: degree of spread, degree of stability, presence of systematicity or structured variation, and degree of linguistic independence.

I found high rates of pattern spread within the sample and high rates of pattern consistency within the individual, and low rates of variation pattern heterogeneity between my speakers (high degrees of stability). Most speakers of Lánnang-uè consistently formed *wh*-question constructions that adhered to the distributional patterns involving *wh*-phrase position – providing support for the argument that Lánnang-uè has high degrees of languageness.

Furthermore, I found that the variation observed in the adherence to the position pattern is sociolinguistically structured or systematic. Certain *wh*-question patterns – embedded with social meaning (Labov 1972; Eckert 1989) – were avoided or preferred in particular social conditions. Constructions where the adjunct *why*-phrase is in the sentence-initial position were, for example, avoided by many younger speakers because they associate the structure with the social characteristic *konyò*; constructions where the phrase is sentence-medial were preferred because they index the characteristic 'proper'. The systematic use of variation to express particular social meanings (e.g., *konyò*, command in a particular source language) can be regarded as evidence for Lánnang-uè's languageness (Weinreich et al. 1968).

I also found mixed evidence of linguistic independence in Lánnang-uè's *wh*-phrase position pattern, assuming that the relationship tested between the source language proficiency variable and variation involves actual proficiency affecting variation. The production results suggest that the pattern is not influenced by Lánnang-uè's source languages. The acceptability

results, on the other hand, imply that the pattern is partially influenced by these languages. Under this assumption, the results overall suggest that the *wh*-phrase pattern is, for the most part, not influenced by (knowledge of *wh*-question constructions in) Lánnang-uè's source languages – evidence in favor of linguistic independence (Hypothesis 4).

A major limitation of this study is the imbalance of elicited *why* and *how/when/where/object who/object what* constructions. There were significantly more *how/when/where/object who/object what* constructions compared to *why* constructions in my dataset. Although I mitigated the token imbalance by adopting regression methods, having a balanced sample would allow me to generalize about the conditioning effect of *wh*-phrase type on *wh*-phrase position as well as the effect of sociolinguistic factors on (non)adherence to the pattern with more certainty. Ideally, I would have designed the game such that the number of 'why' stimuli/cards matched the number of 'how', 'when', 'where', object 'what', and object 'who' cards, dramatically increasing the number of stimuli and time needed for the sessions. However, because participants in my pilot study complained about long sessions, I decided to proceed with a game that allowed me to elicit the same number of questions across all *wh*-types in the least amount of time, which meant an evenly distributed and representative sample of *wh*-questions in Lánnang-uè, but an uneven *why* and *how/when/where/object who/object what* distribution. Because of this imbalance, I hesitate to be too categorical about my findings. Future research should consider attempting to replicate the results of this study by fitting the same models on a dataset that has relatively equal *why* and *how/when/where/object who/object what* constructions. While there are limitations regarding the data, my findings should not be discounted, as they could still provide us insights into the syntactic patterns and variation in Lánnang-uè.

The anecdotally observed high rates of variation in the *wh*-phrase distributional pattern – the impetus of this study – do not pose a major challenge to the argument that Lánnang-uè is highly language-like. A closer examination of the variation patterns in a more representative sample in a controlled experimental setting revealed that the syntactic pattern is highly widespread and stable in Lánnang-uè, similar to features and patterns at the morphological, prosodic, and lexical levels. The findings indicate that the variation is systematic (i.e., socially-conditioned), similarly to what had been found in linguistic varieties labeled as 'languages' (e.g., Singlish, English) (Labov 1972; Starr and Balasubramaniam 2019). They also provide some indication of linguistic independence – only certain aspects of the *wh*-phrase distributional

pattern were influenced by (knowledge of *wh*-question constructions in) Lánnang-uè's source languages. Together with the results of the previous chapters and previous research (Gonzales 2018; Gonzales and Starr 2020), the findings of this chapter provide further evidence that Lánnang-uè has a high degree of languageness.

Chapter 7 : Epilogue

7.1 *Brief review*

There is, to date, no research that comprehensively and systematically investigates Lánnang-uè as used in Manila. Little is known about whether the variety is a product of ad-hoc code-switching between English, Tagalog, Mandarin, and Hokkien – as reported by many members of the community – or whether the mixing is systematic and structured. Not much is also known about the degree of spread and stability of its features. Although some speakers claim that Lánnang-uè is systematic and has features that are consistently used by most of the Lánnang-uè-speaking community, many view the variety as unstructured and unstable, using the existence of extensive and “random” (55-year-old Lannang male, CLIN-18-1:243) intra- and inter-speaker variation as evidence. Many also view Lánnang-uè as a constellation of idiolects and familects with no common pool of features; they think that many of its linguistic features are idiolectal or familectal (i.e., not widespread in the community). In other words, many speakers believe that Lánnang-uè has a low degree of languageness.

The scarce research that has attempted to examine whether systematicity, spread, and stability exist in Lánnang-uè (Gonzales 2017a; Gonzales 2018; Gonzales and Starr 2020), all led by me, provided some evidence of a mismatch between popular folk belief and actual linguistic practice. The results in these studies all suggested that, from the perspective of production, Lánnang-uè is highly conventionalized and stable – at odds with what many speakers reported in my ethnographic work. The speakers also showed variation, but not randomly occurring variation, as many speakers claimed. The variation is instead systematically conditioned by sociolinguistic factors, such as age and sex. However, these investigations only focused on three features: derivational affixes in the nominal domain, vowel monophthongs, and question tags. And as I have emphasized at the outset of this dissertation, a few features are not enough to generalize about conventionalization, stability, and sociolinguistic patterns of variation in

Lánnang-uè. They certainly cannot, in isolation, be used to generalize about the languageness of the variety.

The current project attempted to remedy this. I conducted a comprehensive investigation of Lánnang-uè, analyzing linguistic data (multiple features) with respect to key properties of languageness, focusing on systematicity, (feature) spread, stability, and linguistic independence. In employing a combination of approaches to examining Lánnang-uè (e.g., descriptive, experimental, computational, corpus-based, sociolinguistic), I got closer to answering the question of how language-like Lánnang-uè is.

In Chapter 3, I provided a description of the variety. Based on it, I argued that Lánnang-uè exhibits high degrees of systematicity; there was also ample evidence of many stable and widespread features in the variety. Relying on corpus data, elicitations, and judgments, I showed that Lánnang-uè is far from a random mix of Hokkien, Tagalog, English, and Mandarin. For many of its features, I was able to definitively conclude that the features are widespread and stable due to minimal variation. But for some, I was not able to, due to seemingly high rates of variation. These features also had variation that seemed to be ‘free’, ‘random’, or unstructured, which can be viewed as an argument against high degrees of languageness in Lánnang-uè.

In the next three chapters (Chapter 4 to Chapter 6), I conducted a more systematic investigation of seven of these features, distributed across three levels of language to test whether they pose a significant challenge to Lánnang-uè’s languageness:

Prosody

- Lexical tone
- Duration-cued stress
- CV/CVT tone distributional pattern
- CVR-English/CVR-Tagalog tone distributional pattern

Lexicon

- Conjunction distributional pattern
- Prepositions distributional pattern

Syntax

- *Wh*-phrase position distributional pattern

I found evidence of high degrees of spread and stability in all these features. I also found that the patterns of variation found in these seven features are all conditioned by at least one sociolinguistic factor (e.g., speaker attitudes, age, sex). The variation is structured. Finally, I found mixed evidence of linguistic independence in my investigation – there is both evidence for and against the claim that the features/patterns are influenced by the source languages of Lánnang-uè.

7.2 *Research questions revisited*

So, going back to the research questions posed in Chapter 1,

1. Where does Lánnang-uè fall in the cline of languageness?
 - a. Is Lánnang-uè highly systematic? Are the patterns of variation in it structured? That is, can they be explained by (socio)linguistic factors?
 - b. Are the features of Lánnang-uè used at all by most speakers in the Lánnang-uè-speaking community or only a small subset?
 - c. Is Lánnang-uè highly stable? Will speakers be consistent in the use of its features?
 - d. Is Lánnang-uè linguistically independent? Are the features/patterns of the variety independent from the features/patterns of its source languages?
2. If it has a high degree of languageness, where does it fall in the typology of contact languages? If not, what kind of contact phenomenon is it?

7.2.1 *Lánnang-uè in the cline of languageness*

Pending successful study replications, the findings of this dissertation strongly support the argument that Lánnang-uè has high degrees of languageness. My findings across chapters illustrate a scenario where speakers with varying linguistic ideologies or beliefs have relatively homogeneous linguistic behavior. Regardless of their views towards Lánnang-uè, most speakers follow the conventions with high levels of consistency, only occasionally patterning differently from these norms. Some speakers do not vary at all. The results show that a significant number of tokens that did not conform to the widespread patterns (conventions) of the variety – ‘variants’ – were systematically linked to at least one sociolinguistic factor (e.g., speaker

attitudes, age, sex), illustrating systematicity, and corroborating previous variationist research in Lánnang-uè (Gonzales 2018; Gonzales and Starr 2020). I found some unexpected sociolinguistic patterns, but I was also able to account for most of them using sociolinguistic, historical, and cognitive theories, providing more evidence for systematicity or structured variation in Lánnang-uè. From a variationist sociolinguistic perspective (Eckert 2012), the findings on sociolinguistically-conditioned variation indicate the existence of sociolinguistic conventions on top of linguistic ones. In many cases, there is concrete evidence that variation is used to express particular social meanings, as part of Lánnang-uè's linguistic system. My findings relevant to language proficiency, under a 'linguistic transfer' interpretation of the results, provided some indication that the patterns/features of Lánnang-uè are relatively independent from its source languages.

Overall, there is a relatively high degree of systematicity, spread, stability, and linguistic independence across multiple features in the variety. In addition, there is also evidence of clustering of different kinds of (socio)linguistic features/patterns. For instance, I found that young speakers tended to use certain (system/pattern-non-conforming) conjunctions *and* *wh*-question constructions to avoid sounding *konyò*. I also found that my speakers, in general, used most – if not all – the features described and examined in this dissertation. Furthermore, although many speakers do not perceive it as an independent language, there are those who do, referring to it as 'secret code' and 'mixed language' (see Chapter 5.6.2). With respect to at least six criteria derived from works that adopted a sociolinguistic framework (Haugen 1966; Kloss 1968; Weinreich et al. 1968; Labov 1982; Görlach 2002; Ghyselen and De Vogelaer 2018) – systematicity, spread, stability, independence, clustering, and, partially, attitudes – Lánnang-uè is indeed highly language-like, corroborating my previous research on Lánnang-uè, which makes the same claim (Gonzales 2017a; Gonzales 2018; Gonzales and Starr 2020). My findings do not support that it is an ephemeral, ad-hoc code-switching phenomenon.

7.2.2 *Lánnang-uè in the constellation of contact languages*

Given that Lánnang-uè is highly language-like and that it is the product of contact between Tagalog, Hokkien, English, and Mandarin, what kind of contact language is it?

At the beginning of this project, I identified two possible alternatives for Lánnang-uè's status, based on my preliminary data and ethnographic work conducted between 2017 and 2018

(Gonzales 2018; Gonzales 2020). Lánnang-uè could be (1) a mixed language, as claimed by a small number of users and as suggested by preliminary sociohistorical and linguistic evidence in previous work (Gonzales 2018; Gonzales and Starr 2020), or (2) a variety of Hokkien, something that many Lannangs anecdotally reported (Uytanlet 2014:161).

Note that I did not consider nativized Tagalog, English, or Mandarin varieties as alternatives because, based on my preliminary data, there is inadequate (less than 50%) Tagalog, English, or Mandarin lexicon and structure in the variety to make such a hypothesis plausible. I also did not entertain the possibility of Lánnang-uè being a creole or a pidgin, because the sociohistorical context and linguistic features of Lánnang-uè do not fit classical descriptions of creoles and pidgins. For one, based on the history of the Lannang community, Lánnang-uè did not, for example, emerge out of contexts where “populations have been displaced forcefully via slavery or voluntarily as indentured labor” or contexts where groups of people with different backgrounds needed to find a common ‘makeshift’ language, or other documented contexts related to pidgins and creoles (Baptista 2005:34).⁸⁰ Linguistically, most of the lexicon of Lánnang-uè was not inherited from the socially dominant language (in this context, Tagalog), a characteristic that is present in many creoles like the Philippine creole Chabacano where the lexifier is Spanish (the socially-dominant language in Chabacano’s genesis).

7.2.2.1 Lánnang-uè as a mixed language?

I am unable to reject the first hypothesis, considering the dissertation findings. My examination of both the sociohistorical and linguistic aspects of Lánnang-uè showed that Lánnang-uè has patterns that overlap with patterns that characterize “mixed languages” (Thomason and Kaufman 1988; Meakins 2013:159; O’Shannessy 2020:325; Sippola 2021:255). One characteristic of mixed languages is a “split in the sources of their morphemes” and structure (Sippola 2021:255) or evidence of a composite grammar or lexicon, which I have observed in my data (Chapter 3.12). For instance, in Chapter 3 and Chapter 5, I found that certain (lexical) classes of conjunctions tended to be derived from Tagalog, Hokkien, or English. The nominal derivational system is derived from Tagalog (Chapter 3.4) (Gonzales 2018) while the article system is derived from English. Components of the aspectual system (e.g., inchoative aspect) were derived from Hokkien, while some (e.g., iterative) were derived from Tagalog (Chapter 3.5.2). The stress

⁸⁰ Many creoles emerge independently from these contexts.

system is arguably derived from Tagalog and English (Chapter 3.3.3.2); the tone system is arguably derived in large part from Hokkien (Chapter 3.3.3.1). The adjectival pre-modification structure is derived from Hokkien whereas the post-modification structure is derived from Tagalog and English (Chapter 3.4.7.1). Finally, the component of the *wh*-question system pertaining to *why*-questions (*wh*-phrase sentence-initiality) is arguably derived from Tagalog and English whereas the other components are arguably derived from Hokkien (Chapter 3.7.2).

Another hallmark of mixed languages, according to Matras and Bakker (2003) is the seemingly unconstrained incorporation of grammatical elements (Meakins 2013:190) or the incorporation of elements that in the past have been characterized as “loan proof” (Meakins 2013:190), one common example being inflectional morphology. There is evidence of this in Lánnang-uè, where I observed that the variety incorporates inflectional morphology from Hokkien (i.e., the perfective *-tioh*), Tagalog (i.e., the perfective *kakà-* and inchoative *pà-*) and English (e.g., plural *-s*) (Sections 3.4.9 and 3.5.2).

Another characteristic involves identity functions. Scholars have reported that mixed languages often develop in communities where members want to express membership in a new social group or express a new mixed identity (Meakins 2013; Sippola 2021). In the context of Lánnang-uè, its users, the Lannangs, have been observed to use their mixed code Lánnang-uè to index their mixed identity (Gonzales 2021a). Apart from explicitly linking Lánnang-uè to the hybrid Lannang identity during interviews (Gonzales 2021a), they almost never use Lánnang-uè when conversing with individuals with Hokkien heritage (e.g., Mainland Chinese visitors from the Fujian/Hokkien area); they never use Lánnang-uè to converse with Filipinos without Lannang heritage; they only use it as a ‘secret’ or ‘in-group’ code when conversing with individuals who have mixed Southern Chinese-Filipino (i.e., Lannang) cultural heritage.

These three characteristics, along with other indicators discussed in Gonzales (2018) suggest that Lánnang-uè can be analyzed as a mixed language.

Assuming that Lánnang-uè is a mixed language, one interesting thing to note is that, linguistically, Lánnang-uè does not seem to pattern after varieties that have been labelled as ‘mixed languages’, including Michif, Gurindji Kriol, and Mednyj Aleut (Meakins 2013). For example, Michif systematically derives grammatical elements in the NP domain from French and derived elements in the VP domain from Cree (Bakker 1994). Gurindji Kriol patterns similarly with Michif in that its NP grammatical elements are sourced from one language, Gurindji, while

its VP elements are sourced from Kriol (Meakins 2012). On the other hand, Mednyj Aleut derives its finite inflectional VP morphology from Russian whereas its nonfinite VP morphology, NPs, and 90% of its lexicon are derived from Aleut (Golovko 1994). With respect to social function, Lánnang-uè resembles some mixed languages (e.g., languages of community or ancestral identity, like Light Warlpiri and Gurindji Kriol). However, the linguistic profile of Lánnang-uè diverges from the linguistic profiles of these mixed languages and, to my knowledge, the profiles of other documented varieties identified as ‘mixed languages’ (see Meakins (2013) for a comprehensive list and discussion.) There appears to be no wholesale incorporation of entire paradigms or systems from different source languages (e.g., VP morphemes from Tagalog, NP morphemes from Hokkien, or lexicon from Hokkien and grammatical morphemes from English). Instead, what I observed is a finer-grained fusion of lexicon and grammatical systems. Based on my comparison of Lánnang-uè and varieties regarded as ‘mixed languages’, Lánnang-uè appears to be distinct.

7.2.2.2 Lánnang-uè as a Hokkien variety?

While there is evidence for Lánnang-uè as a mixed language, it is important to carefully assess the alternative possibility – Lánnang-uè as a Hokkien variety. Considering this scenario is justified given that many Lánnang-uè users in my ethnographic fieldwork from 2018 to 2021 claimed that Lánnang-uè is a variety of Hokkien. A comparison of features in Hokkien and Lánnang-uè sheds some light on the status of Lánnang-uè as a potential Hokkien variety:

Similarities

- a. Lexicon. Most of the vocabulary in Lánnang-uè that is least likely to be borrowed (i.e., items on the Swadesh list) (Chapter 3.2) resembles Hokkien vocabulary (MacGowan 1922; Bodman 1987; Tsai 2017).
- b. Phonemic tone inventory. Seven of the eight tones in Lánnang-uè (Chapter 3.3.3.1) resemble Hokkien tone phonemes (Tsai 2017). Only one tone phoneme – the high-III tone (Chapter 3.3.3.1) – is distinctively Lánnang-uè.

- c. Tone sandhi. Most of the tone sandhi rules in Lánnang-uè pattern after Hokkien tone sandhi rules (Chapter 3.3.3.1).
- d. Demonstratives. Like Hokkien (Chappell 2019), Lánnang-uè demonstratives have number and proximal-distal contrasts (Chapter 3.4.1).
- e. Classifiers. Lánnang-uè has a classifier system (Chapter 3.4.1) that resembles Hokkien (Chappell 2019).
- f. Numerals. The rules for numerals in Lánnang-uè (Chapter 3.4.5) resemble the rules for numerals in Hokkien (Bodman 1987).
- g. Adjectivals. The most common strategy to modify nouns using adjectivals in Lánnang-uè – the Adj + =*ē* + N construction (Chapter 3.4.7) – is the default strategy in Hokkien (Chappell 2019).
- h. Personal pronouns. The personal pronouns of Lánnang-uè (Chapter 3.4.8) are near-identical to the pronouns of Hokkien. The pronoun systems in both codes have a number, a person, and a clusivity contrast (Chappell 2019).
- i. Modals. Most of the modal markers in Lánnang-uè (Chapter 3.5.1) pattern after Hokkien (Lin 2015).
- j. Negative markers. All of the negative markers in Lánnang-uè (Chapter 3.5.8) pattern after Hokkien (Chappell 2019).
- k. Benefactive construction. The benefactive construction in the variety (Chapter 3.6) is identical to the benefactive construction in Hokkien (Chappell 2019).

Differences

- a. Phonemic vowel inventory. The phonemic vowel inventory of Lánnang-uè is larger than that of Hokkien. Lánnang-uè has 10 vowel phonemes (Chapter 3.3.1). This is in contrast with Hokkien, which has 8 vowels (Tsai 2017). A closer look at the vowels in both Lánnang-uè and Hokkien reveals that Lánnang-uè does not have three vowels that are reported to be in the Hokkien vowel inventory – it does not have the open-mid back rounded vowel /ɔ/, the mid central vowel /ə/, or the close back unrounded vowel /u/ in its phonemic inventory. The evidence from vowels suggests that, while Lánnang-uè’s vowel inventory is larger than Hokkien’s, its vowel inventory is not best analyzed as an expanded Hokkien vowel inventory.
- b. Consonant clusters. Lánnang-uè permits consonant clusters (e.g., *Spain* ‘Spain’ [spejn⁵¹]). Hokkien does not (Tsai 2017; Chappell 2019). Unlike Lánnang-uè, Hokkien subjects the loanwords to a specific phonological process (i.e., vowel epenthesis) to deal with the cluster (e.g., Hokkien *sē pāng gá* ‘Spain’ [se³³ pang³³ ga³⁵]).
- c. Demonstratives. The Lánnang-uè demonstrative system, derived from Hokkien, does not have a function contrast (Chapter 3.4.1), unlike Hokkien. In Hokkien, for example, the singular proximal demonstrative functioning as a determiner is *chit*⁸ ‘this’; the singular proximal demonstrative functioning as a pronoun is *che*^l ‘this’ (Chappell 2019:199). In Lánnang-uè, the (conventionalized) singular proximal demonstrative – regardless of function – is *tsí* ‘this’ (pronoun function: *Kasí tsí sī concèrn dī=e taitsi* ‘because **this** is a matter that concerns you’, PROT-16-NA:37167; determiner function: *tsí ge lánng ‘this person’, CFH-001).*
- d. Adjectival modification. Lánnang-uè, unlike Hokkien, has post-nominal adjectival modification (Chapter 3.4.7.1).
- e. No nominal marking suffix -a. Lánnang-uè does not have the nominal marking suffix -a, a morpheme found in Hokkien that is generally used to indicate that the

entity is a noun (e.g., *chiao-a* ‘bird’, *hi-a* ‘fish’, *tiu-a* ‘rice paddy’) (Li and Thompson 1981; Chappell 2019). In the examples given in this chapter, none of the words were suffixed by *-a* (e.g., Hokkien-derived *kaû* ‘dog’ in FRST-19-52:9266, FRST-19-114:19407-19408; English-derived *deèr* ‘deer’ in PC0001-FRST18, and Hokkien-derived *hi* ‘fish’ in Appendix A wordlist item 56). There were a few words that appear to be suffixed by nominal marking *-a*, such as *ginâ* ‘kid’ (PC0103-FRST19) and *tshangkabâ* ‘frog’ (FRST-19-119:21745). However, I argue that the *â*’s here are not suffixes, as the root words without the supposed *-a* suffix (i.e., *gin* and *tshangkab*) do not exist in the corpus.⁸¹ It seems that these words ending in *â* were imported from Hokkien into Lánnang-uè as a single morpheme, supporting the argument that Lánnang-uè does not have the suffix *-a*. Overall, Lánnang-uè’s lack of nominal marking suffix *-a* (or nominal marking suffixes in general) is a feature that is distinctively Lánnang-uè.

- f. No gender marking with *-kang* or *-bu*. In Hokkien, a subset of animate nouns such as domestic animals and fowl are marked with the suffix *-kang* for males and the suffix *-bu* for a reproductive female of the species (e.g., *tikang* ‘boar’, *tibu* ‘a sow that has produced a litter’) (Chappell 2019). In Lánnang-uè there are no such suffixes at all.
- g. No triplication of adjectives. Monosyllabic adjectives in Hokkien (e.g., *ang* ‘red’) undergo triplication to increase the intensity of predicate adjectives (e.g., *ang-ang-ang* ‘extremely red’) (Chappell 2019). This process cannot be found in Lánnang-uè, even for monosyllabic adjectives derived from Hokkien (Chapter 3.4.7).
- h. Adverbial modification. In Hokkien, adverbs can be formed by adjectival reduplication followed by the suffix *-a* (e.g., *ban-ban-a* ‘slowly’) (Chappell 2019). In Lánnang-uè, this is not permitted (Chapter 3.4.7.6).

⁸¹ Native speakers either did not understand me or rated these words low when I presented the two words to them.

- i. Progressive aspect marker. The progressive aspect in Lánnang-uè is marked only with Hokkien-derived *lè* /le⁵¹/ (Chapter 3.5.2). This is different from Hokkien, which has three progressive markers *ti*, *leh*, *tileh* (e.g., *ti chhiu-koa* ‘singing’) (Chappell 2019:204–205).
- j. Iterative *tītī*. The morpheme *tītī* in Lánnang-uè only codes iterative aspect (Chapter 3.5.2). This contrasts with Hokkien, where *tī³³tī³³* is multifunctional and can be used for its habitual or iterative function (Tan 2014:24–25). For example, the phrase *tītī kong-uè* only means ‘talk repeatedly (in a single point in time)’ (iterative) and not ‘talk regularly (as a habit)’ (habitual). In Hokkien, both readings are possible (Tan 2014:24–25).
- k. Causative *kay-*. *Kay-* is a causative prefix that is attached to a verb to derive another verb that means ‘cause an entity to be Verb-ed (by a causer)’ (Chapter 3.5.4). It cannot be found in Hokkien or any of Lánnang-uè’s source languages.
- l. The *wh*-phrase position distributional pattern. In Lánnang-uè, *why* phrases and subject argument *wh*-phrases are, by default, placed in sentence-initial position. Object argument *wh*-phrases are placed in the verb complement position (typically sentence-final) while adjunct *wh*-phrases that are not *why* (i.e., *how-*, *when-*, *where-*phrases) are placed in the preverbal position (Chapter 3.7.2). This distributional pattern cannot be found in Hokkien, which has default *wh-in-situ*. *Wh*-fronting in Hokkien is used restrictively in topicalization and wide-scope *why* constructions (Bodman 1987; Sato 2013; Cheung 2014; Yuan and Dugarova 2012).
- m. Restricted use of negative particles in yes/no question formation. One way to form yes/no questions in Lánnang-uè is the clause + negative particle strategy, which is derived from Hokkien. However, unlike Hokkien, not all negative particles (e.g., *m*) can be used (Chapter 3.5.8). The constraint distinguishes

Lánnang-uè from Hokkien, which does not have such restrictions (Chappell 2019).

The comparative evidence shows that Lánnang-uè has many features and patterns that resemble Hokkien. However, it also shows that the language has many elements (both lexical and grammatical) that are distinct from Hokkien, either patterning after Tagalog and/or English (e.g., the lack of a function contrast in Lánnang-uè demonstratives patterning after the English demonstrative system) or developing independently from its source languages (e.g., the causative *kay-*). While the abundance of Hokkien-derived elements appears to support the hypothesis that Lánnang-uè is a Hokkien variety, the plausibility of such a hypothesis is tempered by the profusion of innovations observed. The current evidence from a purely linguistic standpoint suggests that Lánnang-uè is not best viewed as a Hokkien variety. Comments about Lánnang-uè's mutual unintelligibility with Mainland Hokkien users from my fieldwork – from the perspective of both Mainlanders and Lannangs – further support this suggestion.

From a sociolinguistic perspective, the hypothesis that Lánnang-uè is a Hokkien variety appears to be supported, given that many users view the language as Hokkien (Ang See 1990; Uytanlet 2014). However, it is known that linguistic ideologies (e.g., linguistic purism) can influence the perception of contact languages as independent languages. Contact languages historically tend to be viewed by laypeople and linguists as varieties of a dominant variety (commonly the primary lexifier) even if the empirical evidence suggests that they are not. For example, Portuguese-lexifier Cape Verdean Creole (Baptista 2005:37) – a language now known to be independent from Portuguese – was characterized as a Romance dialect in the late 1800s (Coelho 1881:1). A Spanish-lexifier pidgin used in the Philippines between Japanese settlers and Filipinos, a language established to be distinct from Spanish (Fernández and Sippola 2017), was referred to as “Bamboo Spanish” in the 1950s (Reinecke et al. 1975:210). In my interviews with Lánnang-uè users, a common theme in discussions relating to Lánnang-uè is the notion of language purity or protectionism. Many community members expressed the desire to have languages that do not have any form of mixing (e.g., Hokkien as spoken in Xiamen), viewing them as superior languages over languages with ‘explicit’ mixing (e.g., Lánnang-uè). Their ideology seems to have affected their perception of Lánnang-uè: speakers who had this ideology tended to claim that Lánnang-uè is (‘broken’) Hokkien while those who did not tend to view

Lánnang-uè as a language independent from Hokkien. If one takes this language ideology out of the equation, then the findings provide evidence against the hypothesis that Lánnang-uè is a Hokkien variety.

Despite the existing evidence, I am not fully committing to rejecting this hypothesis yet in the absence of a quantitative and more systematic comparison of Hokkien and Lánnang-uè as well as in-depth investigations of the degree of mutual unintelligibility between both languages. Further research is needed to ascertain definitively whether Lánnang-uè is independent from Hokkien, although the linguistic and sociolinguistic findings both suggest that such is the case.

7.2.2.3 Lánnang-uè as a novel contact language

In the previous subsections, I have presented evidence of Lánnang-uè as a mixed language. I have also presented some evidence against Lánnang-uè as a Hokkien variety but remain cautious in committing to reject the hypothesis that Lánnang-uè is Hokkien until further evidence is furnished. So, what type of contact language is Lánnang-uè? Based on preliminary evidence, the most likely scenario is that Lánnang-uè is a mixed language. If we shift away from the idea of rigid typological categories and align with a view of a linguistic continuum (Baptista 2015), then Lánnang-uè is situated somewhere on a continuum between ‘Hokkien’ and a ‘mixed language’, heavily leaning towards ‘mixed language’. A very unlikely scenario (but still possible, pending evidence) based on what I found is that Lánnang-uè is Hokkien or a variety leaning towards Hokkien. If Lánnang-uè is analyzed as such, then my findings point to another variety of Hokkien in the Philippines alongside Manila/Philippine Hokkien (Dy 1972; Tsai 2017). However, if it is analyzed as a mixed language (or a variety leaning towards mixed languages), then my findings point to the existence of the first documented mixed language in the Philippines. They also point to the existence of a mixed language that patterns very differently from other varieties that have been labelled as mixed languages. In any case, Lánnang-uè is a novel contact language that is distinct from documented contact varieties in the Philippines:

- Cavite Chabacano (Lesho 2013) and Lannang Chabacano (Lizbeth Lim, p.c.), examples of creoles;
- Chinese Tagalog Pidgin (Gonzales 2021b) and Chinese Spanish Pidgin (Fernández 2018; Fernández and Sippola 2017), examples of pidgins, and

- Ilokano-Spanish, a variety that, at present, cannot be conclusively classified as a mixed language, a case of code-switching, or a variety with heavy borrowing (Sippola 2021:272).

7.3 *Limitations and future directions*

Like many studies, the scope of this project was limited by time, and there are still quite a few outstanding issues to explore. In fact, I hope that this dissertation raises more questions than it answers – it is, after all, the only comprehensive investigation of the variety to my knowledge. I have already listed some of them previously within the text, but I list some others here.

One path for future research is to expand the description of Lánnang-uè. I had plans to enrich my description in the early stages of my dissertation journey, but I was not able to, partly because the COVID-19 pandemic has made fieldwork and conducting sociolinguistic experiments almost impossible. What I have presented in Chapter 3 is only a snapshot of the variety's lexicon and grammar. I have not, for example, described complex constructions such as embedded *wh*-questions. I have also not done quantitative descriptions for each feature described in the sketch. For instance, in the phonetics chapter, I have only described the vowels and consonants based on corpus data and native speaker intuitions. I have not acoustically measured and described, for example, the vowel quality (e.g., mean formant values) and voice onset time of voiced stops. Furthermore, I have not systematically investigated the patterns of variation for most of the features in the descriptive sketch. Relying on corpus data, I was not able to find any conditioning factors for many patterns (e.g., the use of *shangá* 'who' vs. *siangá* 'who', the use of mid vs. low tone in tone sandhi). However, it is possible that patterns would emerge in an investigation of these features using other sources of data. A more comprehensive description of the variety informed by data sources other than the Lannang Corpus promises to be of great value to the Lánnang-uè-speaking community as well as individuals interested in language contact, multilingualism, and heritage varieties in the Philippines.

Because of time and data constraints, I was not able to quantitatively measure and test the degree of spread and stability in other features of Lánnang-uè in this dissertation (e.g., the use of Hokkien-derived pronouns). I was also unable to analyze the deviations from the potential conventions with respect to multiple factors (e.g., social, cognitive) for the same reasons. I only focused on quantitatively examining spread, stability, and variation in select prosodic, lexical,

and syntactic features (e.g., tone patterns, conjunction distribution pattern). Thus, another potential area for research is conducting similar analyses on Lánnang-uè features that I have not yet systematically investigated. Doing so will widen our current understanding of Lánnang-uè.

This dissertation answered some questions regarding the relationship between a few hypothesized social factors (e.g., sex, age, language proficiency, language attitudes) and linguistic production. However, many questions involving the relationship between other salient social variables (e.g., religion, degree of physical displacement from the historical Lannang enclave of Binondo, degree of Filipino-ness/Chinese-ness, frequency of language use, socio-economic class) and linguistic behavior remain. For instance, do these enumerated social factors correlate with patterns of variation in Lánnang-uè? Are these potential correlations consistent across all features or just a subset of features? What is the relationship between linguistic behavior and non-social (e.g., cognitive) factors? Cognitive factors such as memory and likelihood of block/ignoring variants that are “non-native” have not been formally included in my examination of linguistic variation and feature use but have been shown in cross-linguistic research to correlate with linguistic behavior (Sharma and Sankaran 2011:401; Chambers 2002; Takahesu Tabori et al. 2018). Research has also shown that cognitive factors interact with social factors to give rise to linguistic innovations (Sharma and Sankaran 2011:424). Therefore, another path for further research is to investigate my data with respect to these variables. Considering factors that have not been explored in this dissertation (e.g., cognitive, other social factors) is of paramount importance for a truly comprehensive analysis of linguistic features and variation in Lánnang-uè.

Study replication is also another potential (and crucial) research trajectory, given the relatively small sample of participants in certain analyses. In my prosody chapter, for example, I only investigated the prosody of 20 participants. While I statistically controlled for individual effects by introducing random effects in the model (Konstantopoulos and Hedges 2019) and found sociolinguistic patterns, I remain very cautious about the generalizations I make. Statistical control is no replacement for more data, so future research should test whether my sociolinguistic models apply to other sociolinguistic Lánnang-uè datasets. If the same model yields identical results in another larger set of data (ideally random-sampled), then the findings of that chapter can be generalized to the Lánnang-uè-speaking community. If not, all possible factors that may have contributed to the differences (e.g., experimental setting, methodology)

must be meticulously examined and ruled out before concluding that the findings of that chapter are unreliable or inaccurate. Regardless of the results, study replication is crucial to test the robustness of my (sociolinguistic) models and findings.

While joining some scholars in specifically looking at under-documented Asia-Pacific communities sociolinguistically (Leimgruber et al. 2020; Starr and Wang 2021; Dickson and Durantin 2019), this dissertation barely scratches the surface, and stands in large part as a call for future research.

7.4 Final remarks

Lánnang-uè looks like a hodgepodge of Hokkien, Tagalog, English, and Mandarin elements; it is also perceived by many of its speakers as such. My dissertation looked at Lánnang-uè empirically and holistically (e.g., phonologically, morpho-syntactically) and found that Lánnang-uè has many systematic, widespread, stable, and relatively independent features, providing strong evidence that the variety is highly language-like. A closer look at the features of Lánnang-uè revealed that many of these features resemble Hokkien. However, my systematic investigation also highlighted several cases where Lánnang-uè patterned differently from Hokkien. For instance, I observed the highly consistent and systematic derivation of certain linguistic elements (lexicon and structure) from Hokkien, Tagalog, or English – a feature associated with mixed languages. I also found Lánnang-uè features that resembled none of the features in the source languages. So, given that Lánnang-uè is highly language-like (a contact language, particularly) and given the (socio)linguistic patterns uncovered in this dissertation, does Lánnang-uè belong to the Sino-Tibetan family, specifically Southern Min, or does it belong to the cluster of mixed languages? An examination of the sociolinguistic data involving a pool of roughly 140 Lánnang-uè speakers in Manila provides no *unequivocal* support for Lánnang-uè as a Hokkien variety or as a mixed language. However, based on the social and linguistic evidence so far, there is a high likelihood that Lánnang-uè is mixed language or – if one adopts a non-categorical perspective – a linguistic variety situated in a continuum between ‘Hokkien’ and ‘mixed language’, leaning closer towards mixed language. Pending successful replication and further study, the pieces of evidence presented in this dissertation and in previous work converge on the idea that Lánnang-uè has a high degree of languageness, with features that set it apart from other linguistic varieties

and language types in its ecology. It is rightfully labeled Lánnang-uè – a language that its speakers can truly call their own.

APPENDICES

Appendix A. Completed Expanded Swadesh List (24-year-old Male Lannang)

#	Word	Hokkien-sourced	Tagalog-sourced	English-sourced
1	all	<i>lōng tsòng</i>		
2	and	<i>kâp</i>	<i>tsakà</i>	
3	animal			<i>animàl</i>
4	ashes		<i>abò</i>	<i>ashês</i>
5	at	<i>tī</i>		
6	back (of body)	<i>aūpiâh</i>		<i>bâck</i>
7	bad	<i>phaî</i>		<i>bâd</i>
8	bark (tree)			<i>bârk</i>
9	because		<i>kasî</i>	
10	belly	<i>pâktô</i>		<i>belly</i>
11	big	<i>tuà</i>		
12	bird	<i>tsiaû</i>		<i>bîrd</i>
13	to bite		<i>kagât</i>	
14	black	<i>ō</i>		<i>blâck</i>
15	blood	<i>huîh</i>	<i>dugô</i>	<i>bloôd</i>
16	to blow			<i>blòw</i>
17	bone	<i>kût</i>		<i>bòne</i>
18	breast		<i>dedê</i>	<i>breâst</i>
19	to breathe	<i>hōkhîp</i>		<i>breâthe</i>
20	brother	<i>âhiā</i>		
21	to burn	<i>siō</i>	<i>sunôg</i>	<i>bùrn</i>
22	child	<i>gīnâ</i>		
23	clothing	<i>sā</i>		
24	cloud			<i>clòud</i>
25	claw			<i>clàw</i>

26	cold (weather)	<i>kuá</i>		
27	to come	<i>lai</i>		
28	to cook	<i>tsí</i>	<i>lutó</i>	
29	to count	<i>sùng</i>		<i>coùnt</i>
30	to cut			<i>cút</i>
31	to dance	<i>thiaübū</i>		<i>dânce</i>
32	day (not night)			<i>daytime</i>
33	to die	<i>sí</i>	<i>patây</i>	
34	to dig	<i>iâh</i>		<i>díg</i>
35	dirty	<i>ātsām</i>		
36	dog	<i>kaū</i>		
37	to drink	<i>dīm</i>		
38	dry (substance)	<i>tā</i>	<i>tuyó</i>	<i>dry</i>
39	dull (knife)			<i>dùll</i>
40	dust	<i>thōhún</i>		<i>dúst</i>
41	ear	<i>hīyá</i>	<i>tengà</i>	
42	earth (soil)	<i>thó</i>		
43	to eat	<i>tshiáh</i>		
44	egg			<i>égg</i>
45	eight	<i>poèh</i>		
46	eye	<i>bāktsiū</i>		<i>èye</i>
47	to fall	<i>lâk/sâk</i>	<i>hulôg</i>	
48	far	<i>hūng</i>		
49	fat (grease)	<i>puí</i>		
50	father	<i>pā pá</i>		
51	to fear	<i>kiā</i>		
52	feather			<i>feathèr</i>
53	few	<i>tāmpóh</i>		
54	to fight	<i>ūankē</i>		<i>fight</i>
55	fire	<i>hê</i>		<i>fire</i>
56	fish	<i>hí</i>		<i>fish</i>
57	five	<i>gō</i>		
58	to float			<i>floát</i>
59	to flow			<i>flòw</i>
60	flower	<i>huē</i>		<i>flowèr</i>
61	to fly	<i>pē</i>		
62	fog			<i>fôg</i>
63	foot	<i>khā</i>		
64	four	<i>sì</i>		

65	fruit	<i>kētsī</i>		<i>frùit</i>
66	full		<i>punô</i>	<i>fùll</i>
67	to give	<i>hō</i>		
68	good	<i>hō</i>		
69	grass	<i>tshaû</i>		<i>grâss</i>
70	green			<i>greèn</i>
71	guts		<i>bitukâ</i>	<i>intestìne</i>
72	hair	<i>thaūmúng</i>		
73	hand	<i>tshiû</i>		
74	he	<i>î</i>		
75	head	<i>thau</i>		
76	to hear	<i>thiāh</i>		
77	heart	<i>sīm</i>		<i>heârt</i>
78	heavy	<i>tāng</i>	<i>bigât</i>	
79	here	<i>tsiá</i>		
80	to hit	<i>phāh</i>		
81	to hold (in hand)		<i>hawâk</i>	
82	horn (of an animal)			<i>hòrn</i>
83	how	<i>tsiuā</i>		
84	hundred	<i>bāh</i>		<i>hundrêd</i>
85	to hunt			<i>hùnt</i>
86	husband		<i>asawà</i>	
87	I	<i>goā</i>		
88	ice			<i>íce</i>
89	if	<i>nā</i>	<i>pâg</i>	
90	inside	<i>laībín</i>		
91	to kill	<i>phāhsi</i>		
92	knee			<i>kneè</i>
93	to know (facts)	<i>tsā iā</i>		
94	lake			<i>lâke</i>
95	to laugh	<i>tshiò</i>		
96	leaf			<i>leáf</i>
97	left (hand)	<i>tō</i>		<i>lêft</i>
98	leg			<i>lêg</i>
99	to lie down	<i>tō</i>	<i>higâ</i>	
100	to live (be alive)	<i>úah</i>		
101	liver			<i>livèr</i>

102	long	<i>túng</i>		
103	louse		<i>kutò</i>	
104	man (male)	<i>lāmè</i>		
105	many	<i>tsuè</i>		
106	meat (flesh)	<i>mâh</i>		
107	moon			<i>moòn</i>
108	mother	<i>māmá</i>		
109	mountain			<i>mountain</i>
110	mouth	<i>tshuì</i>		
111	name	<i>miá</i>		
112	narrow			<i>narròw</i>
113	near	<i>kūn</i>		
114	neck	<i>āmkuī</i>		<i>nêck</i>
115	new	<i>sīn</i>		
116	night	<i>âmi</i>		
117	nine	<i>kaû</i>		
118	nose	<i>phì</i>		<i>nôse</i>
119	not	<i>bō</i>		
120	old	<i>kù/laū</i>		
121	one	<i>tsít</i>		
122	other	<i>pādé</i>		
123	person	<i>láng</i>		
124	to play	<i>thīthó</i>		
125	to pull		<i>haták</i>	<i>pùll</i>
126	to push		<i>tulák</i>	<i>pûsh</i>
127	rain	<i>hō</i>		
128	red	<i>áng</i>		<i>rêd</i>
129	right (correct)	<i>tióh</i>		
130	right (hand)			<i>ríght</i>
131	river			<i>rivèr</i>
132	road	<i>lò</i>		
133	root		<i>ugât</i>	<i>roôt</i>
134	rope			<i>rôpe</i>
135	rotten (log)		<i>bulôk</i>	
136	to run	<i>tsaû</i>		
137	salt	<i>íam</i>		<i>sâlt</i>
138	sand	<i>soā</i>		<i>sând</i>
139	to say	<i>kōng</i>		
140	to scrach		<i>kamôt</i>	<i>scrâch</i>

141	sea (ocean)	<i>haī</i>		
142	to see	<i>khuà</i>		
143	seed			<i>seéd</i>
144	seven	<i>tshīt</i>		
145	to sew		<i>tahī</i>	
146	sharp (knife)	<i>tsiām</i>		<i>shârp</i>
147	to shoot			<i>shoôt</i>
148	short	<i>tē</i>		
149	to sing	<i>tshiū</i>		
150	sister	<i>atsī</i>		
151	to sit	<i>tsē</i>		
152	six	<i>lák</i>		
153	skin (of person)	<i>phé</i>	<i>balât</i>	
154	sky	<i>thī</i>		<i>skỳ</i>
155	to sleep	<i>khùn</i>		
156	small	<i>suè</i>		
157	to smell (perceive odor)	<i>phī</i>		
158	smoke			<i>smôke</i>
159	smooth		<i>kinis</i>	<i>smoôth</i>
160	snake		<i>ahâs</i>	<i>snâke</i>
161	snow			<i>snòw</i>
162	some	<i>ū-é</i>	<i>mgà</i>	
163	to spit		<i>durâ</i>	<i>spīt</i>
164	to split		<i>hatī</i>	<i>splīt</i>
165	to squeeze			<i>squeêze</i>
166	to stab (pierce)		<i>tusôk</i>	<i>stâb</i>
167	to stand	<i>têkhiā</i>		
168	star			<i>stâr</i>
169	stick (of wood)			<i>stīck</i>
170	stone			<i>stòne</i>
171	straight	<i>tít</i>		<i>straīght</i>
172	to suck	<i>súh</i>		<i>súck</i>
173	sun	<i>dītthau</i>		<i>sùn</i>
174	to swell	<i>tsiêng</i>	<i>magâ</i>	
175	to swim	<i>shūtsuī</i>		
176	tail			<i>tail</i>
177	ten	<i>tsáp</i>		
178	there	<i>hiá</i>		

179	that	<i>hîgé</i>		
180	they	<i>īn</i>		
181	thick	<i>kaū</i>	<i>kapâl</i>	
182	thin	<i>sān</i>		
183	to think	<i>siū</i>		
184	this	<i>tsîgé</i>		
185	thou (you, sg.)	<i>dī</i>		
186	three	<i>sā</i>		
187	to throw	<i>tānsāk</i>		
188	to tie	<i>pāk</i>	<i>talī</i>	
189	tongue		<i>dilā</i>	<i>tonguè</i>
190	tooth	<i>tshuīkhī</i>		
191	tree	<i>tshiu</i>		<i>tree</i>
192	to turn		<i>ikôt</i>	<i>tùrn</i>
193	twenty	<i>dītsáp</i>	<i>bentè</i>	
194	two	<i>nūng</i>		
195	to vomit	<i>thò</i>		
196	to walk	<i>kiá</i>		
197	warm (weather)	<i>duáh</i>	<i>inīt</i>	
198	to wash	<i>suē</i>	<i>labà/hugás</i>	
199	water	<i>tsuí</i>		
200	we	<i>gūn/dân</i>		
201	wet		<i>basá</i>	
202	what?	<i>siāmmih/siāmih</i>		
203	when?	<i>tīsí</i>		
204	where?	<i>tōlôh</i>		
205	white	<i>pēh</i>		
206	who?	<i>siāngá</i>		
207	wide			<i>wīde</i>
208	wife	<i>bô/khāntshiu</i>	<i>asawà</i>	
209	wind (breeze)	<i>huāng</i>		
210	wing		<i>pakpāk</i>	<i>wīngs</i>
211	to wipe	<i>tshīt</i>	<i>punás</i>	<i>wīpe</i>
212	with (accompanying)	<i>kâp</i>		
213	woman	<i>dūwè</i>		
214	wood		<i>kahôy</i>	
215	to work	<i>tsuêkāng</i>		
216	worm		<i>uôd</i>	<i>wòrm</i>

217	ye (you, plu.)	<i>dîn</i>		
218	year	<i>ní</i>		
219	yellow	<i>úng</i>		<i>yellòw</i>

Appendix B. Means, Standard Deviations and Standard Errors of Sixteen Lannangs' Likelihood of Using Lánnang-uè Word by Word Origin (1 = very likely, 0 = unlikely), Based on a 219-word Expanded Swadesh List.

Data was collected online in 2020. Shaded cells are those above the average score of 0.5.

#	gloss	Hokkien			Tagalog			English		
		Mean	SD	SE	Mean	SD	SE	Mean	SD	SE
1	all	1.00	0.00	0.00	0.38	0.50	0.13	0.13	0.34	0.09
2	and	1.00	0.00	0.00	0.50	0.52	0.13	0.19	0.40	0.10
3	animal	0.56	0.51	0.13	0.31	0.48	0.12	0.69	0.48	0.12
4	ashes	0.19	0.40	0.10	0.31	0.48	0.12	0.81	0.40	0.10
5	at	0.94	0.25	0.06	0.38	0.50	0.13	0.13	0.34	0.09
6	back (of body)	0.81	0.40	0.10	0.44	0.51	0.13	0.31	0.48	0.12
7	bad	0.94	0.25	0.06	0.31	0.48	0.12	0.38	0.50	0.13
8	bark (tree)	0.13	0.34	0.09	0.06	0.25	0.06	0.88	0.34	0.09
9	because	0.75	0.45	0.11	0.81	0.40	0.10	0.25	0.45	0.11
10	belly	0.94	0.25	0.06	0.31	0.48	0.12	0.38	0.50	0.13
11	big	1.00	0.00	0.00	0.38	0.50	0.13	0.25	0.45	0.11
12	bird	0.63	0.50	0.13	0.44	0.51	0.13	0.56	0.51	0.13
13	to bite	0.69	0.48	0.12	0.56	0.51	0.13	0.25	0.45	0.11
14	black	1.00	0.00	0.00	0.31	0.48	0.12	0.38	0.50	0.13
15	blood	0.69	0.48	0.12	0.44	0.51	0.13	0.38	0.50	0.13
16	to blow	0.44	0.51	0.13	0.06	0.25	0.06	0.75	0.45	0.11
17	bone	0.69	0.48	0.12	0.50	0.52	0.13	0.38	0.50	0.13
18	breast	0.31	0.48	0.12	0.19	0.40	0.10	0.75	0.45	0.11
19	to breathe	0.63	0.50	0.13	0.38	0.50	0.13	0.44	0.51	0.13
20	brother	1.00	0.00	0.00	0.19	0.40	0.10	0.31	0.48	0.12
21	to burn	0.75	0.45	0.11	0.44	0.51	0.13	0.50	0.52	0.13
22	child	1.00	0.00	0.00	0.31	0.48	0.12	0.06	0.25	0.06
23	clothing	0.88	0.34	0.09	0.50	0.52	0.13	0.19	0.40	0.10
24	cloud	0.19	0.40	0.10	0.19	0.40	0.10	0.94	0.25	0.06
25	claw	0.13	0.34	0.09	0.06	0.25	0.06	0.88	0.34	0.09
26	cold (weather)	0.88	0.34	0.09	0.31	0.48	0.12	0.19	0.40	0.10

27	to come	1.00	0.00	0.00	0.25	0.45	0.11	0.13	0.34	0.09
28	to cook	0.88	0.34	0.09	0.56	0.51	0.13	0.13	0.34	0.09
29	to count	0.81	0.40	0.10	0.19	0.40	0.10	0.56	0.51	0.13
30	to cut	0.81	0.40	0.10	0.13	0.34	0.09	0.50	0.52	0.13
31	to dance	0.81	0.40	0.10	0.38	0.50	0.13	0.31	0.48	0.12
32	day (not night)	0.75	0.45	0.11	0.13	0.34	0.09	0.19	0.40	0.10
33	to die	1.00	0.00	0.00	0.31	0.48	0.12	0.13	0.34	0.09
34	to dig	0.38	0.50	0.13	0.25	0.45	0.11	0.69	0.48	0.12
35	dirty	1.00	0.00	0.00	0.31	0.48	0.12	0.38	0.50	0.13
36	dog	1.00	0.00	0.00	0.25	0.45	0.11	0.38	0.50	0.13
37	to drink	0.94	0.25	0.06	0.25	0.45	0.11	0.13	0.34	0.09
38	dry (substance)	0.75	0.45	0.11	0.38	0.50	0.13	0.56	0.51	0.13
39	dull (knife)	0.38	0.50	0.13	0.44	0.51	0.13	0.50	0.52	0.13
40	dust	0.56	0.51	0.13	0.44	0.51	0.13	0.38	0.50	0.13
41	ear	0.88	0.34	0.09	0.31	0.48	0.12	0.19	0.40	0.10
42	earth (soil)	0.44	0.51	0.13	0.31	0.48	0.12	0.50	0.52	0.13
43	to eat	1.00	0.00	0.00	0.25	0.45	0.11	0.13	0.34	0.09
44	egg	0.69	0.48	0.12	0.44	0.51	0.13	0.50	0.52	0.13
45	eight	1.00	0.00	0.00	0.25	0.45	0.11	0.25	0.45	0.11
46	eye	0.88	0.34	0.09	0.38	0.50	0.13	0.38	0.50	0.13
47	to fall	0.81	0.40	0.10	0.50	0.52	0.13	0.13	0.34	0.09
48	far	0.81	0.40	0.10	0.38	0.50	0.13	0.19	0.40	0.10
49	fat (grease)	0.69	0.48	0.12	0.44	0.51	0.13	0.31	0.48	0.12
50	father	1.00	0.00	0.00	0.31	0.48	0.12	0.25	0.45	0.11
51	to fear	0.88	0.34	0.09	0.38	0.50	0.13	0.13	0.34	0.09
52	feather	0.13	0.34	0.09	0.06	0.25	0.06	1.00	0.00	0.00
53	few	0.88	0.34	0.09	0.63	0.50	0.13	0.13	0.34	0.09
54	to fight	0.81	0.40	0.10	0.44	0.51	0.13	0.38	0.50	0.13
55	fire	0.63	0.50	0.13	0.19	0.40	0.10	0.56	0.51	0.13
56	fish	0.94	0.25	0.06	0.25	0.45	0.11	0.38	0.50	0.13
57	five	1.00	0.00	0.00	0.25	0.45	0.11	0.25	0.45	0.11
58	to float	0.44	0.51	0.13	0.25	0.45	0.11	0.75	0.45	0.11
59	to flow	0.31	0.48	0.12	0.06	0.25	0.06	0.69	0.48	0.12
60	flower	0.75	0.45	0.11	0.25	0.45	0.11	0.56	0.51	0.13
61	to fly	0.75	0.45	0.11	0.44	0.51	0.13	0.25	0.45	0.11
62	fog	0.31	0.48	0.12	0.13	0.34	0.09	1.00	0.00	0.00
63	foot	1.00	0.00	0.00	0.25	0.45	0.11	0.19	0.40	0.10
64	four	1.00	0.00	0.00	0.31	0.48	0.12	0.25	0.45	0.11
65	fruit	0.69	0.48	0.12	0.38	0.50	0.13	0.50	0.52	0.13

66	full	0.56	0.51	0.13	0.63	0.50	0.13	0.25	0.45	0.11
67	to give	0.94	0.25	0.06	0.31	0.48	0.12	0.06	0.25	0.06
68	good	0.94	0.25	0.06	0.19	0.40	0.10	0.13	0.34	0.09
69	grass	0.50	0.52	0.13	0.13	0.34	0.09	0.75	0.45	0.11
70	green	0.50	0.52	0.13	0.00	0.00	0.00	0.81	0.40	0.10
71	guts	0.31	0.48	0.12	0.19	0.40	0.10	0.69	0.48	0.12
72	hair	0.88	0.34	0.09	0.38	0.50	0.13	0.19	0.40	0.10
73	hand	1.00	0.00	0.00	0.25	0.45	0.11	0.13	0.34	0.09
74	he	0.94	0.25	0.06	0.13	0.34	0.09	0.13	0.34	0.09
75	head	0.94	0.25	0.06	0.25	0.45	0.11	0.19	0.40	0.10
76	to hear	1.00	0.00	0.00	0.13	0.34	0.09	0.13	0.34	0.09
77	heart	0.75	0.45	0.11	0.13	0.34	0.09	0.50	0.52	0.13
78	heavy	0.88	0.34	0.09	0.50	0.52	0.13	0.19	0.40	0.10
79	here	1.00	0.00	0.00	0.25	0.45	0.11	0.19	0.40	0.10
80	to hit	0.81	0.40	0.10	0.19	0.40	0.10	0.25	0.45	0.11
81	to hold (in hand)	0.81	0.40	0.10	0.38	0.50	0.13	0.13	0.34	0.09
82	horn (of an animal)	0.13	0.34	0.09	0.19	0.40	0.10	0.75	0.45	0.11
83	how	1.00	0.00	0.00	0.38	0.50	0.13	0.06	0.25	0.06
84	hundred	1.00	0.00	0.00	0.00	0.00	0.00	0.31	0.48	0.12
85	to hunt	0.25	0.45	0.11	0.00	0.00	0.00	0.81	0.40	0.10
86	husband	0.50	0.52	0.13	0.38	0.50	0.13	0.56	0.51	0.13
87	I	1.00	0.00	0.00	0.19	0.40	0.10	0.13	0.34	0.09
88	ice	0.44	0.51	0.13	0.19	0.40	0.10	0.81	0.40	0.10
89	if	0.56	0.51	0.13	0.75	0.45	0.11	0.38	0.50	0.13
90	inside	0.94	0.25	0.06	0.38	0.50	0.13	0.19	0.40	0.10
91	to kill	0.88	0.34	0.09	0.31	0.48	0.12	0.13	0.34	0.09
92	knee	0.13	0.34	0.09	0.44	0.51	0.13	0.69	0.48	0.12
93	to know (facts)	0.94	0.25	0.06	0.19	0.40	0.10	0.13	0.34	0.09
94	lake	0.13	0.34	0.09	0.06	0.25	0.06	0.94	0.25	0.06
95	to laugh	0.81	0.40	0.10	0.19	0.40	0.10	0.06	0.25	0.06
96	leaf	0.25	0.45	0.11	0.38	0.50	0.13	0.63	0.50	0.13
97	left (hand)	0.81	0.40	0.10	0.19	0.40	0.10	0.56	0.51	0.13
98	leg	0.50	0.52	0.13	0.06	0.25	0.06	0.69	0.48	0.12
99	to lie down	0.81	0.40	0.10	0.50	0.52	0.13	0.06	0.25	0.06
100	to live (be alive)	0.69	0.48	0.12	0.44	0.51	0.13	0.13	0.34	0.09
101	liver	0.38	0.50	0.13	0.06	0.25	0.06	0.75	0.45	0.11
102	long	0.81	0.40	0.10	0.38	0.50	0.13	0.13	0.34	0.09
103	louse	0.13	0.34	0.09	0.50	0.52	0.13	0.50	0.52	0.13
104	man (male)	1.00	0.00	0.00	0.25	0.45	0.11	0.13	0.34	0.09

105	many	0.94	0.25	0.06	0.31	0.48	0.12	0.19	0.40	0.10
106	meat (flesh)	0.88	0.34	0.09	0.00	0.00	0.00	0.19	0.40	0.10
107	moon	0.44	0.51	0.13	0.13	0.34	0.09	0.75	0.45	0.11
108	mother	1.00	0.00	0.00	0.25	0.45	0.11	0.25	0.45	0.11
109	mountain	0.56	0.51	0.13	0.31	0.48	0.12	0.63	0.50	0.13
110	mouth	0.88	0.34	0.09	0.31	0.48	0.12	0.25	0.45	0.11
111	name	1.00	0.00	0.00	0.19	0.40	0.10	0.31	0.48	0.12
112	narrow	0.50	0.52	0.13	0.19	0.40	0.10	0.69	0.48	0.12
113	near	0.75	0.45	0.11	0.38	0.50	0.13	0.25	0.45	0.11
114	neck	0.50	0.52	0.13	0.25	0.45	0.11	0.56	0.51	0.13
115	new	0.94	0.25	0.06	0.38	0.50	0.13	0.19	0.40	0.10
116	night	0.88	0.34	0.09	0.31	0.48	0.12	0.13	0.34	0.09
117	nine	1.00	0.00	0.00	0.13	0.34	0.09	0.19	0.40	0.10
118	nose	0.88	0.34	0.09	0.25	0.45	0.11	0.50	0.52	0.13
119	not	1.00	0.00	0.00	0.25	0.45	0.11	0.19	0.40	0.10
120	old	0.81	0.40	0.10	0.38	0.50	0.13	0.25	0.45	0.11
121	one	1.00	0.00	0.00	0.25	0.45	0.11	0.25	0.45	0.11
122	other	0.94	0.25	0.06	0.31	0.48	0.12	0.13	0.34	0.09
123	person	1.00	0.00	0.00	0.19	0.40	0.10	0.19	0.40	0.10
124	to play	0.88	0.34	0.09	0.31	0.48	0.12	0.19	0.40	0.10
125	to pull	0.75	0.45	0.11	0.44	0.51	0.13	0.38	0.50	0.13
126	to push	0.75	0.45	0.11	0.44	0.51	0.13	0.44	0.51	0.13
127	rain	0.94	0.25	0.06	0.38	0.50	0.13	0.31	0.48	0.12
128	red	1.00	0.00	0.00	0.19	0.40	0.10	0.31	0.48	0.12
129	right (correct)	0.94	0.25	0.06	0.31	0.48	0.12	0.25	0.45	0.11
130	right (hand)	0.75	0.45	0.11	0.31	0.48	0.12	0.44	0.51	0.13
131	river	0.25	0.45	0.11	0.13	0.34	0.09	0.88	0.34	0.09
132	road	0.63	0.50	0.13	0.31	0.48	0.12	0.38	0.50	0.13
133	root	0.13	0.34	0.09	0.31	0.48	0.12	0.69	0.48	0.12
134	rope	0.44	0.51	0.13	0.31	0.48	0.12	0.63	0.50	0.13
135	rotten (log)	0.25	0.45	0.11	0.50	0.52	0.13	0.44	0.51	0.13
136	to run	0.81	0.40	0.10	0.31	0.48	0.12	0.19	0.40	0.10
137	salt	0.75	0.45	0.11	0.31	0.48	0.12	0.31	0.48	0.12
138	sand	0.63	0.50	0.13	0.13	0.34	0.09	0.50	0.52	0.13
139	to say	1.00	0.00	0.00	0.25	0.45	0.11	0.13	0.34	0.09
140	to scrach	0.50	0.52	0.13	0.50	0.52	0.13	0.56	0.51	0.13
141	sea (ocean)	0.56	0.51	0.13	0.31	0.48	0.12	0.38	0.50	0.13
142	to see	1.00	0.00	0.00	0.19	0.40	0.10	0.19	0.40	0.10
143	seed	0.13	0.34	0.09	0.19	0.40	0.10	0.81	0.40	0.10

144	seven	1.00	0.00	0.00	0.19	0.40	0.10	0.25	0.45	0.11
145	to sew	0.44	0.51	0.13	0.63	0.50	0.13	0.19	0.40	0.10
146	sharp (knife)	0.63	0.50	0.13	0.31	0.48	0.12	0.50	0.52	0.13
147	to shoot	0.31	0.48	0.12	0.13	0.34	0.09	0.63	0.50	0.13
148	short	0.81	0.40	0.10	0.25	0.45	0.11	0.38	0.50	0.13
149	to sing	0.88	0.34	0.09	0.25	0.45	0.11	0.19	0.40	0.10
150	sister	1.00	0.00	0.00	0.25	0.45	0.11	0.25	0.45	0.11
151	to sit	0.88	0.34	0.09	0.25	0.45	0.11	0.13	0.34	0.09
152	six	1.00	0.00	0.00	0.19	0.40	0.10	0.19	0.40	0.10
153	skin (of person)	0.75	0.45	0.11	0.31	0.48	0.12	0.38	0.50	0.13
154	sky	0.63	0.50	0.13	0.25	0.45	0.11	0.63	0.50	0.13
155	to sleep	1.00	0.00	0.00	0.25	0.45	0.11	0.19	0.40	0.10
156	small	1.00	0.00	0.00	0.25	0.45	0.11	0.19	0.40	0.10
157	to smell (perceive odor)	0.75	0.45	0.11	0.31	0.48	0.12	0.19	0.40	0.10
158	smoke	0.44	0.51	0.13	0.25	0.45	0.11	0.56	0.51	0.13
159	smooth	0.19	0.40	0.10	0.25	0.45	0.11	0.88	0.34	0.09
160	snake	0.50	0.52	0.13	0.38	0.50	0.13	0.69	0.48	0.12
161	snow	0.44	0.51	0.13	0.00	0.00	0.00	0.94	0.25	0.06
162	some	0.88	0.34	0.09	0.25	0.45	0.11	0.38	0.50	0.13
163	to spit	0.56	0.51	0.13	0.31	0.48	0.12	0.44	0.51	0.13
164	to split	0.25	0.45	0.11	0.50	0.52	0.13	0.50	0.52	0.13
165	to squeeze	0.19	0.40	0.10	0.25	0.45	0.11	0.75	0.45	0.11
166	to stab (pierce)	0.44	0.51	0.13	0.56	0.51	0.13	0.38	0.50	0.13
167	to stand	0.88	0.34	0.09	0.19	0.40	0.10	0.13	0.34	0.09
168	star	0.19	0.40	0.10	0.00	0.00	0.00	0.94	0.25	0.06
169	stick (of wood)	0.06	0.25	0.06	0.00	0.00	0.00	1.00	0.00	0.00
170	stone	0.56	0.51	0.13	0.44	0.51	0.13	0.38	0.50	0.13
171	straight	0.56	0.51	0.13	0.31	0.48	0.12	0.56	0.51	0.13
172	to suck	0.56	0.51	0.13	0.13	0.34	0.09	0.44	0.51	0.13
173	sun	0.56	0.51	0.13	0.31	0.48	0.12	0.63	0.50	0.13
174	to swell	0.38	0.50	0.13	0.50	0.52	0.13	0.44	0.51	0.13
175	to swim	0.81	0.40	0.10	0.19	0.40	0.10	0.31	0.48	0.12
176	tail	0.31	0.48	0.12	0.44	0.51	0.13	0.69	0.48	0.12
177	ten	1.00	0.00	0.00	0.19	0.40	0.10	0.19	0.40	0.10
178	there	0.94	0.25	0.06	0.25	0.45	0.11	0.13	0.34	0.09
179	that	1.00	0.00	0.00	0.19	0.40	0.10	0.06	0.25	0.06
180	they	0.94	0.25	0.06	0.19	0.40	0.10	0.06	0.25	0.06
181	thick	0.81	0.40	0.10	0.44	0.51	0.13	0.31	0.48	0.12

182	thin	0.75	0.45	0.11	0.31	0.48	0.12	0.19	0.40	0.10
183	to think	0.94	0.25	0.06	0.25	0.45	0.11	0.13	0.34	0.09
184	this	1.00	0.00	0.00	0.13	0.34	0.09	0.13	0.34	0.09
185	thou (you, sg.)	1.00	0.00	0.00	0.06	0.25	0.06	0.06	0.25	0.06
186	three	1.00	0.00	0.00	0.19	0.40	0.10	0.19	0.40	0.10
187	to throw	0.75	0.45	0.11	0.25	0.45	0.11	0.19	0.40	0.10
188	to tie	0.56	0.51	0.13	0.38	0.50	0.13	0.25	0.45	0.11
189	tongue	0.63	0.50	0.13	0.44	0.51	0.13	0.50	0.52	0.13
190	tooth	0.69	0.48	0.12	0.25	0.45	0.11	0.25	0.45	0.11
191	tree	0.69	0.48	0.12	0.38	0.50	0.13	0.44	0.51	0.13
192	to turn	0.44	0.51	0.13	0.31	0.48	0.12	0.50	0.52	0.13
193	twenty	1.00	0.00	0.00	0.19	0.40	0.10	0.25	0.45	0.11
194	two	1.00	0.00	0.00	0.25	0.45	0.11	0.25	0.45	0.11
195	to vomit	0.69	0.48	0.12	0.44	0.51	0.13	0.13	0.34	0.09
196	to walk	0.88	0.34	0.09	0.31	0.48	0.12	0.13	0.34	0.09
197	warm (weather)	0.88	0.34	0.09	0.31	0.48	0.12	0.38	0.50	0.13
198	to wash	0.94	0.25	0.06	0.31	0.48	0.12	0.25	0.45	0.11
199	water	1.00	0.00	0.00	0.25	0.45	0.11	0.31	0.48	0.12
200	we	1.00	0.00	0.00	0.13	0.34	0.09	0.13	0.34	0.09
201	wet	0.75	0.45	0.11	0.44	0.51	0.13	0.19	0.40	0.10
202	what?	1.00	0.00	0.00	0.19	0.40	0.10	0.13	0.34	0.09
203	when?	0.75	0.45	0.11	0.31	0.48	0.12	0.13	0.34	0.09
204	where?	1.00	0.00	0.00	0.13	0.34	0.09	0.13	0.34	0.09
205	white	1.00	0.00	0.00	0.13	0.34	0.09	0.13	0.34	0.09
206	who?	1.00	0.00	0.00	0.19	0.40	0.10	0.06	0.25	0.06
207	wide	0.38	0.50	0.13	0.25	0.45	0.11	0.63	0.50	0.13
208	wife	0.75	0.45	0.11	0.31	0.48	0.12	0.38	0.50	0.13
209	wind (breeze)	0.69	0.48	0.12	0.25	0.45	0.11	0.38	0.50	0.13
210	wing	0.25	0.45	0.11	0.56	0.51	0.13	0.69	0.48	0.12
211	to wipe	0.75	0.45	0.11	0.38	0.50	0.13	0.19	0.40	0.10
212	with (accompanying)	1.00	0.00	0.00	0.25	0.45	0.11	0.19	0.40	0.10
213	woman	1.00	0.00	0.00	0.25	0.45	0.11	0.13	0.34	0.09
214	wood	0.56	0.51	0.13	0.25	0.45	0.11	0.31	0.48	0.12
215	to work	0.88	0.34	0.09	0.25	0.45	0.11	0.25	0.45	0.11
216	worm	0.31	0.48	0.12	0.50	0.52	0.13	0.69	0.48	0.12
217	ye (you, plu.)	1.00	0.00	0.00	0.06	0.25	0.06	0.13	0.34	0.09
218	year	0.94	0.25	0.06	0.13	0.34	0.09	0.38	0.50	0.13
219	yellow	0.88	0.34	0.09	0.06	0.25	0.06	0.63	0.50	0.13

Appendix C. Sample Wordlist for Stress and Tone Production Experiment

stimno	word
1	shiqi
2	siopao
3	halo
4	Beijing
5	kuko
6	sisig
7	hotdog
8	bato
9	pantoh
10	giraffe
11	xuexiao
12	suka
13	toufa
14	massage
15	lechon
16	duster
17	plateau
18	shiba
19	depot
20	sandal
21	hotel
22	pancake
23	biko
24	taho
25	bangus
26	balloon
27	baseball
28	siomai
29	laser
30	buto

31	gata
32	salon
33	tauge
34	ditsap
35	tshiathau
36	siensi
37	atay
38	buko
39	puto
40	isaw
41	bouquet
42	suman
43	debut
44	cookie
45	ballpen
46	yanjing
47	toga
48	bazaar
49	taxi
50	salad
51	laubu
52	gumah
53	lugaw
54	dessert
55	turon
56	taxi
57	massage
58	buko
59	pantoh
60	shiba
61	tshiathau

62	ballpen
63	isaw
64	bato
65	Beijing
66	laubu
67	plateau
68	atay
69	sisig
70	suka
71	baseball
72	cookie
73	bouquet
74	turon
75	kuko
76	siensi
77	siopao
78	pancake
79	balloon
80	dessert
81	gumah
82	salon
83	salad
84	hotdog
85	halo
86	duster
87	gata
88	toga
89	bangus
90	lechon
91	tauge
92	yanjing

93	buto
94	shisan
95	laser
96	depot
97	siomai
98	debut
99	giraffe
100	taho
101	shiqi
102	toufa
103	sandal
104	hotel
105	biko
106	suman
107	puto
108	ditsap
109	lugaw
110	bazaar
111	yanjing
112	buko
113	atay
114	laser
115	suka
116	salon
117	siensi
118	bouquet
119	Beijing
120	taho
121	baseball
122	sandal
123	bazaar

124	lugaw
125	lechon
126	laubu
127	duster
128	salad
129	shiba
130	cookie
131	bangus
132	plateau
133	shisan
134	biko

135	gata
136	ballpen
137	giraffe
138	gumah
139	puto
140	siomai
141	massage
142	buto
143	toufa
144	dessert
145	turon

146	hotel
147	depot
148	pancake
149	ditsap
150	tshiathau
151	bato
152	shiqi
153	toga
154	halo
155	kuko
156	hotdog

157	pantoh
158	balloon
159	sisig
160	taxi
161	siopao
162	isaw
163	debut
164	tauge
165	suman

Appendix D. Interview Questions

Community

- Do you still feel racial discrimination from non-Chinese Philippine society? How about in the past?
- What is the place of the Lannangs/ Chinese Filipinos/ Filipino-Chinese in Philippine society now as opposed to the past? Do you think our community is at threat of being dissolved or not? Why?
- What are your thoughts on the new mainland immigrants in general? Do you consider them as part of the Lannangs /Chinese Filipino/Filipino-Chinese community? What makes us different from them?
- How do you feel when media or people conflate them with Lannangs /Chinese Filipinos/Filipino-Chinese?

Identity

- How do you feel when people call you Filipino?
- How do you feel when people call you Chinese?
- For you, what does it mean to be Lannang? Chinese Filipino? Filipino-Chinese? Instead of being just Filipino or Chinese?
- Is there a difference between being Filipino-Chinese and Chinese Filipino? What is the difference for you?
- Would you consider yourself an overseas Chinese or Huakiaú? Why or why not?

Language

- If you had a child, what language would you use to speak to him or her? Why?
- What are your thoughts on the unmixed Philippine Hokkien?
- How about the mixing Hokkien?
- For you, what is the status of Hokkien use in the Manila? Why do you think such is the case?
- What are your thoughts on Hybrid Hokkien as a distinct language from Hokkien and Philippine Hokkien?
- Do you think Hybrid Hokkien has a right way of mixing? Or is it random?
- Is Hokkien important to our community? Why?
- Is Hybrid Hokkien important to our community? Why?
- Is Mandarin important to our community? Do you think it should replace Hokkien as the lingua franca? Why or why not?
- When do you use unmixed Philippine Hokkien?
- When do you use the hybrid variety of Hokkien?

- What are your thoughts on documenting the mixed language or variety that we speak?
- What are your thoughts on new immigrants learning Hybrid Hokkien?
- For younger people: Why do you not use Hybrid Hokkien with peers
- How do you feel about the younger generation not using Hokkien anymore? What can you do about it?
- Personally, do you think we should maintain Hokkien, or do you think shifting to Tagalog and English is better?
- Do you think Hybrid Hokkien is “Conyo”? Funny? Why?

Education

- Should Hokkien be taught in schools?
- Should Hybrid Hokkien be taught in schools? Used in other domains than home?
- The younger generation do not know how to write Hokkien characters anymore, but they still understand and speak Hybrid Hokkien, what are your thoughts on romanizing the language for access and maintenance?
- What do you feel about standardizing the mixed variety of Hokkien, this new language? How about transmission?
- What do you think about materials written in PHH?

Appendix E. Supplementary Linear Regression Results - Stress Perception

Dependent variable = perceived lexical stress (observations = 4,759, conditional $R^2 = 0.354$, random intercepts for participant and item). Sociolinguistic factors are based on speaker information rather than listener information

Predictors	Estimates	SE	CI	<i>p</i>
(Intercept)	0.5962	0.4213	-0.2295 – 1.4220	0.157
duration	3.274	0.2705	2.7438 – 3.8043	< 0.001
age	-0.0043	0.0107	-0.0253 – 0.0167	0.686
sex (male vs. female)	0.0319	0.1505	-0.2631 – 0.3270	0.832
proficiency (languages without stress)	-0.0907	0.2325	-0.5463 – 0.3650	0.697
proficiency (languages with stress)	0.0767	0.093	-0.1055 – 0.2589	0.409
F0 (mean)	-0.0009	0.0033	-0.0073 – 0.0055	0.781
intensity (mean)	-0.0032	0.0057	-0.0144 – 0.0080	0.576
vowel quality	-0.0369	0.0222	-0.0804 – 0.0066	0.096
F0 (initial)	-0.0048	0.0033	-0.0113 – 0.0018	0.151
F0 (slope)	-0.0086	0.0141	-0.0363 – 0.0190	0.541
duration * age	0.0136	0.0071	-0.0003 – 0.0274	0.055
duration * sex	-0.1248	0.1028	-0.3263 – 0.0767	0.225
duration * proficiency (languages without stress)	-0.3115	0.1609	-0.6269 – 0.0038	0.053
duration * proficiency (languages with stress)	-0.0741	0.0725	-0.2162 – 0.0679	0.307
age * F0 (mean)	0.0000	0.0001	-0.0002 – 0.0002	0.982
sex * F0 (mean)	-0.0003	0.0015	-0.0032 – 0.0027	0.866
proficiency (languages without stress) * F0 (mean)	0.0018	0.002	-0.0021 – 0.0057	0.374
proficiency (languages with stress) * F0 (mean)	-0.0006	0.0009	-0.0024 – 0.0012	0.493
age * intensity (mean)	-0.0001	0.0001	-0.0003 – 0.0002	0.668
sex * intensity (mean)	0.0000	0.0021	-0.0041 – 0.0041	0.993
proficiency (languages without stress) * intensity (mean)	0.0018	0.0032	-0.0045 – 0.0080	0.583

proficiency (languages with stress) * intensity (mean)	-0.0009	0.0011	-0.0031 – 0.0014	0.454
age * vowel quality	-0.0005	0.0005	-0.0016 – 0.0006	0.348
sex * vowel quality	0.0054	0.0075	-0.0094 – 0.0202	0.473
proficiency (languages without stress) * vowel quality	0.0041	0.0114	-0.0183 – 0.0265	0.721
proficiency (languages with stress) * vowel quality	0.0007	0.0066	-0.0121 – 0.0136	0.912
age * F0 (initial)	0.0001	0.0001	-0.0001 – 0.0002	0.433
sex * F0 (initial)	0.0009	0.0015	-0.0020 – 0.0039	0.535
proficiency (languages without stress) * F0 (initial)	-0.0021	0.002	-0.0061 – 0.0019	0.3
proficiency (languages with stress) * F0 (initial)	0.0007	0.0009	-0.0011 – 0.0025	0.473
age * F0 (slope)	0.0005	0.0003	-0.0001 – 0.0012	0.122
sex * F0 (slope)	-0.0007	0.0058	-0.0121 – 0.0107	0.903
proficiency (languages without stress) * F0 (slope)	-0.0061	0.008	-0.0219 – 0.0096	0.444
proficiency (languages with stress) * F0 (slope)	0.0016	0.0041	-0.0064 – 0.0096	0.694

Appendix F. Supplementary Statistical Tables for Chapter 4 : Stress and Tone Features

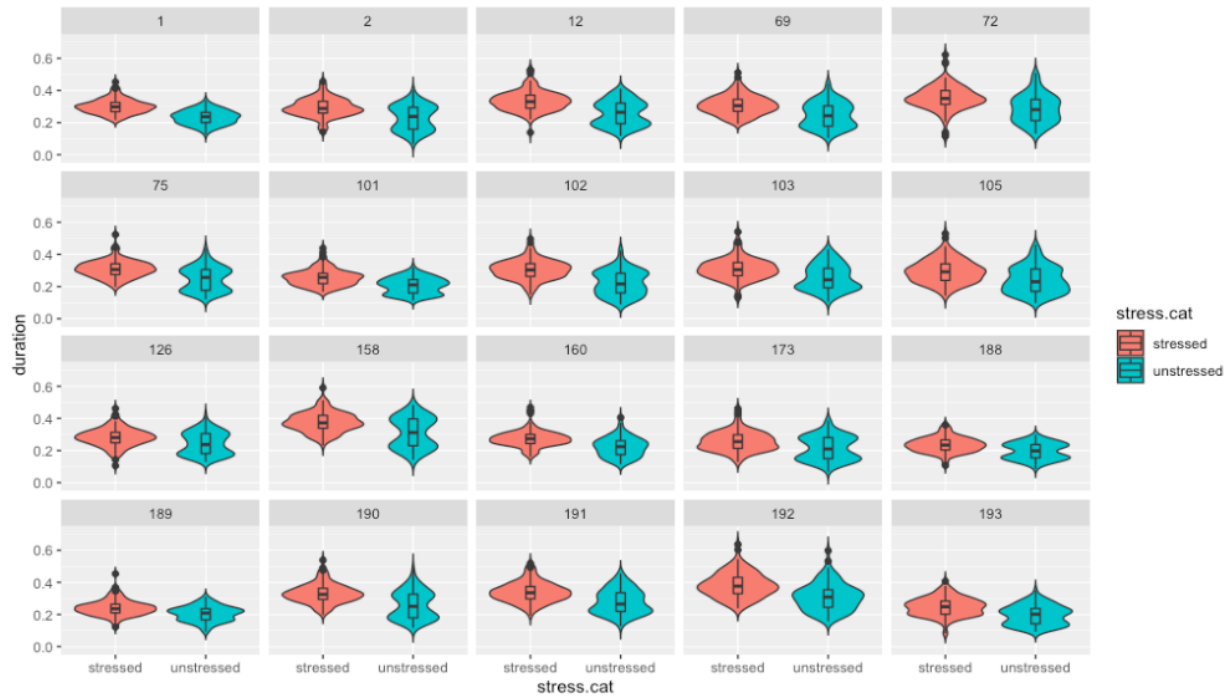


Figure 38. Syllable duration (indicative of actual stress) in syllables expected to be stressed/unstressed for each participant

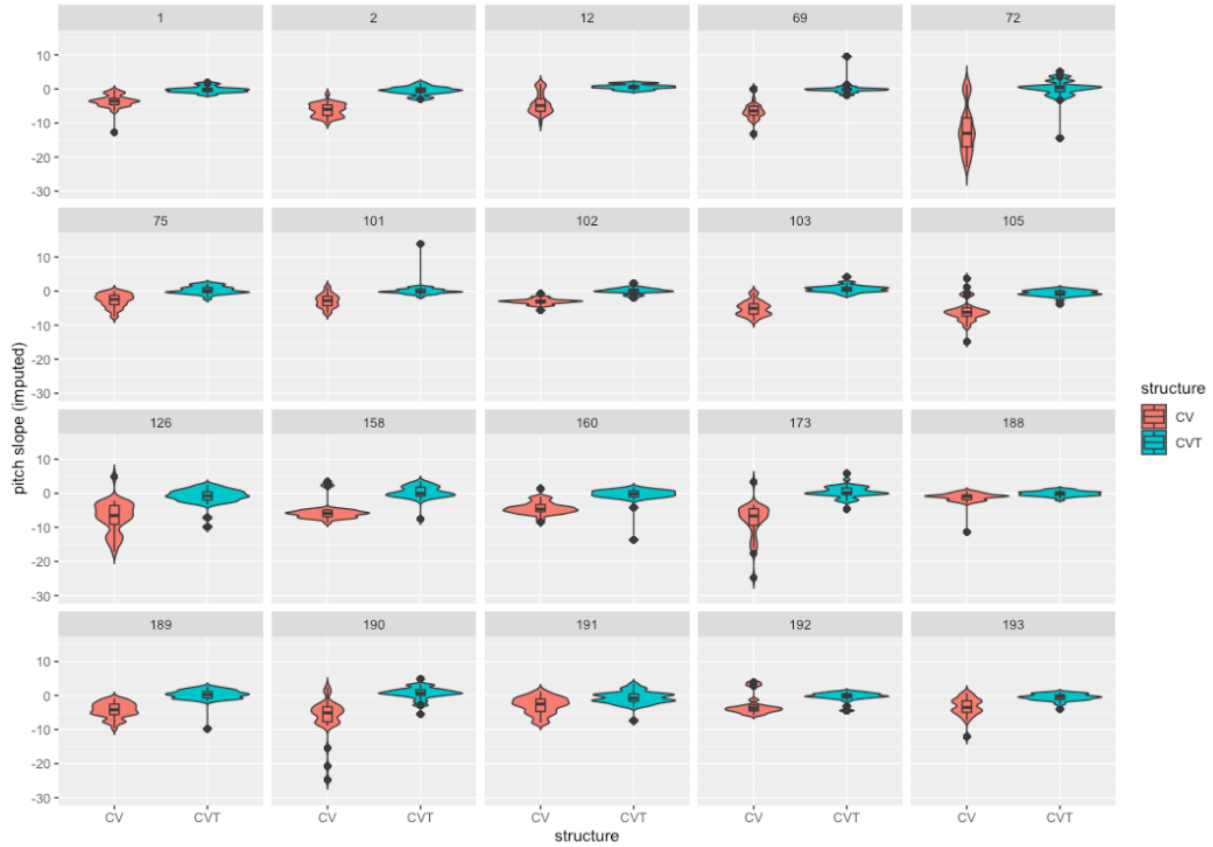


Figure 39. Pitch slope (imputed) in CV and CVT syllables in Tagalog- and English-origin words for each participant; slope means close to zero indicate use of high tone, slope means -5 and below indicate use of falling tone

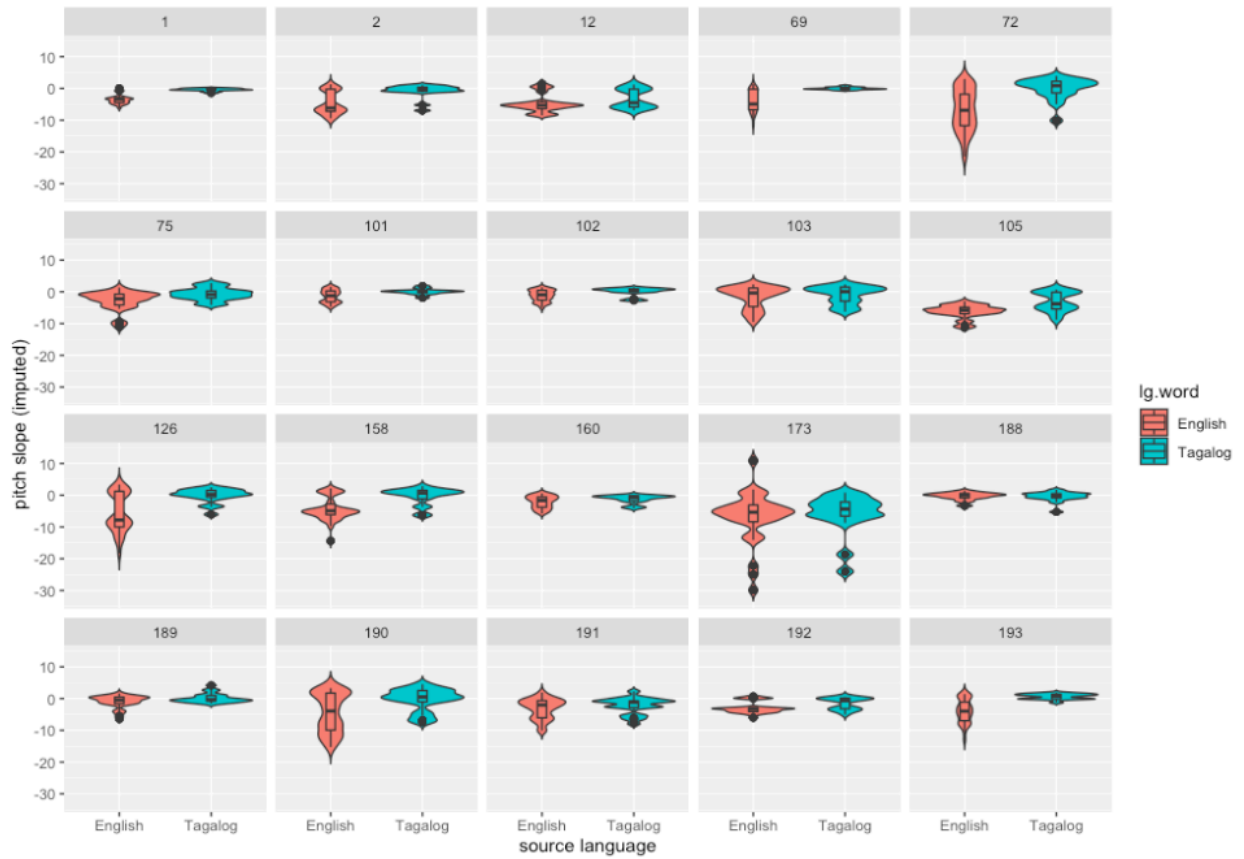


Figure 40. Pitch slope (imputed) in CVR-English and CVR-Tagalog syllables for each participant; slope means close to zero indicate use of high tone, slope means -5 and below indicate use of falling tone

Appendix G. Supplementary statistical tables for Chapter 5 : Conjunction and Preposition
Lexical Patterns

Table 52. Spread scores of conjunctions by class - *Speaker (with)*: speakers who derived at least one conjunction from that class from the expected language; *Speakers (all)*: speakers who had at least one conjunction from that class

Class	Supposed to be derived from	Speakers (with)	Speakers (all)	Spread score
Adversative	Tagalog	65	85	0.7647
Cumulative (non-emphatic)	Hokkien	89	90	0.9889
Cumulative (emphatic)	Tagalog	31	40	0.775
Disjunctive	Hokkien	84	88	0.9545
Conditional (general)	Hokkien	87	87	1
Conditional (specific)	Tagalog	25	33	0.7576
Conditional ('unless')	English	8	12	0.6667
Concession	Tagalog	30	41	0.7317
Result	Tagalog	32	32	1
Location	Tagalog	0	0	NA
Manner	Hokkien	58	75	0.7733
Reason	Tagalog	69	82	0.8415
Substitution	Tagalog	14	15	0.9333
Temporal ('after')	English	9	10	0.9
Temporal (general)	Tagalog	23	31	0.7419
Relativizer/Complementizer (general)	Tagalog	41	41	1
Relativizer/Complementizer (specific)	Tagalog	0	0	NA

Some speakers did not derive any conjunction in a particular class from the expected language at all (e.g., in the first row, I report that some speakers – 23.53% of 85 speakers who used at least one adversative conjunction –⁸² did not source their adversative conjunctions from Tagalog at all): this is indicated by the rates in Table 52, column 5 that do not have the score of 1. Speakers who produced at least one general conditional conjunction (e.g., *nā* ‘if’), conjunction of result (e.g., *kayâ* ‘that is why), or general relativizer/complementizer (e.g., *nà*) derived these conjunctions consistently from the from the expected language (see Table 52 column 2). Scores

⁸² This value was derived by subtracting the spread rate in the fifth column from 1 (i.e., $1 - 0.7647 = 0.2353$) and multiplying it by 100 to get a percentage.

of 1 for those classes indicate this (see column 5). None of the speakers used any conjunctions of location (e.g., *kungsaân* ‘where’) or relativizers and complementizers that are time-, manner-, or location-specific (e.g., *nûng* ‘when’).

Table 53. Mean intraspeaker consistency scores and interspeaker pattern inconsistency scores (conjunctions) by class

Class	n	Mean intraspeaker consistency score	SD	Interspeaker pattern inconsistency score
Adversative	65	0.6115	0.3458	0.5655
Cumulative (non-emphatic)	89	0.881	0.1806	0.2049
Cumulative (emphatic)	31	0.8327	0.2798	0.336
Disjunctive	84	0.9267	0.1384	0.1494
Conditional (general)	87	0.9484	0.1139	0.1201
Conditional (specific)	25	0.98	0.1	0.102
Conditional ('unless')	8	1	0	0
Concession	30	0.7429	0.352	0.4739
Result	32	0.9984	0.0093	0.0093
Location	0	<i>NA</i>	<i>NA</i>	<i>NA</i>
Manner	58	0.897	0.2106	0.2348
Reason	69	0.6491	0.3436	0.5294
Substitution	14	0.8899	0.2509	0.2819
Temporal ('after')	9	1	0	0
Temporal (general)	23	0.8124	0.2847	0.3504
Relativizer/Complementizer (general)	41	1	0	0
Relativizer/Complementizer (specific)	0	<i>NA</i>	<i>NA</i>	<i>NA</i>

Table 54. Regression results – likelihood to use Hokkien-derived conjunctions (observations = 15,901, $R^2 = 0.632$, random intercepts for speaker)

Predictors	Log-Odds	SE	CI	<i>p</i>
(Intercept)	-0.98	0.33	-1.62 – -0.34	0.003
Classes of conjunctions supposed to be from Hokkien (part vs. non-part)	3.65	0.12	3.41 – 3.89	<0.001
Age (young vs. old)	0.34	0.47	-0.57 – 1.25	0.467
Sex (male vs. female)	0.24	0.35	-0.44 – 0.92	0.492
Proficiency (Hokkien)	0.87	0.29	0.30 – 1.44	0.003
proficiency (Non-Hokkien)	0.03	0.2	-0.36 – 0.42	0.897
Classes * Age	-0.41	0.2	-0.81 – -0.01	0.045
Classes * Sex	0.4	0.14	0.13 – 0.67	0.004
Classes * Proficiency (Hokkien)	-0.21	0.12	-0.44 – 0.02	0.069
Classes * Proficiency (Non-Hokkien)	0.06	0.08	-0.10 – 0.21	0.462

Table 55. Regression results – likelihood to use Tagalog-derived conjunctions (observations = 15,901, $R^2 = 0.447$, no random effects)

Predictors	Log-Odds	SE	CI	<i>p</i>
(Intercept)	-4.13	0.14	-4.41 – -3.88	<0.001
Classes of conjunctions supposed to be from Tagalog (part vs. non-part)	4.45	0.14	4.18 – 4.74	<0.001
Age (young vs. old)	0.51	0.17	0.17 – 0.84	0.003
Sex (male vs. female)	-0.55	0.17	-0.88 – -0.22	0.001
Proficiency (Tagalog)	0.48	0.13	0.23 – 0.73	<0.001
proficiency (Non-Tagalog)	-0.65	0.09	-0.81 – -0.48	<0.001
Classes * Age	-0.01	0.18	-0.38 – 0.34	0.936
Classes * Sex	0.17	0.18	-0.18 – 0.53	0.333
Classes * Proficiency (Tagalog)	-0.5	0.14	-0.77 – -0.22	<0.001
Classes * Proficiency (Non-Tagalog)	0.49	0.09	0.31 – 0.67	<0.001

Table 56. Regression results – likelihood to use English-derived conjunctions (observations = 15,901, $R^2 = 0.038$, no random effects)

Predictors	Log-Odds	SE	CI	<i>p</i>
Intercept	-3.59	0.09	-3.78 – -3.42	<0.001
Classes of conjunctions supposed to be from English (part vs. non-part)	5.21	0.69	3.97 – 6.78	<0.001
Age (young vs. old)	0.7	0.11	0.48 – 0.91	<0.001
Sex (male vs. female)	0.76	0.08	0.60 – 0.92	<0.001
Proficiency (English)	-0.06	0.07	-0.20 – 0.08	0.386
Proficiency (Non-English)	0.08	0.05	-0.02 – 0.18	0.101
Classes * Age	14.37	229.27	-8.34 – NA	0.95
Classes * Sex	-4.18	1.33	-7.46 – -1.92	0.002
Classes * Proficiency (English)	1.06	1.15	-1.08 – 3.65	0.355
Classes * Proficiency (Non-English)	0.49	0.99	-1.53 – 2.57	0.625

Table 57. Spread scores of prepositions by class - *Speaker (with)*: speakers who derived at least one preposition from that class from the expected language; *Speakers (all)*: speakers who had at least one preposition from that class

Class	Supposed to be derived from	Speakers (with)	Speakers (all)	Spread
Accompaniment	English	25	25	1
‘Of’	English	57	62	0.9194
Location	Hokkien	89	90	0.9889
Orientation	Hokkien	68	69	0.9855
Range/path	Hokkien	90	91	0.989
Temporal	English	15	18	0.8333
Spatial	English	50	81	0.6173

Out of the seven classes, only the first class – prepositions of accompaniment (e.g., *with* ‘with’) – is fully diffused: all 25 speakers who had accompaniment prepositions sourced at least one of these prepositions from English. This is not the case for the other six classes. For these six, most – not all – speakers derived the class of prepositions from the expected language (Table 57, column 2) at least once; some speakers would not do this at all. The spread rates in the last column of Table 57 that are not 1 indicate this. For instance, the majority of speakers (91.94%) who produced at least one preposition meaning ‘of’ (e.g., *ôf* ‘of’) derived the preposition(s) from English at least once. The rest – 8.06% – did not do this at all. Overall, spread rates are consistently above 50% (mean = 90.47%, SD = 14.01).

Table 58. Mean intraspeaker consistency scores and interspeaker pattern inconsistency scores (prepositions) by class

Class	n	Mean intraspeaker consistency score	SD	Interspeaker pattern inconsistency score
Accompaniment	25	0.9964	0.0182	0.0182
'Of'	57	0.9551	0.1284	0.1345
Location	89	0.9409	0.0851	0.0904
Orientation	68	0.9788	0.089	0.091
Range/path	90	0.8459	0.1559	0.1843
Temporal	15	0.9867	0.0516	0.0523
Spatial	50	0.6817	0.3009	0.4413

Table 59. Regression results – likelihood to use Hokkien-derived prepositions (observations = 8,134, $R^2 = 0.506$, random intercepts for participant)

Predictors	Log-Odds	SE	CI	<i>p</i>
(Intercept)	-1.4	0.32	-2.04 – -0.77	<0.001
Classes of prepositions supposed to be from Hokkien (part vs. non-part)	3.46	0.2	3.07 – 3.85	<0.001
Age (young vs. old)	0.02	0.4	-0.77 – 0.81	0.961
Sex (male vs. female)	0.65	0.31	0.04 – 1.27	0.036
Proficiency (Hokkien)	1.28	0.26	0.76 – 1.79	<0.001
Proficiency (Non-Hokkien)	-0.28	0.18	-0.63 – 0.08	0.131
Classes * Age	0.55	0.27	0.02 – 1.09	0.043
Classes * Sex	-0.73	0.2	-1.13 – -0.33	<0.001
Classes * Proficiency (Hokkien)	-0.57	0.18	-0.91 – -0.23	0.001
Classes * Proficiency (Non-Hokkien)	0.24	0.12	0.01 – 0.48	0.045

Table 60. Regression results – likelihood to use Tagalog-derived prepositions (observations = 8,134, $R^2 = 0.368$, random intercepts for participant)

Predictors	Log-Odds	SE	CI	<i>p</i>
(Intercept)	-4.31	0.3	-4.90 – -3.72	<0.001
Age (young vs. old)	0.17	0.36	-0.53 – 0.87	0.632
Sex (male vs. female)	-0.44	0.35	-1.13 – 0.24	0.207
Proficiency (Tagalog)	1.07	0.29	0.51 – 1.63	<0.001
Proficiency (Non-Tagalog)	-0.54	0.18	-0.89 – -0.18	0.003

Table 61. Logistic regression results – likelihood to use English-derived prepositions (observations = 8,134, $R^2 = 0.473$, random intercepts for participant)

Predictors	Log-Odds	SE	CI	<i>p</i>
(Intercept)	-2.79	0.23	-3.24 – -2.33	<0.001
Classes of prepositions supposed to be from English (part vs. non-part)	2.37	0.17	2.04 – 2.70	<0.001
Age (young vs. old)	0.09	0.34	-0.56 – 0.75	0.784
Sex (male vs. female)	-0.02	0.29	-0.57 – 0.54	0.957
Proficiency (English)	0	0.23	-0.45 – 0.44	0.984
Proficiency (Non-English)	0.02	0.17	-0.32 – 0.36	0.919
Classes * Age	0.13	0.24	-0.33 – 0.59	0.583
Classes * Sex	0.8	0.2	0.42 – 1.19	<0.001
Classes * Proficiency (English)	0.58	0.17	0.24 – 0.92	0.001
Classes * Proficiency (Non-English)	-0.37	0.13	-0.62 – -0.13	0.003

Table 62. Logistic regression results – likelihood to use Mandarin-derived prepositions (observations = 8,134, $R^2 = 0.733$, random intercepts for participant)

Predictors	Log-Odds	SE	CI	<i>p</i>
(Intercept)	-12.46	3.66	-19.63 – -5.28	0.001
Age (young vs. old)	0.06	0.05	-0.03 – 0.15	0.226
Sex (male vs. female)	-0.87	1.53	-3.86 – 2.13	0.571
Proficiency (Mandarin)	-0.81	1.06	-2.88 – 1.26	0.444
Proficiency (Non-Mandarin)	1.2	0.87	-0.51 – 2.91	0.168

REFERENCES

- ALDRIDGE, EDITH. 2002. Nominalization and wh-movement in Seediq and Tagalog. *Language and Linguistics* 3.393–426.
- ANDERSON, VICTORIA. 2006. Lexical stress without postlexical head-marking: Evidence from Tagalog. *The Journal of the Acoustical Society of America* 120.3092.
- ANG SEE, TERESITA. 1990. *The Chinese in the Philippines: Problems and perspectives*. Vol. 1. Manila: Kaisa Para Sa Kaunlaran.
- ATKINSON, EMILY.; AARON APPLE.; KYLE RAWLINS.; and AKIRA OMAKI. 2016. Similarity of wh-phrases and acceptability variation in wh-islands. *Frontiers in Psychology* 6.2048.
- BAKKER, PETER. 1994. Michif, the Cree-French mixed language of the Metis buffalo hunters in Canada. *Mixed Languages: 15 Case Studies in Language Intertwining*, ed. by Peter Bakker and Maarten Mous, 13–33. Amsterdam: Uitgave IFOTT.
- BAKKER, PETER. 1997. *A language of our own: the genesis of Michif, the mixed Cree-French language of the Canadian Métis*. Rev. ed. Oxford studies in anthropological linguistics 10. New York: Oxford University Press.
- BALARI, SERGIO.; and GUILLERMO LORENZO. 2018. The internal, the external and the hybrid: The state of the art and a new characterization of language as a natural object. *Glossa: a journal of general linguistics* 3. doi:10.5334/gjgl.330. <https://www.glossa-journal.org/article/id/4990/>.
- BAPTISTA, MARLYSE. 2005. New directions in pidgin and creole studies. *Annual Review of Anthropology* 34.33–42.
- BAPTISTA, MARLYSE. 2015. Continuum and variation in Creoles: Out of many voices, one language. *Journal of Pidgin and Creole Languages* 30.225–264.
- BAPTISTA, MARLYSE. 2020. Competition, selection, and the role of congruence in creole genesis and development. *Language* 96.160–199.
- BATES, DOUGLAS.; MARTIN MÄCHLER.; BEN BOLKER.; and STEVE WALKER. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.
- BAUTISTA, MA. LOURDES S. 2004. Researching English in the Philippines: Bibliographical resources. *World Englishes* 23.199–210.
- BENOR, SARAH BUNIN. 2010. Ethnolinguistic repertoire: Shifting the analytic focus in language and ethnicity. *Journal of Sociolinguistics* 14.159–183.
- BIBLICAL SEMINARY OF THE PHILIPPINES 2011. The Amoy Audio Bible Project. *Biblical Seminary of the Philippines*. <https://bsop.edu.ph/the-amoy-bible-project/>.
- BODMAN, NICHOLAS C. 1987. *Spoken Amoy Hokkien*. New York: Language Services.
- BOERSMA, PAUL.; and DAVID WEENINK. 2021. *Praat: doing phonetics by computer*. <http://www.praat.org/>.

- BORBÉLY, ANNA. 1995. Attitudes as a factor of language choice: A sociolinguistic investigation in a bilingual community of Romanian-Hungarians. *Acta Linguistica Hungarica* 43.311–321.
- BOSMA, EVELYN.; and ELMA BLOM. 2019. A code-switching asymmetry in bilingual children: Code-switching from Dutch to Frisian requires more cognitive control than code-switching from Frisian to Dutch. *International Journal of Bilingualism* 23.1431–1447.
- BUCHSTALLER, ISABELLE.; ANNE KRAUSE-LERCHE.; and JOHANNA MECHLER. 2021. Exploring the effect of linguistic architecture and heuristic method in panel analysis. *Language Variation and Language Change Across the Lifespan: Theoretical and Empirical Perspectives from Panel Studies*, ed. by Karen V. Beaman and Isabelle Buchstaller, 185–208. 1st ed. Routledge. doi:10.4324/9780429030314. <https://www.taylorfrancis.com/books/9780429641695>.
- BUSSMANN, HADUMOD. 2006. *Routledge dictionary of language and linguistics*. 2nd ed. London: Routledge.
- CALUDE, ANDREEA SIMONA.; STEVEN MILLER.; and MARK PAGEL. 2020. Modelling loanword success – a sociolinguistic quantitative study of Māori loanwords in New Zealand English. *Corpus Linguistics and Linguistic Theory* 16.29–66.
- CHAMBERS, J. K. 2002. Dynamics of dialect convergence. *Journal of Sociolinguistics* 6.117–130. doi:10.1111/1467-9481.00180.
- CHAO, YUAN REN. 1968. *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.
- CHAPPELL, HILARY. 2019. Southern Min. *The Mainland Southeast Asia linguistic area*, ed. by A. Vittrant and J. Watkins. Berlin: De Gruyter Mouton.
- CHELLIAH, SHOBHANA L.; and WILLEM J. DE REUSE. 2011. *Handbook of Descriptive Linguistic Fieldwork*. Dordrecht: Springer Netherlands. doi:10.1007/978-90-481-9026-3. <http://link.springer.com/10.1007/978-90-481-9026-3>.
- CHEN, HONGJUAN. 2006. Utterance final lengthening and focus induced lengthening in standard Chinese bi-syllabic words. Edinburgh, United Kingdom: The University of Edinburgh master's thesis.
- CHENG, CHIN-CHUAN. 1973. A quantitative study of tone in Chinese. *Journal of Chinese Linguistics* 1.93–110.
- CHEUNG, CANDICE CHI-HANG. 2014. Wh-fronting and the left periphery in Mandarin. *Journal of East Asian Linguistics* 23.393–431.
- CHOMSKY, NOAM. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- CHOW, UNA Y. 2016. L2 transfer of stress, tones, and intonation from Mandarin: A case study. *Calgary Working Papers in Linguistics* 29.19–40.
- CHU, RICHARD. 2021. From 'sangley' to 'Chinaman', 'Chinese Mestizo' to 'Tsinoy': Unpacking 'Chinese' identities in the Philippines at the turn of the Twentieth-Century. *Asian Ethnicity*. doi:10.1080/14631369.2021.1941755.
- CHUA, DOROTHY ANG. 2004. From Chinese to Filipino: Changing identities of the Chinese in the Philippines. Vancouver, Canada: The University of British Columbia master's thesis.
- CHUAUNSU, REBECCA. 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. Manila: University of the Philippines master's thesis.

- CLARKE, SANDRA.; and ANDREW ERSKINE. 2010. Language attitudes and language change. *Newfoundland and Labrador English*, ed. by Sandra Clarke. Oxford: Oxford University Press.
- COELHO, F. ADOLPHO. 1881. *Os dialectos românicos ou neo-latinos na África, Ásia e América [The Romance or Neo-Latin dialects of Africa, Asia, and America: The study of the creole of the island of Santiago]*. Lisboa, Portugal: Casa da Sociedade de Geographia. <https://www.gutenberg.org/files/33159/33159-h/33159-h.htm>.
- COHEN, CLARA.; SYED WAQAR NABI.; CATHERINE F HIGHAM.; MICHAEL PUTNAM.; and GERRIT JAN KOOTSTRA. 2021. Individual variation in the structure of bilingual grammars. *Language* 97.752–792.
- COMRIE, BERNARD. 1976. *Aspect: An Introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.
- CRYSTAL, DAVID. 2008. *A dictionary of linguistics and phonetics*. 6th ed. Malden, MA: Blackwell.
- DA SILVA, AUGUSTO SOARES. 2010. Replication, selection and language change. Why an evolutionary approach to language variation and change? *Revista Portuguesa de Filosofia* 66.803–818.
- DAGANZO, CARLOS F.; VIKASH V. GAYAH.; and ERIC J. GONZALES. 2012. The potential of parsimonious models for understanding large scale transportation systems and answering big picture questions. *EURO Journal on Transportation and Logistics* 1.47–65.
- D'ARCY, ALEX. 2005. The development of linguistic constraints: Phonological innovations in St. John's English. *Language Variation and Change* 17. doi:10.1017/S0954394505050131. http://www.journals.cambridge.org/abstract_S0954394505050131.
- DEFIOR, SYLVIA.; and JÉSUS ALEGRIA. 2005. Morphosyntax and spelling connection: When phonology is (almost) enough to spell. *Revista de Logopedia, Foniatria y Audiología* 25.51–61.
- DICKSON, GREG.; and GAUTIER DURANTIN. 2019. Variation in the reflexive in Australian Kriol. *Asia-Pacific Language Variation* 5.171–207.
- DODSWORTH, ROBIN M. 2005. Linguistic variation and sociological consciousness. Columbus, Ohio: The Ohio State University ph.d. dissertation.
- DOEPPERS, DANIEL. 1986. Destination, selection and turnover among Chinese migrants to Philippine cities in the nineteenth century. *Journal of Historical Geography* 12.235–260.
- DUANMU, SAN. 2007. *The phonology of standard Chinese*. 2nd ed. Oxford: Oxford University Press.
- DY, CARMEN J. 1972. The syntactic structures of Amoy as used in the Philippines. *Philippine Journal of Linguistics* 3.75–94.
- ECKERT, PENELOPE. 1989. The whole woman: Sex and gender differences in variation. *Language Variation and Change* 1.245–267.
- ECKERT, PENELOPE. 2005. Variation, convention, and social meaning. Paper presented at the Annual Meeting of the Linguistic Society of America, Oakland, CA.
- ECKERT, PENELOPE. 2012. Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual Review of Anthropology* 41.87–100.
- ECKERT, PENELOPE.; and SALLY MCCONNELL-GINET. 2003. *Language and Gender*. Cambridge University Press.

- ESPOSITO, LEWIS.; and CHANTAL GRATTON. 2020. Prosody and ideologies of embodiment: Variation in the use of pitch and articulation rate among fitness instructors. *Language in Society* 1–26. doi:10.1017/S0047404520000809.
- FERNÁNDEZ, MAURO. 2018. El pidgin chino-español de Manila a principios del siglo XVIII. *Zeitschrift für Romanische Philologie* 134.137–170.
- FERNÁNDEZ, MAURO.; and EEVA SIPPOLA. 2017. A new window into the history of Chabacano: Two unknown mid-19th century texts. *Journal of Pidgin and Creole Languages* 32.304–338. doi:10.1075/jpcl.32.2.04fer.
- FISHER, CYNTHIA.; D. GEOFFREY HALL.; SUSAN RAKOWITZ.; and LILA GLEITMAN. 1994. When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua* 92.333–375.
- GHYSELEN, ANNE-SOPHIE.; and GUNTHER DE VOGELAER. 2018. Seeking Systematicity in Variation: Theoretical and Methodological Considerations on the “Variety” Concept. *Frontiers in Psychology* 9.385. doi:10.3389/fpsyg.2018.00385.
- GOLDBERG, YOAV. 2017. *Neural network methods for Natural Language Processing*. (Ed.) Graeme Hirst. California, USA: Morgan and Claypool.
- GOLOVKO, EVGENIY V. 1994. Mednyj Aleut or Copper Island Aleut: An Aleut-Russian mixed language. *Mixed Languages: 15 Case Studies in Language Intertwining*, ed. by Peter Bakker and Maarten Mous, 113–121. Amsterdam: Uitgave IFOTT.
- GONZALES, WILKINSON DANIEL WONG. 2016. Trilingual code-switching using quantitative lenses: An exploratory study on Hokaglish. *Philippine Journal of Linguistics* 47.106–128.
- GONZALES, WILKINSON DANIEL WONG. 2017a. Chinese Filipinos tag their questions, kiam si? Some notes on tag questions in Philippine Hybrid Hokkien. *Philippine ESL Journal* 19.84–99.
- GONZALES, WILKINSON DANIEL WONG. 2017b. Philippine Englishes. *Asian Englishes* 19.79–95. doi:10.1080/13488678.2016.1274574.
- GONZALES, WILKINSON DANIEL WONG. 2017c. 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* 1.185–212. doi:10.1075/le.1.2.04gon.
- GONZALES, WILKINSON DANIEL WONG. 2018. Philippine Hybrid Hokkien as a postcolonial mixed language: Evidence from nominal derivational affixation mixing. Singapore: National University of Singapore master’s thesis.
- GONZALES, WILKINSON DANIEL WONG. 2020. Tone in Tagalog and English? Prosodic adaption in Philippine Hybrid Hokkien. Poster. Paper presented at the Annual Meeting of the Linguistic Society of America. <https://www.linguisticsociety.org/abstract/tone-tagalog-and-english-prosodic-adaption-philippine-hybrid-hokkien>.
- GONZALES, WILKINSON DANIEL WONG. 2021a. Filipino, Chinese, neither, or both? The Lannang identity and its relationship with language. *Language & Communication* 77.
- GONZALES, WILKINSON DANIEL WONG. 2021b. Interactions of Sinitic languages in the Philippines: Sinicization, Filipinization, and Sino-Philippine language creation. *The Palgrave handbook of Chinese language studies*, ed. by Zhengdao Ye. London: Palgrave-MacMillan.
- GONZALES, WILKINSON DANIEL WONG. 2022a. The Lannang Corpus (LanCorp): A POS-tagged, sociolinguistic corpus containing recordings and transcriptions of Lannang speech

- collected from the metropolitan Manila Lannangs between 2016 and 2020. *University of Michigan - Deep Blue Repository*. doi:10.7302/66g9-e028.
- GONZALES, WILKINSON DANIEL WONG. 2022b. Hybridization. *Philippine English: Development, Structure, and Sociology of English in the Philippines*, ed. by Ariane Macalinga Borlongan. London: Routledge.
- GONZALES, WILKINSON DANIEL WONG.; and MIE HIRAMOTO. 2020. Two Englishes diverged in the Philippines? A substratist account of Manila Chinese English: A substratist account of Manila Chinese English. *Journal of Pidgin and Creole Languages* 35.125–159. doi:10.1075/jpcl.00057.gon.
- GONZALES, WILKINSON DANIEL WONG.; and REBECCA LURIE STARR. 2020. Vowel system or vowel systems? Variation in the monophthongs of Philippine Hybrid Hokkien in Manila. *Journal of Pidgin and Creole Languages* 35.253–292.
- GONZALEZ, ANDREW. 1970. Acoustic correlates of accent, rhythm, and intonation in Tagalog. *Phonetica* 22.11–44.
- GORDON, ELIZABETH.; LYLE CAMPBELL.; JENNIFER HAY.; MARGARET MACLAGAN.; ANDREA SUDBURY.; and PETER TRUDGILL. 2004. *New Zealand English: Its origins and evolution*. Cambridge, the United Kingdom: Cambridge University Press.
- GORDON, MATTHEW. 2001. A typology of contour tone restrictions. *Studies in Language* 25.423–462.
- GORDON, MATTHEW.; and TIMO ROETTGER. 2017. Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard* 3. doi:10.1515/lingvan-2017-0007. <https://www.degruyter.com/document/doi/10.1515/lingvan-2017-0007/html>.
- GÖRLACH, MANFRED. 2002. Ulster Scots - a language? *Still more Englishes*, 69–86. Amsterdam, the Netherlands: John Benjamins Publishing Company.
- GREEN, DAVID D. 1998. Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition* 1.67–81.
- HALL-LEW, LAUREN.; AMANDA CARDOSO.; and EMMA DAVIES. 2021. Social Meaning and Sound Change. *Social Meaning and Linguistic Variation: Theorizing the Third Wave*, ed. by Emma Moore, Lauren Hall-Lew, and Robert J. Podesva, 27–53. Cambridge: Cambridge University Press. doi:10.1017/9781108578684.002. <https://www.cambridge.org/core/books/social-meaning-and-linguistic-variation/social-meaning-and-sound-change/2B40EF08A0FCD55F4B087CB80FCAC595>.
- HAUGEN, EINAR. 1966. Dialect, language, nation. *American Anthropologist* 68.922–935.
- HEBDIGE, DICK. 1979. *Subculture: The meaning of style*. London: Routledge.
- HERMANS, DAAN.; THEO BONGAERTS.; KEES DE BOT.; and ROBERT SCHREUDER. 2003. Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition* 1.213–229.
- HURST, ELLEN. 2009. Tsotsitaal, global culture and local style: identity and recontextualisation in twenty-first century South African townships. *Social Dynamics* 35. Routledge.244–257. doi:10.1080/02533950903076196.
- HWANG, HYUN KYUN.; NAONORI NAGAYA.; and JULIÁN VILLEGAS. 2019. Cue weighting in the perception of Tagalog stress. *The Journal of the Acoustical Society of America* 146.3052.
- JIN, DAWEI. 2014. Contradiction, why-questions and complex islands in Chinese. *University of Pennsylvania Working Papers in Linguistics* 21.1–8.
- JUN, SUN-AH. 2007. Prosodic typology. *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford Univeristy Press.

- KAHN, JULIETTE.; NICOLAS AUDIBERT.; JEAN-FRANÇOIS BONASTRE.; and SOLANGE ROSSATO. 2011. Inter and intra-speaker variability in French: An analysis of oral vowels and its implication for automatic speaker verification. Paper presented at the ICPhS XVII, Hong Kong, SAR, People's Republic of China.
- KINGSTON, JOHN. 2011. Tonogenesis. *The Blackwell companion to phonology*, ed. by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice. New Jersey: John Wiley & Sons.
- KLAUS, JANA.; KRISTIN LEMHÖFER.; and HERBERT SCHRIEFERS. 2018. The second language interferes with picture naming in the first language: evidence for L2 activation during L1 production. *Language, Cognition and Neuroscience* 33.867–877. doi:10.1080/23273798.2018.1430837.
- KLOSS, HEINZ. 1968. *Die Entwicklung neuer germanischer Kultursprachen seit 1800*. Düsseldorf: Bagel.
- KONSTANTOPOULOS, SPYROS.; and LARRY HEDGES. 2019. Statistically analyzing effect sizes: Fixed- and random-effects models. *The handbook of research synthesis and meta-analysis*, ed. by Harris Cooper, Larry V. Hedges, and Jeffrey C. Valentine, 246–278. 3rd ed. New York: Russell Sage Foundation.
- KOSSMANN, MAARTEN. 2019. Is Dutch Straattaal a mixed multiethnolect? A Moroccan perspective. *Applied Linguistics Review* 10.293–316. doi:10.1515/applirev-2017-0050.
- KUO, YU-CHING.; STUART ROSEN.; and ANDREW FAULKNER. 2008. Acoustic cues to tonal contrasts in Mandarin: Implications for cochlear implants. *The Journal of the Acoustical Society of America* 123.2824.
- KUZNETSOVA, ALEXANDRA.; PER BRUUN BROCKHOFF.; and RUNE HAUBO BOJESEN CHRISTENSEN. 2019. *Tests in linear mixed effects models: Package 'lmerTest.'*
- LABOV, WILLIAM. 1972. The social motivation of a sound change. *Sociolinguistic patterns*, 251–265. New York: Academic.
- LABOV, WILLIAM. 1982. Building on empirical foundations. *Perspectives on historical linguistics*, ed. by Winfred Lehmann and Yakov Malkiel, 17–92. Amsterdam, the Netherlands: John Benjamins Publishing Company.
- LAFFERTY, JOHN.; ANDREW MCCALLUM.; and FERNANDO C. N. PEREIRA. 2001. Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data. *Proceedings of the 18th International Conference on Machine Learning*.282–289.
- LAGROU, EVELYNE.; ROBERT J. HARTSUIKER.; and WOUTER DUYCK. 2011. Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37.952–965.
- LEE, NALA HUIYING. 2014. A grammar of Baba Malay with sociophonetics considerations. Honolulu, HI: University of Hawai'i at Mānoa ph.d. dissertation.
- LEIMGRUBER, JAKOB.; JUN JIE LIM.; WILKINSON DANIEL WONG GONZALES.; and MIE HIRAMOTO. 2020. Ethnic and gender variation in the use of Colloquial Singapore English discourse particles. *English Language and Linguistics*.
- LESHO, MARIVIC. 2013. The sociophonetics and phonology of the Cavite Chabacano vowel system. Columbus, Ohio: The Ohio State University ph.d. dissertation.
- LESHO, MARIVIC. 2018. Philippine English (Metro Manila acrolect). *Journal of the International Phonetic Association* 48.357–370.
- LI, CHARLES N.; and SANDRA A. THOMPSON. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.

- LIN, PHILIP. 2015. *Taiwanese Grammar: A concise reference*. Durnham, NC: Greenhorn Media.
- LIPSKI, JOHN. 2020. Reconstructing the life-cycle of a mixed language: An exploration of Ecuadoran Media Lengua. *International Journal of Bilingualism* 24.410–436.
- LÜDECKE, DANIEL. 2018a. ggeffects - Estimated Marginal Means and Adjusted Predictions from Regression Models. *ggeffects*. <https://strengjacke.github.io/ggeffects/>.
- LÜDECKE, DANIEL. 2018b. ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software* 3.772. doi:10.21105/joss.00772.
- MACGOWAN, JOHN. 1922. *Dictionary of the Amoy Dialect*. London: Trubner & Co.
- MACLAGAN, MARGARET.; ELIZABETH GORDON.; and GILLIAN LEWIS. 1999. Women and sound change: Conservative and innovative behavior by the same speakers. *Language Variation and Change* 11.19–41.
- MAKONI, SINFREE, and ALASTAIR PENNYCOOK (eds.) 2007. *Disinventing and reconstituting languages*. Bilingual education and bilingualism 62. Clevedon, [England]: Buffalo ; Multilingual Matters.
- MATRAS, YARON.; and PETER BAKKER. 2003. The study of mixed languages. *The mixed language debate: Theoretical and empirical advances*, ed. by Yaron Matras and Peter Bakker, 1022. Berlin, Germany & New York: Mouton de Gruyter.
- MATRAS, YARON.; and JEANETTE SAKEL. 2007. Investigating the mechanism of pattern replication in language convergence. *Studies in Language* 31.825–865.
- MATTHEWS, P. H. 2007. *The concise Oxford dictionary of linguistics*. 2nd ed. Oxford: Oxford University Press.
- MEAKINS, FELICITY. 2012. Which Mix — code-switching or a mixed language? — Gurindji Kriol. *Journal of Pidgin and Creole Languages* 27.105–140. doi:10.1075/jpcl.27.1.03mea.
- MEAKINS, FELICITY. 2013. Mixed languages. *Contact languages*, ed. by Peter Bakker and Yaron Matras, 159–228. De Gruyter Mouton.
- MEAKINS, FELICITY.; and CARMEL O'SHANNESSY. 2010. Ordering arguments about: Word order and discourse motivations in the development and use of the ergative marker in two Australian mixed languages. *Lingua* 120.1693–1713. doi:10.1016/j.lingua.2009.05.013.
- MECHLER, JOHANNA.; and ISABELLE BUCHSTALLER. 2019. [In]stability in the use of a stable variable. *Linguistics Vanguard* 5. doi:10.1515/lingvan-2018-0024. <https://www.degruyter.com/document/doi/10.1515/lingvan-2018-0024/html>.
- MEIR, IRIT.; and WENDY SANDLER. 2019. Variation and conventionalization in language emergence: The case of two young sign language of Israel. *Linguistik Aktuell/Linguistics Today*, ed. by Edit Doron, Malka Rappaport Hovav, Yael Reshef, and Moshe Taube, 256:337–363. Amsterdam: John Benjamins Publishing Company. doi:10.1075/la.256.13mei. <https://benjamins.com/catalog/la.256.13mei>.
- MILLER, JOHN E.; TIAGO TRESOLDI.; ROBERTO ZARIQUIEY.; CÉSAR A. BELTRÁN CASTAÑÓN.; NATALIA MOROZOVA.; and JOHANN-MATTIS LIST. 2020. Using lexical language models to detect borrowings in monolingual wordlists. *PLOS ONE* 15.e0242709.
- MILROY, JAMES.; and LESLEY MILROY. 1985. Linguistic Change, social network and speaker innovation. *Journal of Linguistics* 21.339–384.
- MINTZ, TOBEN H.; and DANI BYRD. 2010. *Discovering speech, words, and mind*. 1st ed. Oxford: Wiley-Blackwell.
- MONAGHAN, PADRAIC.; and SEÁN G. ROBERTS. 2019. Cognitive influences in language evolution: Psycholinguistic predictors of loan word borrowing. *Cognition* 186.147–158.

- OBEIDAT, HUSSEIN.; and KHADIDJA HAMMOUDI. 2019. Men as leaders of linguistic change: Case of Tlemcen speech community. *Jordan Journal of Modern Languages and Literature* 11.371–384.
- OBERAUER, KLAUS. 2009. Interference between storage and processing in working memory: Feature overwriting, not similarity-based competition. *Memory & Cognition* 37.346–357.
- OHALA, JOHN. 1978. Production of tone. *Tone: A linguistic survey*, ed. by V. A. Fromkin, 3–39. New York: Academic Press.
- O'SHANNESY, CARMEL. 2020. Mixed languages. *The Routledge Handbook of Language Contact*, ed. by Evangelia Adamou and Yaron Matras, 325–248. 1st ed. London: Routledge.
- PALACIOS-MARTÍNEZ, IGNACIO M. 2018. Lexical innovation in the language of teenagers: A cross-linguistic perspective. *Jugendsprachen/Youth languages*. Berlin: De Gruyter Mouton.
- PÉLABON, CHRISTOPHE.; CHRISTOFFER H. HILDE.; SIGURD EINUM.; and MARLÈNE GAMELON. 2020. On the use of the coefficient of variation to quantify and compare trait variation. *Evolution Letters* 4.180–188.
- PHAM, GIANG.; DANAE DONOVAN.; QUYNH DAM.; and AMY CONTANT. 2018. Learning Words and Definitions in Two Languages: What Promotes Cross-Language Transfer? *Language Learning* 68.206–233. doi:10.1111/lang.12274.
- PLASS, JAN L.; DOROTHY M. CHUN.; RICHARD E. MAYER.; and DETLEV LEUTNER. 2003. Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior* 19.221–243.
- POA, DORY. 2004. From Quadrilingual to Bilingual: On the Multilingual Teaching in the Chinese Schools in the Philippines). *Bilingual Studies*, 2:96–107. Beijing: Minzu Chubanshe.
- PRINCE, ALAN.; and PAUL SMOLENSKY. 2008. *Optimality Theory: Constraint interaction in generative grammar*. New Jersey: Wiley-Blackwell.
- PRIOR, ANAT.; and BRIAN MACWHINNEY. 2009. A bilingual advantage in task switching. *Bilingualism: Language and Cognition* 13.253–262.
- R CORE TEAM 2015. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. <http://www.R-project.org>.
- REINECKE, JOHN E.; STANLEY M. TSUZAKI.; DAVID DECAMP.; IAN F. HANCOCK.; and RICHARD E. WOOD. 1975. Chabacano (Philippine Creole Spanish). *A bibliography of pidgin and creole languages*, 210–216. Oceanic Linguistics Special Publication 14.
- REMIJSEN, BERT. 2014. The study of tone in languages with a quantity contrast. *Language Documentation & Conservation* 8.672–689.
- REYES, ANGELA. 2017. Inventing Postcolonial Elites: Race, Language, Mix, Excess. *Journal of Linguistic Anthropology* 27.210–231.
- RIEDER, MARIA. 2018. Irish Travellers' views on Cant: what folk criteria of languageness tell us about the community. *Language Awareness* 27.40–60. doi:10.1080/09658416.2018.1431242.
- SANDE, HANNAH. 2015. Nouchi as a Distinct Language: The Morphological Evidence. *Selected Proceedings of the 44th Annual Conference on African Linguistics*.243–253.
- SANKOFF, GILLIAN. 2006. Age: Apparent time and real time. *The encyclopedia of Language & Linguistics*, ed. by Keith Brown, 110–116. 2nd ed. Oxford: Elsevier.

- SATO, YOSUKE. 2013. Wh-Questions in Colloquial Singapore English: Adaptive traits from vernacular Malay and typological congruence. *Journal of Pidgin and Creole Languages* 28.299–322.
- SCHACHTER, PAUL.; and FE T. OTANES. 1972. *Tagalog reference grammar*. Berkeley, CA: Univeristy of California Press.
- SCHACHTER, PAUL.; and TIMOTHY SHOPEN. 2007. Parts-of-speech systems. *Language typology and syntactic description*, ed. by Timothy Shopen, I:1–60. Cambridge, the United Kingdom: Cambridge University Press.
- SCHMID, HANS-JÖRG. 2020. *The dynamics of the linguistic system*. Oxford: Oxford Univeristy Press.
- SHARMA, DEVYANI.; and LAVANYA SANKARAN. 2011. Cognitive and social forces in dialect shift: Gradual change in London Asian speech. *Language Variation and Change* 23.399–428. doi:10.1017/S0954394511000159.
- SHETEWI, OUROOBA. 2018. Acquisition of Sociolinguistic Variation in a Dialect Contact Situation: The Case of Palestinian Children and Adolescents in Syria. Tyne, England: Newcastle University ph.d. dissertation.
- SHIN, NAOMI LAPIDUS. 2013. Women as leaders of language change: A qualification from the bilingual perspective. *Selected Proceedings of the 6th Workshop on Spanish Sociolinguistics*.135–147.
- SHOPEN, TIMOTHY. 2007. *Language typology and syntactic description: Clause structure*. Vol. I. Cambridge, the United Kingdom: Cambridge University Press.
- SIEGEL, JEFF. 2012. Two Types of Functional Transfer in Language Contact. *Journal of Language Contact* 5.187–215. doi:10.1163/187740912X639247.
- SIPPOLA, EEVA. 2021. Ilokano-Spanish: Borrowing, code-switching or a mixed language? *New Perspectives on Mixed Languages*, ed. by Maria Mazzoli and Eeva Sippola, 253–276. De Gruyter Mouton.
- SONDEREGGER, MORGAN. 2022. *Regression modeling for linguistic data*. 1st ed. Open Science Framework. <https://osf.io/pnumg/>.
- STARR, REBECCA LURIE.; and BRINDA BALASUBRAMANIAM. 2019. Variation and change in English /r/ among Tamil Indian Singaporeans. *World Englishes* 38.630–643. doi:10.1111/weng.12357.
- STARR, REBECCA LURIE.; and TIANXIAO WANG. 2021. Navigating variation amid contested norms and societal shifts: A case study of two L2 Mandarin speakers in Singapore. *Sociolinguistic Variation and Language Acquisition Across the Lifespan*, ed. by Anna Ghimenton, Aurélie Nardy, and Jean-Pierre Chevrot, 199–226. Amsterdam, the Netherlands: John Benjamins Publishing Company.
- STYLER, WILL. 2011. *Using Praat for linguistic research*. <https://wstyler.ucsd.edu/praat/>.
- TAKAHESU TABORI, ANDREA A.; EMILY N. MECH.; and NATSUKI ATAGI. 2018. Exploiting Language Variation to Better Understand the Cognitive Consequences of Bilingualism. *Frontiers in Psychology* 9.1686. doi:10.3389/fpsyg.2018.01686.
- TAN, CHERYL XUE ER. 2014. The aspectual system in Singapore Hokkien. Singapore: Nanyang Technological University unpublished bachelor's thesis.
- TAN, SUSAN VILLANUEVA. 1993. The education of Chinese in the Philippines and Koreans in Japan. Hong Kong, SAR, People's Republic of China: University of Hong Kong.
- TANG, C.-C. JANE. 1988. *Wh-topicalization in Chinese*. (manuscript). Ithaca, NY: Cornell University.

- TAY, LI CHIA.; MEI YUIT CHAN.; NGEE THAI YAP.; and BEE ENG WONG. 2016. Discourse particles in Malaysian English What do they mean? *Bidragen Tot De Taal* 172.479–509.
- THE LANNANG ARCHIVES 2020. Lannang Orthography. *The Lannang Archives*. The Lannang Archives. <https://www.lannangarchives.org>.
- THOMASON, SARAH. 2001. *Language contact: An introduction*. Washington, D. C.: Georgetown University Press.
- THOMASON, SARAH. 2007. Language contact and deliberate change. *Journal of Language Contact* 1.41–62.
- THOMASON, SARAH. 2008. Social and linguistic factors as predictors of contact-induced change. *Journal of Language Contact* 2.42–56.
- THOMASON, SARAH.; and TERRENCE KAUFMAN. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley, CA: Univeristy of California Press.
- THOMPSON, SANDRA A.; ROBERT E. LONGACRE.; and SHIN JA J. HWANG. 2007. Adverbial clauses. *Language typology and syntactic description: Complex constructions*, ed. by Timothy Shopen, II:237–300. Cambridge, the United Kingdom: Cambridge University Press.
- TKACHMAN, OKSANA.; and IRIT MEIR. 2018. Novel compounding and the emergence of structure in two young sign languages. *Glossa: a journal of general linguistics* 3. doi:10.5334/gjgl.632. <https://www.glossa-journal.org/article/id/5104/>.
- TOSCO, MAURO. 2021. Measuring languageness: Fact-checking and debunking a few common myths. *DIVE-IN – An International Journal on Diversity and Inclusion*. DIVE-IN – An International Journal on Diversity and Inclusion.43-51 Pages. doi:10.6092/ISSN.2785-3233/13893.
- TRUDGILL, PETER. 1974. Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography. *Language in Society* 3.215–246. doi:10.1017/S0047404500004358.
- TSAI, HUI-MING. 2017. A study of Philippine Hokkien language. Taipei, Taiwan: National Taiwan Normal University phd dissertation.
- UYTANLET, JULIET LEE. 2014. The Hybrid Tsinoys: Challenges of hybridity and homogeneity as sociocultural constructs among the Chinese in the Philippines. Wilmore, KY: Ashbury Theological Seminary ph.d. dissertation.
- VEJDEMO, SUSANNE.; and THOMAS HÖRBERG. 2016. Semantic factors predict the rate of lexical replacement in content words. *PLOS ONE* 11.
- VERHAGEN, VÉRONIQUE.; MARIA MOS.; JOOST SCHILPEROORD.; and AD BACKUS. 2020. Variation is information: Analyses of variation across items, participants, time, and methods in metalinguistic judgment data. *Linguistics* 58.37–81. doi:10.1515/ling-2018-0036.
- VERREY, NELE.; EVY WOUMANS.; DAVY VANDELANOTTE.; ARNAUD SZMALEC.; and WOUTER DUYCK. 2016. The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition* 19.181–190.
- WANG, PING. 2015. Phonological prominence and its interaction with tone in Chinese dialects. Pittsburgh, PA: University of Pittsburgh ph.d. dissertation.
- WEINREICH, URIEL.; WILLIAM LABOV.; and MARVIN I. HERZOG. 1968. Empirical foundations for a theory of language change. *Directions for historical linguistics*, ed. by Winfred P. Lehmann and Yakov Malkiel, 100. Austin: University of Texas Press.

- WICHMANN, SØREN. 2020. How to Distinguish Languages and Dialects. *Computational Linguistics* 45.823–831. doi:10.1162/coli_a_00366.
- WINFORD, DONALD. 2013. Social factors in contact languages. *Contact languages: A comprehensive guide*, ed. by Peter Bakker and Yaron Matras, 363–416. Berlin: De Gruyter Mouton.
- WU, JIANXIN. 1999. *Syntax and semantics of quantification in Chinese*. manuscript. College Park: MD: The University of Maryland, College Park.
- YIP, MOIRA. 2002. *Tone*. Cambridge, the United Kingdom: Cambridge University Press.
- YUAN, BOPING.; and ESUNA DUGAROVA. 2012. Wh-topicalization at the syntax-discourse interface in English speakers' L2 Chinese grammars. *Studies in Second Language Acquisition* 34.533–560. doi:10.1017/S0272263112000332.
- ZHANG, HANG. 2016. Dissimilation in the second language acquisition of Mandarin Chinese tones. *Second Language Research* 32.427–451.
- ZHANG, JIE. 2002. *The effects of duration and sonority on contour tone distribution: Typological survey and formal analysis*. New York: Routledge.
- ZHANG, QING. 2005. A Chinese yuppie in Beijing: Phonological variation and the construction of a new professional identity. *Language in Society* 34.431–466.
- ZHANG, YIN.; RONG JIN.; and ZHI-HUA ZHOU. 2010. Understanding bag-of-words model: A statistical framework. *International Journal of Machine Learning and Cybernetics* 1.43–52.
- ZWICKY, ARNOLD M. 1985. Clitics and particles. *Language* 61.283–305.
- ZWICKY, ARNOLD M.; and GEOFFREY K. PULLUM. 1983. Cliticization vs. inflection: English N'T. *Language* 59.502–513.