

Reclassifying ISO 639-3 [nan]: An Empirical Approach to Mutual Intelligibility and Ethnolinguistic Distinctions

Supporting documentation for the proposal put forth by the Tâi-oân Lô-má-jī Hiáp-Hōe that Min Nan Chinese [nan] be split into 11 language codes: Hainanese [hnm], Haklau [hlh], Hokkien (incl. Chawan and the Datian Frontlect) [hbl], Kelantan Peranakan Hokkien [hkl], Liongna [lnx], Longdu [zld], Luichew [luh], Namlong [znl], Samheung [zsh], Taiwanese [oan], and Teochew [tws].

1—Introduction

Around 2008, there was a push to create a Teochew Wikipedia. The project moved to the incubator stage. Around August 2008, the would-be founder(s) received a message from Joan Spanne of the ISO 639-3 Registration Authority stating that “Teochew Wikipedia currently lacks one thing: a valid ISO-639-3 language code.”¹

A request was submitted. In January 2009,² the Registration Authority responded:

The request to retire the code element for Min Nan and divide it into two distinct languages, Xiamen and Chaozhou, is rejected because it is incomplete. The Registration Authority agrees with the submitters that, based on the evidence submitted, Chaozhou and Xiamen (the latter including the group of very closely related varieties sometimes grouped under the name “Hokkien”) should be recognized with two separate code elements. However, the request does not address other Min Nan varieties at all that are also part of the denotation of “[nan] Min Nan.” Within the discussions and websites cited in support (as well as other sources), there are mentions of the Southern Min varieties of Hainan (Qiongwen) and Longyan, respectively, that are not directly addressed in the change request. The ISO 639-3 registration authority must account for the full breadth of denotation of a code element to be retired when it considers such a significant change, but it is not the work of the Registration Authority to fill in such large gaps of a request.

The Registration Authority invites the submitter to submit a new request in which

the full breadth of the Min Nan varieties are taken into consideration.³

Perhaps daunted by the requirement that they competently describe and analyze several of the most reclusive, understudied languages in the industrial world, the requesters of 2008 hesitated and were never heard from again. For their part, the Registration Authority certainly had no earthly reason to suspect the depth of the man-made rabbit hole of [nan].

In due time, Teochew Wikipedia was blocked on the grounds of there being “no ISO 639-1/3 code, and no evidence that one will be forthcoming any time soon.”⁴

In August 2019, a request was made by one Kirk Miller to split [nan].⁵ This was rejected based on lack of “evidence of specifics,” and “preponderance of scholarship.”⁶

1.1—Motivation

To be sure, this is an unusually large document written to support a seemingly simple request. However, we wanted to ensure that the amount of evidence provided was clearly sufficient for all languages involved, such that an uninvolved third party would be left with little doubt about the situation on the ground.

And also to be sure, we are not uninvolved third parties. Our motivation comes through the groups and individuals who have contributed to this project, based out of Taiwan (Formosa) where our language, Taiwanese, has faced significant political, military, and bureaucratic setbacks over the past century. We fear that despite having many millions of living native speakers, if we cannot right the ship, Taiwanese will continue its swift march to an early grave. For, despite very significant efforts made by activists in the last four decades especially, the rate of passing of older speakers continues to outpace acquisition by children and young people. As technology today intersects with every field, every industry, and practically every facet of our lives, one of the most significant barriers we face continues to be the lack of an

1 https://incubator.wikimedia.org/wiki/User_talk:Sarn-jaava#How_to_obtain_a_Teochew_ISO-639-3_language_code (last accessed Aug 2021)

2 <https://iso639-3.sil.org/request/2008-083> (last accessed Aug 2021)

3 https://iso639-3.sil.org/sites/iso639-3/files/change_requests/2008/CR_Comments_2008-083.pdf (last accessed Aug 2021)

4 https://meta.wikimedia.org/wiki/Requests_for_new_languages/Wikipedia_Teochew#The_outcome_.28January_26.2C_2009.29 (last accessed Aug 2021)

5 <https://iso639-3.sil.org/request/2019-062> (last accessed Aug 2021)

6 https://iso639-3.sil.org/sites/iso639-3/files/change_requests/2019/CR_Comments_2019-062.pdf (last accessed Aug 2021)

appropriate ISO code with which to tag our products, websites, publications, software, and works of all kinds. It is precisely the young people of today who are most important for stabilizing the language to prevent decline, and it goes without saying that, for them, technology is simply an integral fact of life. We hope that the ISO issue can be resolved forthwith, so that we may retrain our focus on the enormous mountain of work that lies before us, to educate, normalize and bring the language back up to speed with modern life.

As it stands, [nan] is confusing, disorienting, and frustrating for both the speakers of the languages that are purported to be subsumed by it, and the businesses, products, and services that are attempting to support them. And yet to an outsider, it is surely difficult to imagine that such issues have been overlooked, discounted, or disregarded by the many hundreds or thousands of scholars and researchers that investigate such matters. What we must make clear is that [nan] has often not been subjected to the rigors of linguistic scholarship that are taken for granted in linguistics departments the world over. Rather, [nan] is typically studied as a branch of Sinology—that is, the study of China and Chinese culture. While the languages that comprise [nan] are undoubtedly an important area of study for Sinologists, and while Sinology is certainly an important field of study, we believe that the ISO 639-3 Registration Authority seeks to apply the same *linguistic* criteria to the languages of all nations and cultures, and that therefore [nan] must be viewed through the lens of linguistic science rather than solely Sinology. We are not aware of a large body of scholarship that analyzes the languages of [nan] in the same way as do those scholars of African, Australian, American, European, or most other Asian languages.

Through the rest of this document, we provide significant, factual evidence to support our position that [nan] should be divided into 12 distinct languages under 11 codes. We understand that classifying languages is not a mechanical process with well-defined, technical procedures resulting in hard, objective boundaries. We also understand that we are not authoritative experts in any of these languages. Given what we have come to understand about the vast differences between them as a result of this work, we would not be surprised to find that no single linguist or research group could properly be considered authoritative across the entire field. That being said, as the 12 languages of [nan] have been bound together for a complex assortment of cultural and political reasons which the Registration Authority is understandably neither willing nor able

to sort through unassisted, we have been left with no alternative but to take up the task ourselves. We have carried out this project in good faith, and have taken the utmost care to examine all constituents of [nan] fairly and without favoritism. It is unavoidable that we are more knowledgeable about certain languages than others, but to the extent possible, we have tried to prevent this from influencing our judgments.

We begin by introducing 13 different varieties currently associated with [nan]. For each variety, we address issues related to naming, political and social status, literacy, availability or existence of mass media and intra-variety intelligibility. We then examine pair-wise mutual intelligibility and ethno-linguistic contrast between varieties. We recommend that [nan] be reapportioned into 12 languages under 11 ISO 639-3 codes, and offer suggestions on which codes would be most suitable. In [Appendix A](#), we further address issues of naming and outline the reasons supporting our code suggestions.

To curb fatigue, we hereinafter abbreviate both “mutually intelligible” and “mutual intelligibility” to “m.i.”

1.2—A Note on Proper Nouns and Romanization

Since this paper is in English, and chiefly for this reason, if there has ever been a customary English proper noun for a place or ethnolinguistic group, we will use it. We privilege current custom over antique custom, and antique custom over no custom.

What a place or group is called in English, what it is called in Mandarin, and what it is called natively are three related but distinct questions. All three questions—especially the last—are relevant, and will be dealt with in greater detail in [Appendix A](#).

We are aware that most Sino-topolectologists in the early 21st century believe that modern English (and Spanish, Indonesian, etc.) usage should mirror Mandarin usage, without regard to either past English usage or native custom. We find this imperative to be unwarranted and exceptional in the study of human language the world over.

Where translation of proper nouns is routed through a third language in linguistic study in general, this is typically incidental. There is arguably no blanket default practice; if there is one, it would be direct translation from native usage to the target language.

We suspect the silencing of both native usage and past English usage—which typically reflects past native usage—prejudices the linguistic inquiry in subtle, cumulative ways. For the possible conve-

nience of the reader, though, we also provide the equivalent in Mandarin Pinyin wherever our usage diverges from the dominant Sino-topolectologist practice.

In a case like that of Datian, we detect no customary usage in English. We default to the Mandarin-derived, Pinyin-based form.

Our approach would be unremarkable in any context outside of early 21st century Sino-topolectology. We hope it is clear that we have chosen this approach not out of ignorance or nostalgia, but rather with fairness and decency in mind.

2—Unpacking [nan]

In this section, we detail each variety currently subsumed by ISO 639-3 [nan].

2.1—Introducing Taiwanese

Taiwanese is spoken actively by about 15 million in Taiwan (Formosa) and the Pescadores, the great majority native speakers. There are probably a few million more dormant speakers, also largely native.

English Wikipedia calls Taiwanese “Taiwanese Hokkien.” The naming issue—which sprawls across English, Mandarin, and Taiwanese itself—has been contentious in recent years. The language has generally been called “Taiwanese” in English. We use this *de facto* name throughout. We address the naming controversy in [Appendix A-1](#).

Taiwanese is the main wet-market language in the lowlands on the west coast of Taiwan, and in most of the north, but another language—Mandarin—dominates in education, mass media, software, government, and white-collar, “middle-class” environments in general throughout Taiwan. Language shift to Mandarin has been under way since the 1980s. In what is a highly urbanized country, it is exceptional to hear children speaking Taiwanese in the cities today. It is not clear that Taiwanese will be viable into the 22nd century.

Under the Development of National Languages Act (2019) of the Republic of China—which has occupied Taiwan and the Pescadores since 1945—Taiwanese is now one of close to two dozen so-called *national languages*.⁷ Mandarin remains the only *official* language.

Grade school students can elect Taiwanese as a subject for an hour a week. Mandarin is the sole language of instruction in all subjects.

The Republic of China has not directly restricted Taiwanese-language broadcasting and publishing since the late 1980s. At present there is a wide variety of TV programming in Taiwanese, although there is relatively little of it. Taiwanese has a more constant presence on the radio, where audiences skew older.

*Ethnologue*⁸ implies that Taiwanese is the “*de facto* language of national identity” in Taiwan. This a complex and furiously contested half-truth. There is no Taiwanese nation yet in any case, and as Japanese-educated and illiterate older Taiwanese (Formosans) age and pass, *Taiwanese Mandarin* and the Republic of China orthographic standard (for Mandarin) have become unambiguous markers of “Republic of China on Taiwan” identity. Today it is wishful—and politically incorrect—to identify Taiwan with the Taiwanese language. However, the Taiwanese language remains powerfully, unambiguously, and exclusively associated with Taiwan.

Most Taiwanese speakers born since WWII write and read in Mandarin by default, but—for an unofficial language in East Asia—Taiwanese has an impressive body of literature. There are a handful of literary periodicals in Taiwanese. The *Church News* (*Kàu-hōe Kong-pò* 教會公報), which was Formosa’s first periodical in 1885, was forced to switch from Taiwanese to Mandarin by the Republic of China in 1969; today it is a Mandarin publication with a modest Taiwanese section.

Taiwanese has multiple orthographies, the Church-derived Latin script being the most standardized. A traditional Sino-script—conceptually akin to both Vietnamese *Hán Nôm* 漢喃 and Japanese *man’yōgana* 万葉名—has arguably been much more widespread, and underlies the common written form of place names in most of Taiwan. Mandarin-educated Taiwanese have applied the Mandarin script to Taiwanese, mostly in an ad hoc manner, but a scheme for doing this systematically was devised in the 2000s. These scripts are now often combined in a variety of ways.

We note that while Taiwanese traditionally existed in a diglossic relationship with the Classical Chinese written koine—the “Latin” or “Fusha” (Arabic) of East Asia⁹—many Taiwanese-speaking Christians, typically of indigenous extraction, attained literacy in romanized Taiwanese and did not take part even passively¹⁰ in diglossia with Clas-

7 Article 3 reads: “‘National language’ as referred to in this Act shall mean the natural languages and sign languages used by the different ethnic groups in Taiwan.” “Taiwan” in the Act is a term of art metonymically referring to the Republic of China.

8 22nd edition. Ditto hereinafter unless otherwise stated.

9 The traditional Sino-script is based on the Classical Chinese script, and to this day has always existed marginally.

10 Illiteracy was prevalent in 19th and 20th century Formosa, but Classical Chinese played (and plays) an intimate role in

sical Chinese. In other words, they were outside of the basis for the ISO 639-1 [zh] code, but they were Taiwanese speakers.

Taiwanese has seen massive dialect leveling as a by-product of the modernization and urbanization of Taiwan. With fringe exceptions, m.i. is nearly perfect within Taiwanese. The Pescadorean dialects, while highly intelligible to mainstream Taiwanese speakers, require effort initially.¹¹ The “Seaward” (*Hái-kháu* 海口, Mand. *Hǎikǒu*) dialect centered on Lokkang 鹿港 (Mand. *Lùgǎng*) may be the same way in full trim, as spoken by the elderly.

There is a uniform sense that Taiwanese, including the Pescadorean dialects, is a unit. The unit does not extend to Hokkien—not even Quemoy 金門~金門 (Mand. *Jīnmén*, sometimes “Kinmen”)¹² Hokkien, despite the last seven decades of political circumstance.¹³

Linguistically, roughly speaking, Hokkien is the closest entity to Taiwanese. We discuss the relationship between Taiwanese and Hokkien in §4.1.

2.2—Introducing Hainanese

Hainanese is spoken by several million on the island of Hainan 海南¹⁴ and many thousands more in ASEAN lands, especially Malaysia, Singapore, Thailand, and Vietnam. The Joshua Project estimates an ethnic population of 6.68 million, including 6.06 million in China.¹⁵

Hainanese is the most widely spoken of the many languages of Hainan, and the only one associated with [nan]. Hainanese was lingua franca on Hainan till the 1980s. Some older and middle-aged islanders speak it as a second or third language. It is strictly a minority, “ethnic” language off Hainan.

Language shift to Mandarin is well under way. In the capital, people under 15 are rarely able to speak Hainanese.¹⁶

mainstream Taiwanese-speaking (and Hakka-speaking) religious life; illiterate individuals were almost invariably keen to introduce literacy in Classical Chinese into the family line as soon as they had the means.

11 <https://dwhy0929.pixnet.net/blog/post/24953479> (last accessed Aug 2021; in Mandarin).

12 Administered by the Republic of China to this day.

13 Nor does Hokkien become Taiwanese when it is spoken by Hokkien-speaking immigrants from the People’s Republic of China or Malaysia. Interestingly, while post-WWII Teochew-, Fuzhou-, and Hakka-speaking refugees sometimes shifted to Taiwanese, Hokkien-speaking immigrants seem to have overwhelmingly shifted to Mandarin.

14 “Hainan” or “Hailam” via Hainanese, but the Mandarin anglicization “Hainan” seems to have prevailed in English for quite some time.

15 https://joshuaproject.net/people_groups/12053/CH (last accessed Aug 2021)

16 <http://hi.people.com.cn/GB/374508/375779/376068/index.html> (last accessed July 2021; in Mandarin); <https://www.zhihu.com/question/278258575/answer/1268601018> (last accessed July 2021; in Mandarin)

Hainanese is not official anywhere. It was probably used informally as a language of instruction on Hainan through the turn of this century—see the Luichew City school anecdote in §2.3.

Hainanese acquired a romanized missionary script in the 19th century. This was used in Christian circles,¹⁷ and seems to have fallen out of use at some point in the 20th century. We are not aware of any other vernacular literature in Hainanese, but we would not be surprised if it existed in some unconventional form on the margins of Hainanese society.

There is news in Hainanese on TV on Hainan, and on the radio in Singapore (FM 95.8) and Malaysia. There is local entertainment media in Hainanese on Hainan. Namewee—an internationally known, native Hainanese-speaking singer from Muar, Malaysia—has recorded in Hainanese.

We are not sure to what extent the dialects of the Hainanese language are strictly m.i.¹⁸ Says @yuemeigui on Reddit, in English:¹⁹

All the dialects of Hainanese are mutually comprehensible with each other but, at the same time, they are far enough apart to be classed as regional dialects rather than regional accents. My understanding (mostly on the basis of every Hainanese speaker I know being at least somewhat capable in Wenchang dialect) is that Wenchang Hainanese was the island wide standard for communicating with people “not from around these parts”. Despite populations also coming from places like Qionghai and Dongfang, Wenchang Hainanese is the Hainanese spoken today in Malaysia and parts of Singapore.

The dialect-switching suggests stiff dialect differences.²⁰ At the same time, all anecdotal evidence—and the ethnolinguistic breakdown of the Chinese in Indochina and Nusantara²¹—suggests a united Hainanese ethnolinguistic identity, regardless of intelligibility.

[zhihu.com/question/278258575/answer/1268601018](https://www.zhihu.com/question/278258575/answer/1268601018) (last accessed July 2021; in Mandarin)

17 <https://zh.wikipedia.org/wiki/海南話白話字#历史> (last accessed July 2021; in Mandarin)

18 See <https://www.zhihu.com/question/24516606> (last accessed July 2021; in Mandarin).

19 Accessible via cache: <https://webcache.googleusercontent.com/search?q=cache:sFeJyeTLaqQJ:https://www.reddit.com/user/yuemeigui/+> (last accessed July 2021)

20 Her use of past tense suggests that people nowadays have another method—probably Mandarin—for bridging dialect differences. The dialects in question most likely still exist.

21 Indonesia, Malaysia-Singapore-Brunei, and the Philippines. We use this term in its broadest sense.

Linguistically, the closest entity to Hainanese is Luichew.

2.3—Introducing Luichew

Luichew is spoken in most of Luichew²² 雷州 (Mand. *Léizhōu*), the peninsular region at the southern tip of continental China. It is also spoken several dozen kilometers up the coast in the district of Tinpak 電白 (Mand. *Diànbái*). English Wikipedia quotes 2.8 million speakers as of 2004 for “Leizhou Min.” This does not seem to have included the million or so speakers in Tinpak. Luichew may have over 4 million speakers today, nearly all within China. We are not aware of any enclaves abroad where Luichew is used in public.²³

Luichew consists of two main dialects. “Loi” (黎话)²⁴—often called simply (the equivalent of) “Luichew”—is spoken in most of Luichew and on the near side of Tinpak. “Sea Speech” (海话)²⁵ is spoken on the far side of Tinpak. The two dialects are m.i., but different enough that Sea Speech- and Loi-speaking teenagers in the county seat of Tinpak—among classmates, at least—now favor Mandarin over mixed-dialect conversation.²⁶

Luichew seems to be holding its own in the rural districts. In Tinpak, though, many families in the county seat are raising their children in Mandarin; elementary school students now favor Mandarin amongst themselves.²⁷ And language shift to Mandarin and Cantonese has been under way for at least a generation in Luichew City, the cultural heart of the Luichew-speaking region.

One netizen²⁸ recalls growing up in central Luichew City 雷州市—formerly Hoihong 海康—in

the 90s and 00s. In grade school, classes and administration were theoretically in Mandarin; the kids spoke among themselves in Cantonese. Her parents were from out of town; she never learned Luichew. Likewise for many of her classmates. The offspring of Luichew-speaking families—such as her eventual husband—often wound up not learning Luichew as well.

Through the early 00s, many older teachers were not comfortable in Mandarin, but a lot of schoolkids didn’t know Luichew. Cantonese served as *lingua franca*. During this period, the Luichew language was seen as rural and backward; Mandarin was felt to be stilted and distant. By the mid 00s, the older local teachers had been replaced with teachers from provinces to the north; students still spoke Cantonese among themselves.

Our raconteur came of age and moved to the Pearl Delta. What inspired her to reminisce was a visit to her hometown some years later. It was a familiar shock to experience the streets full of people speaking *Cantonese*.²⁹ She marvels that to this day Baidu Baike—the Chinese equivalent of Wikipedia—portrays Luichew City as a Luichew-speaking town.³⁰

Luichew is m.i. with all of itself,³¹ although—as discussed—the Sea Speech dialect is different enough from mainstream Luichew to trigger switching to Mandarin among very young people.

Along with mainstream Hainanese, Be [onb], and Vietnamese, the southernmost dialects of Luichew seem to have taken part in an imploding consonant shift several hundred years ago. M.i. within Luichew does not seem to have been affected.

Luichew’s linguistic next of kin is Hainanese.

22 This usage is current. See, for example, the Lui Chew Association of Muar, Malaysia. Also spelled “Luichiu.”

23 However, Cancao 港口—a semi-independent 18th century maritime state located in the Mekong Delta—was founded by Luichew native Mok Kiu 鄭玖 (Vietnamese *Mạc Cửu*). See https://en.wikipedia.org/wiki/Mạc_Cửu (last accessed Aug 2021).

24 See page 167 of the 雷州方言詞典 put out by the 江蘇教育出版社; “黎” is pronounced [loi] in this context.

25 Not to be confused with at least two other nearby, very distinct varieties called “Sea Speech” (海话). One is a Kra-Dai language indexed as “Jizhao language” in English Wikipedia (https://en.wikipedia.org/wiki/Jizhao_language, last accessed Aug 2021). Another, called 廉江海话, has a superstrate derived from a Cantonese-like language (<https://baike.baidu.com/item/廉江方言/14720719>, last accessed Aug 2021; in Mandarin). We have not come across more satisfying descriptions of this tongue.

26 <https://www.zhihu.com/question/278258575/answer/1554937160> (last accessed Aug 2021; in Mandarin)

27 Id.

28 See <https://www.getit01.com/p20180106726349979> (last accessed July 2021), where the story has been copied and stashed. Date, authorship and original context are lost to us, unfortunately.

29 The Pearl Delta has more transplants from other provinces; evidently it is a step ahead of Luichew in the push to not put other languages before the national language.

30 Typical, perhaps. The Sino-topolects are often conceptualized as a carpet of incrementally varying nodes, syntactically and lexically homogeneous, geographically exclusive by default and more or less static. Some less rigorous—or more intensely top-down—studies deal with “messy” multilingual locales by designating “representative topolects” (代表方言) and mentally erasing non-representative ones. The notion of Luichew City as a Luichew-speaking city seems to be an instance of such thinking.

31 <https://kknews.cc/zh-hk/news/j96ykbe.html> (last accessed Aug 2021)

2.4—Introducing Teochew

Teochew is spoken natively by roughly 10 million in China and two to five million overseas, according to Vietnamese Wikipedia.³² These numbers are plausible³³; their imprecision hints at how almost nobody anywhere is actually keeping track of how many people speak Teochew 潮州 (Mand. *Cháozhōu*).³⁴

The Teochew-speaking homeland—traditionally of the same name—is a stretch of coastal plain in the province of Guangdong 廣東 running from the border with the province of Fujian 福建 down to about half way to Hong Kong. The region is about the size of Puerto Rico and was likewise long a fountainhead of massive, sustained emigration. It is ordinary to hear Teochew spoken in Singapore, Indonesia (Riau, Medan, Pontianak), or almost anywhere in Malaysia; or in a broad belt from southern Vietnam through Cambodia into eastern and central Thailand, and down the gulf side of southern Thailand; or in the Pearl Delta megalopolis; or in the West by way of Vietnam, Cambodia, Malaysia, etc. There are non-native speakers in some ASEAN locales.

English Wikipedia calls Teochew both “Teochew dialect” and “Chaoshan Min.” As of July 2021, there is a warring article under each of the two names. The “Teochew dialect” piece is clearly the original. Apparently, since the word “Teochew” is the anglicization of the *native* name of a *pre-modern* administrative region, it runs afoul of a certain hyper-hierarchical point of view. Meanwhile, “Chaoshan” (潮汕) is the anglicization of a *Mandarin* usage coined in terms of the *current* administrative hierarchy; and “Min” (閩~閩) ballpark Teochew’s place in the parallel *linguistic* hierarchy that the Sino-topolectologists have sculpted.

The “Teochew” guard has not yielded to the would-be overlords. Hence the dueling articles on Wikipedia. For this paper, we use the natively derived, customary “Teochew.” We revisit the naming issue in [Appendix A-4](#).

Teochew is not official anywhere, although it was the *de facto* language of instruction at public schools throughout the Teochew region at least through the

90s.³⁵ Teochew-language instruction persisted in rural districts at least into the 10s.³⁶

With Cantonese having gained in prestige for much of the post-Reform era, young adults in the Teochew region are commonly fluent in Cantonese. Unlike with Luichew in Luichew City, though, there has been no shift from Teochew to Cantonese. Rather, the “clash of titans” between Cantonese and Mandarin has sheltered Teochew (in China), at least until the last few years. Language shift to Mandarin in the Teochew region may have now begun. See “爲何小孩開始不會講潮汕話 (潮州話) 了呢”³⁷ (roughly “Why are there now children that don’t speak Teochew?”). That this video was made in 2019, though, underscores how vigorous or well-situated the Teochew language is in its homeland: Hainanese, Luichew and Taiwanese were at this “unenviable” stage decades ago.

Overseas, Teochew is fading in Thailand (shift to Thai), Singapore (shift to Mandarin and English), and Johor in Malaysia (same). A heavily Vietnamized Teochew remains vigorous in pockets in southern Vietnam.³⁸ A Teochew-speaking enclave remains in Phnom Penh despite the purges of the late 20th century.

Teochew was the second most (natively) spoken language in Singapore circa 1980, at the beginning of the Speak Mandarin Campaign. The role of Teochew in Singapore has been greatly diminished. In the West—in Paris and California—heritage speakers often study Mandarin as a sort of proxy for maintenance of Teochew, which goes by the wayside: an interesting kind of “sunk language shift.”³⁹

Teochew is robust in its enclaves in Indonesia, for social and political reasons. On Batam, practically in the shadow of the Singapore skyline, Teochew (along with Hokkien) remains a go-to language for heritage speakers of all ages and socioeconomic levels.

In terms of volume and variety of media and broadcasting, Teochew in its home region is doubly overshadowed by Mandarin and Cantonese,

32 https://vi.wikipedia.org/wiki/Tiếng_Triều_Châu (last accessed July 2021; in Vietnamese)

33 Ten million is about three quarters of the current population of the region traditionally called “Teochew.” The lowlands have traditionally been Teochew-speaking; the highlands, Hakka-speaking.

34 Often romanized as “Chiu Chow” via Cantonese.

35 This could be inferred from the situation at schools in the Luichew region. We’ve also heard through a Taiwanese visitor to the Teochew region in the 90s that teachers taught in Teochew at that time.

36 <https://www.zhihu.com/question/278258575/answer/398936261> (last accessed Aug 2021; in Mandarin); <https://www.zhihu.com/question/278258575/answer/2000837125> (last accessed Aug 2021; in Mandarin)

37 <https://youtu.be/Gc1EeSB2iGI> (last accessed July 2021; in Teochew with Mandarin subtitles)

38 <https://youtu.be/VFRVfRlHTE> (see comments; last accessed Aug 2021; in Vietnamese); <https://youtu.be/bNrk45rxFvE> (last accessed Aug 2021; in Vietnamese and Teochew)

39 <https://storymaps.arcgis.com/stories/aa6abe118b-394c40a5ded25d08c27384> (last accessed Aug 2021)

although it is clearly better situated than Hainanese or Luichew.

In Singapore, there is public service broadcasting in Teochew to this day.⁴⁰ There are news broadcasts in Teochew on the radio in Singapore (FM 95.8) and Malaysia.

Teochew has vernacular literature going at least as far back as the 1400s,⁴¹ in a Sino-script that is to Teochew what *Hán Nôm* is to Vietnamese. There was a Teochew newspaper in the late 19th or early 20th century,⁴² at the dawn of modernity. Like Hainanese, Teochew acquired a romanized missionary script in the 19th century. This was used in the Church and does not seem to have taken hold beyond it.

We are not sure if either Teochew script is actively in use. The short-lived rise of Teochew Wikipedia, reviewed in §1, suggests there is a community of activists or enthusiasts who write in some form(s) of romanized Teochew.⁴³ Given the strong foundation and relatively mild sociolinguistic situation of the Teochew language, we suspect the customary Sino-script survives in marginal contexts that do not draw attention from the mainstream.⁴⁴

Indonesian Teochew speakers write informally in a makeshift, Indonesian-based (Rumi) romanized Teochew, sometimes seen online.⁴⁵

Pontianak Teochew is heavily localized; even Malaysian Teochew speakers have found it incomprehensible.⁴⁶ There are reports of non-intelligibility within China Teochew, specifically the dialect of Teoyeo 潮阳~潮陽 (Mand. *Cháoyáng*)—down the coast from Swatow—being incomprehensible to speakers of the other dialects.⁴⁷ The Teoyeo dialect comes up constantly when Teochew speakers talk about

Teochew. Most only go as far as to say it sounds harsh or takes getting used to. Intra-intelligibility is moot in any case. The Teochew linguistic identity is strong and unambiguous, although there are edge cases from a technical point of view—all of which we discuss in turn.

Linguistically, Haklau and Chawan are Teochew's next of kin. Haklau is ethnolinguistically distinct from Teochew; Chawan, arguably, is not. In China, m.i. between Teochew and Hokkien is limited; in certain ASEAN locales, though—notably Singapore—the local Teochew has Hokkienized, and the Hokkien has Teochewized, so that Teochew speakers from China can understand Singaporean Hokkien better than China Hokkien. We discuss how Teochew is related to Haklau, Chawan and Hokkien, respectively, in §4.4.

2.5—Introducing Haklau

Haklau is traditionally called “Hoklo” in English, based on the name of the language in Cantonese. “Hoklo”⁴⁸ has other meanings in English. It can also refer to Teochew, Hokkien or even Taiwanese, or all of the above in aggregate. The same is true of the Cantonese form of the word. Sources with an Indonesian nexus use the toponym “Hai Lok Hong” (海陸丰~海陸豐), based on the names of the home counties of the language—Hai Hong 海丰~海豐 (Mand. *Hǎifēng*) and Lok Hong 陸丰~陸豐 (Mand. *Lùfēng*).⁴⁹ Similarly, English Wikipedia calls this language “Haifeng dialect”; “Lufeng dialect” and “Hailufeng dialect” redirect to the same article.⁵⁰ Hai Lok Hong lies on the coast about half way between Hong Kong and Swatow 汕头~汕頭 (Mand. *Shàntóu*).

Strictly for this paper, we take the liberty of calling the language “Haklau,” based on the “Hoklo”-cognate name for the language in the language itself⁵¹—akin to adopting a people's autonym when

40 See “Why you don't need to panic buy” (<https://youtu.be/2pY9eJFkuGw>) (last accessed July 2021; in Teochew).

41 <https://kknews.cc/culture/e6b5j5r.html> (last accessed July 2021; in Mandarin)

42 <https://kknews.cc/zh-tw/culture/naxjje3.html> (last accessed July 2021; in Mandarin)

43 In Tatoeba (<https://tatoeba.org>), the hugely useful translation database, Teochew has incarnated first in the name of “Min Nan Chinese” (in the English interface). The Indonesian interface calls it “Hokkien”—an exquisite and telling translation error. There is probably nothing stopping people from adding Taiwanese or Hokkien data to the same node, but the result would serve no one. There is no real life situation where either a Teochew or Taiwanese translation will do.

44 潮州方言詞匯, a Teochew-Mandarin dictionary published by 蔡俊明 in Hong Kong in 1991, writes its Teochew in what appears to be the traditional vernacular Sino-script.

45 <https://twitter.com/scxhee/status/542223681779298304> (last accessed July 2021)

46 See <https://youtu.be/e-rY2zkCNtU> (last accessed Aug 2021). There are comments criticizing the diction while confirming that this is how Pontianak Teochew is spoken.

47 <https://www.zhihu.com/question/24516606> (last accessed July 2021; in Mandarin: “我能说作为一个讲市区汕头话的孩纸, 是去潮阳上了三年高中才能勉强听懂潮阳话的嘛”); <https://www.getit01.com/p20180122868764110> (last accessed July 2021; in Mandarin)

48 Hakka *Hók-ló*. This ethnonym has long defied consistent Sino-graphic representation. The most stable, neutral formula to date has been 學老 (Mand. *Xuélǎo*). Both graphs here are sound-borrowings, which some modern scholars find intolerable.

49 The traditional region of Hai Lok Hong is larger than the modern-day counties of Hai Hong and Lok Hong.

50 Hai Lok Hong is Haklau-speaking on the coast and Hakka-speaking in the hills. “Hailufeng” (海陸豐) in the Taiwan context tends to refer to the dialect of Taiwan *Hakka* derived from the hills of Hai Lok Hong and the vicinity. To state what may not be obvious, “Hai Lok Hong” is derived from the *Haklau* name of the region, and we use it in this paper because of the subject matter, with no prejudice to the Hakka language or culture. We do not understand Anglicized Mandarin to be a *neutral* option anyway.

51 See page 56 of 海丰方言 by 罗志海.

they call themselves “the humans” in a language of their own. We address the naming issue specifically in [Appendix A-5](#).

The regional government estimates 2.65 million speakers for Haklau, out of a population of 3.56 million in the Hai Lok Hong region.⁵² This certainly refers to native speakers, or ethnic speakers. Haklau is spoken non-natively by some of the minority Hakka of Hai Lok Hong.⁵³ Haklau is also spoken in diaspora in the Pearl Delta, although many of those speakers were probably included in the count for the Hai Lok Hong region. There was Haklau emigration to northern Sumatra in the past.⁵⁴ We do not know how much of a Haklau-speaking community there is there today.

Haklau is not official anywhere. It was used unofficially as a language of instruction⁵⁵ until probably some point in the 00s. There is local media in Haklau. We are not aware of vernacular literature in Haklau, nor a romanized missionary script.

There is language shift from Haklau to Mandarin and Cantonese. The heart of the Haklau-speaking region is only about 170 km into inner-city Shenzhen 深圳 and the gates of Hong Kong. Cantonese ability is more de rigueur here than it is in the Teochew region. Haklau-speaking migrants to the Pearl Delta may quickly and perhaps naturally stop speaking Haklau.⁵⁶ At the same time, many local couples now raise their children in Mandarin.⁵⁷

Haklau speakers are familiar with Cantonese-language TV.⁵⁸ A Taiwanese correspondent notes that in the recent past it was not uncommon to hear Taiwanese pop music in Shenzhen and Canton City 廣州 (Mand. *Guǎngzhōu*), typically when Haklau speakers were around.

Haklau speakers have traditionally formed a distinct ethnolinguistic identity. Under the current regime, the Haklau-speaking region is sometimes included in the aforementioned “Chaoshan” region, together with the land of Teochew. In social reality, Haklau speakers are clearly oriented towards Hong

Kong, Shenzhen and Canton City instead of towards the Teochew region.

Haklau is often likened to Teochew for administrative reasons, but the two are not the same language even in a functional sense. We quote from the Hong Kong-based *South China Morning Post* (“Dialect confusion delays start as judge consults chief interpreter,” 9 Oct 2001⁵⁹):

However, the trial suffered a setback when five of the accused said they had difficulty understanding an interpreter.

The part-time interpreter who was recruited to translate Cantonese into the Hoklo dialect spoke with a Chiuchow [Teochew] accent [i.e. *s/he was a Cantonese-Teochew interpreter*], which some of the accused said they could only slightly understand. The five defendants spoke Hoklo but with a Hoifung accent [i.e. *they spoke Haklau*], their lawyers told Mr Justice Thomas Gall in the Court of First Instance.

...

The morning was spent on legal arguments, but counsel for five of the accused returned in the afternoon to inform Mr Justice Gall of the interpretation problem.

The other two accused - Sze and Pang - understand Cantonese.

William Lee, for Chau, said his client could understand only 20 to 30 per cent of what the Hoklo-Cantonese interpreter had said during the morning.

William Stirling, for Yeung Yee-yim, said his client could understand, but only if the interpreter spoke slowly. John Hagon, for Yeung Yee-ping, said his client had similar difficulties.

Richard Wong Tat-wah said his client, Lam, coped by switching to the Cantonese channel as he could understand only about 20 to 30 per cent of the Hoklo interpretation. David Tolliday-Wright, for Fu, said his client was in a similar position.

It is worth pointing out that an interpreter would not have been provided—even for the wrong language—had the legal proceeding taken place in a less special administrative region of China. Throughout our discussion—and not just with regards to Haklau—it should be kept in mind that Chinese institutions do not produce or accommodate the incidental empiri-

52 http://www.shanwei.gov.cn/shanwei/swsq/swfm/rkmzyy/content/post_620223.html (last accessed July 2021; in Mandarin)

53 <http://www.hakkaonline.com/thread-57881-1-1.html> (last accessed July 2021; in Mandarin)

54 See chapter 16 of *Memories of a Nonya* by Queeny Chang. Also see: <https://yusrinlie.wordpress.com/2010/04/04/novel-acek-botak-mendorong-pembauran-bangsa/> (last accessed July 2021; in Indonesian)

55 Accessible via internet archive: <https://archive.ph/tN3rj> (last accessed August 2021; in Mandarin)

56 Id.

57 Id.

58 Id.

59 <https://www.scmp.com/article/359691/dialect-confusion-delays-start-judge-consults-chief-interpreter> (last accessed July 2021)

cal inquiries into m.i. that accumulate and are taken for granted in the West and possibly elsewhere.

M.i. within Haklau is high, to the best of our knowledge. Haklau's next of kin are Hokkien and Teochew.

2.6—Introducing Hokkien

Hokkien is spoken natively by perhaps 20 million in China and Southeast Asia. The homeland of the Hokkien language is the southern, seaward quarter of the province of Fujian, largely corresponding to the pre-modern prefectures of Chiangchew 漳州 (Mand. *Zhāngzhōu*) and Choanchew⁶⁰ 泉州 (Mand. *Quánzhōu*),⁶¹ including the island of Quemoy. Inland Zhangping 漳平 is contiguous and belonged to Chiangchew at one point; the local dialect is apparently almost identical to Chiangchew Hokkien.⁶² The eastern half of the island of Namoa 南澳 (Mand. *Nán'ào*),⁶³ administered by Chiangchew in late antiquity, is also Hokkien-speaking.

Hokkien is spoken widely in diaspora, notably up the coast of China in southeastern Zhejiang 浙江; in the Philippine port cities, especially Manila and Cebu; in Singapore and Brunei; in Indonesia (especially Medan and Riau); in Malaysia, most notably in and around Kuching, Muar, Klang and—above all—in and around Penang, from Taiping 太平 to the Thai border.⁶⁴

Hokkien is a majority or plurality language in various districts in diaspora; non-native speakers must number in the millions even today. There are traces of Hokkien-speaking communities at various other Southeast Asian ports where the language was once widely heard—Saigon, Phuket, and Yangon, to name a few.

The name “Hokkien” is customary in English. English Wikipedia also calls it this. “Fukien”—via Mandarin—is sometimes used in the Philippines, while “Amoy” has fallen out of use as a name for the Hokkien language. Sino-topolectologists prefer

“Southern Min” or “Minnan” (闽南~閩南). “Hokkien” runs afoul of hyper-hierarchical sensibilities because it was anglicized directly from the Hokkien, bypassing Mandarin; and because of its “expat” or “maritime” semantics, incongruous from the viewpoint of the administrative hierarchy. Since this paper is in English, we use “Hokkien” throughout.

Amoy Hokkien was the prestige dialect of Hokkien for a century, culminating in the Voyager missions of 1977.⁶⁵ This is the grain of truth in the note in *Ethnologue* stating that “Amoy is the prestige dialect” of [nan]. (At face value, the statement is analogous to “Madrid is the prestige dialect of Western Romance.”⁶⁶)

The Hokkien news on the radio in Singapore and Malaysia may be a last stand for the idea of Amoy as prestige dialect. Capital 95.8 FM in Singapore broadcasts the news in six “Chinese dialects” daily: Cantonese, Hakka, Hokchew, Hainanese, Teochew... and Hokkien, which they refer to as (the equivalent of) “Amoy” (厦语), in a poetic register.⁶⁷ The name is not used in vain: the broadcasts are technically delivered using Amoy readings.⁶⁸

In a general and functional sense, Amoy 厦门~厦門 (Mand. *Xiàmén*) is no longer the prestige dialect of Hokkien, neither within the home prefectures nor beyond. Hokkien does not have a prestige dialect at the moment. Hokkien-language programming in China reflects the dialect of the locale. The Hokkien news in the city of Chiangchew is broadcast using Chiangchew readings; using Amoy readings would be unthinkable, although central Chiangchew City

65 Hokkien was one of 55 languages in which greetings to extraterrestrial intelligence were recorded and stashed on board. The Amoy dialect was chosen for this—literally the high point of Amoy Hokkien's career as a standard dialect.

66 Hokkien was simply mistaken for [nan]. This trap for the unwary was set when Sino-topolectologists—sifting through existing usages—decided to call Hokkien “Southern Min” (or the Mandarin equivalent), then gave the same name to the bundle of languages that would become [nan]. The literature is unapologetically treacherous. Note that the bundle of speech varieties under [nan] have never formed an ethnolinguistic whole; even in its heyday, Amoy Hokkien never held sway over Teochew or Luichew.

67 Probably as a matter of style, Capital 95.8 FM refers to each of these languages in this context using a one-syllable identifier together with 语 (Mand. *yǔ*, “language”). Cantonese and Hakka have one-syllable identifiers that “naturally occur” in spoken language; Hokkien does not, and 厦 (Hokkien *Hā*)—one-syllable identifier for the city of Amoy—is poetically pressed into service.

68 The newscasts are in a kind of Mando-Hokkien—a kind of guided, on-the-fly Hokkien semi-translation of a Mandarin script. We describe this kind of semi-translation in greater detail in §4. Amoy readings—not at all mainstream in Singapore—are uniformly used. Beyond readings, though, Amoy usages are not adhered to. Years, for example, are pronounced using the count numbers as is the custom in Singapore, not using the book numbers as is customary in Amoy.

60 Choanchew was often called “Chinchew” in English for some time. “Chinchew” was apparently derived from an old Spanish or Portuguese name for Chiangchew. We use “Choanchew” because “Chinchew” is opaque and bewildering.

61 You would not know it from reading much of the literature, but there is considerable dialect variation within both Chiangchew and Choanchew.

62 <https://youtu.be/-TnWLhVuabc> (last accessed Aug 2021; see comments, in Mandarin); <http://www.nytuans.com/view/322825.html> (last accessed Aug 2021; in Mandarin)

63 Namoa is over the province line in Guangdong, but it was bifurcated from 1575 to 1912, with Fujian administering the eastern half of the island. Eastern Namoa is Hokkien-speaking to this day.

64 Past the border in Hat Yai, small merchants and workers in service industries regularly learn Penang Hokkien to better gain and serve visiting Malaysian customers.

is just 50 km out of Amoy. Hokkien-language TV in Malaysia routinely serves a salad of Hokkien brogues from around Malaysia, none of them Amoy-like. Cantonese- and Mandarin-language broadcasters in Malaysia, on the other hand, routinely switch to standard “international” dialects that are not spoken locally save by tourists.

In its original range on the Fujian seaboard, Hokkien can be thought of as a dialect continuum, with strained m.i. between far-apart varieties (on which more later). Taking inland and diaspora varieties into account, though, we believe it is more accurate to think of Hokkien as a network of speech varieties bound under a common, mutually acknowledged ethnolinguistic identity.

A nuanced example of the indivisibility of Hokkien is the dialect cluster spoken in northwestern Malaysia from Taiping up through Kedah to the Thai border⁶⁹; and across the water in northern Sumatra, most notably in Medan. Phuket Hokkien, which is moribund, seems to have been part of the same cluster; likewise Yangon Hokkien, for which we lack data.

This cluster was and still is centered on the island city of Penang. We call it Penang-Medan Hokkien, although English Wikipedia has separate articles for Penang Hokkien and Medan Hokkien. Medan Hokkien is a 19th-century offshoot of Penang Hokkien.⁷⁰ As Bernitone observes on Wordpress: “With ... Hakka magnates being influential figures [in Medan], the prominence of [the] Chiangchew dialect of Hokkien used in Medan reflected how important Penang was to the economy of Deli/Medan.”⁷¹ We have heard from several sources that Medan and Penang Hokkien typically cannot be told apart until a decent ways into a conversation, bringing to mind the fine distinction between Macau and Hong Kong Cantonese, or the generic forms of Canadian and U.S. English.

Intelligibility depends on register, subject matter, and the powers of the listener, but unfiltered, “basilectal” Penang-Medan Hokkien seems safely beyond the functional comprehension of a Chinese or Philippine speaker of Hokkien, at least when s/he is

not being spoken to directly. Take for example a “graphic audio drama series” posted on Facebook by Unleashed.⁷² An untutored speaker of Chinese or Philippine Hokkien would do well to understand half of the audio. There are too many loanwords from (pre-modern) Malay and (modern) Indonesian. There are differences in syntax as well.

Yet Penang-Medan Hokkien—regardless of its past⁷³—is without exception considered to be a form of Hokkien by those acquainted with it, whether speakers of the dialect, or speakers of neighboring dialects,⁷⁴ or Hokkien speakers from China.⁷⁵

To be sure, the academic distinction between loanwords and code-switching is not well understood by most people in Hokkien-speaking Nusantara, if anywhere. A Penang Hokkien speaker may in effect believe s/he is code-switching to Malay constantly when speaking Hokkien naturally; s/he might not consider a Malay loan to be a Hokkien word, even if a heritage speaker from Perth with no Malay could not speak Hokkien without it. On the other hand, spontaneous literate borrowings from Mandarin are considered to be Hokkien—even “deep” Hokkien—as a matter of course. Such beliefs are not irrelevant.

For complex reasons, many Penang-Medan Hokkien speakers today are able to filter the Malay, Indonesian, and English loans out of their Hokkien at will, giving them the ability to present a more “mainstream” Hokkien in encounters with speakers from distant lands. In turn, Chinese Hokkien speakers may come to think Penang Hokkien is just

72 <https://www.facebook.com/sutomounleashed/videos/1386484071389050> (last accessed July 2021; in Hokkien)

73 Churchman’s account of the genesis of Penang Hokkien in “The Eclectic Nature of Penang Hokkien Vocabulary, Its Historical Background and Implications for Character Writing” is enlightening. Penang Hokkien is apparently a Hokkien creole. Earlier waves of Hokkien-speaking settlers in the region had gone to the late stages of shifting to a Baba Malay with a Hokkien substrate. What remained of their Hokkien was useful in dealings with newer waves of migrants from the Hokkien homeland. In the late 19th century, though, the waves of migration became so massive that the Baba Malay speakers shifted to—and largely created—a Hokkien creole with the local Baba Malay as substrate. This Hokkien creole—what we now call Penang Hokkien—continued to converge to metropolitan Hokkien as immigration continued unabated through the mid 20th century. Thus, while the common assumption is that Penang-Medan Hokkien is a “divergent” dialect of Hokkien, it is likely that it is a product of *convergence*. See also §2.12 on Kelantan Local Hokkien.

74 Other dialects of Hokkien are spoken in Malaysia and Indonesia (and Singapore), all of which—with one nominal exception—are closer than Penang-Medan Hokkien to “metropolitan” Hokkien.

75 One correspondent, a functional Taiwanese speaker, was so bewildered on hearing Penang Hokkien for the first time that, after some time, he asked one of the gentlemen (in English) if he was speaking Hokkien. Equally bewildered, the man nodded and said he was.

69 Similar varieties are spoken in the towns on the east coast of the peninsula.

70 See Catherine Churchman (2017), “The Eclectic Nature of Penang Hokkien Vocabulary, Its Historical Background and Implications for Character Writing.”

71 See “Medan Hokkien: When Sinkeh Preserved Peranakan Language” (<https://bernitone.wordpress.com/2020/09/29/medan-hokkien-when-sinkeh-preserved-peranakan-language/>, last accessed July 2021). See also Catherine Churchman (2017), “The Eclectic Nature of Penang Hokkien Vocabulary, Its Historical Background and Implications for Character Writing.”

old-fashioned Chiangchew Hokkien with an extra lilt and “optional code-switching.”

In a real sense, such phenomena—nuances and all—support the idea of Hokkien being one language instead of many. What unites Hokkien is ultimately ethnolinguistic identity, not m.i. And the question of m.i. in Hokkien is getting to be theoretical: a 40-year-old from Amoy and a 40-year-old from Quemoy would likely converse in Mandarin if they met; if they visited Penang, they would probably speak and be spoken to mostly in Mandarin.

At the margins, it is questionable whether the speech of Chawan 詔安~詔安 (Mand. *Zhào'ān*), the speech of Lengna 龙岩~龍巖 (Mand. *Lóngyán*), the “Frontlect” of Datian 大田前路話 (Mand. *Dàtián Qián-lùhuà*), Penang-Medan Hokkien, and the exotic Kelantan Local Hokkien can be considered dialects of Hokkien. As an empirical matter, we believe Lengna and Kelantan Local Hokkien are languages in their own right; Chawan is linguistically and ethnolinguistically transitional between Hokkien and Teochew, with affinities to both. The Datian Frontlect does not seem to have had a firm identity, and appears to be regarded more and more as a strain of Hokkien. We address these varieties separately.

Hokkien is not official anywhere, although—as the local *Sprache* of the island of Quemoy—it has technically been one of the approximately two dozen “national” languages of the Republic of China since 2019⁷⁶; Mandarin remains the sole official language. Grade school students can take an hour a week of Hokkien as a subject.⁷⁷ The same is true in Amoy, although under a different regime.⁷⁸

Excluding Quemoy, schools in Hokkien-speaking parts of China would have taught in Hokkien de facto until most likely some time in the 1980s.⁷⁹

Whatever Hokkien-language instruction there had been in Malaysia was terminated by the 1980s. From the 1980s till the mid 2010s, students and teachers at Chinese schools in the region around Penang were penalized if they spoke Hokkien on site.⁸⁰

76 Article 3 of the Development of National Languages Act (2019) reads: “National language’ as referred to in this Act shall mean the natural languages and sign languages used by the different ethnic groups in Taiwan.” “Taiwan” as used here is a metonymic term of art referring to the Republic of China. The Fujianese island of Quemoy—not part of Taiwan in the conventional sense—is certainly included.

77 <https://opinion.cw.com.tw/blog/profile/266/article/9421> (last accessed Aug 2021; in Mandarin)

78 <https://www.facebook.com/uegugu/posts/3967764086584151> (last accessed Aug 2021; in Mandarin)

79 <http://hokkienese.com/?p=288> (last accessed July 2021; in Hokkien)

80 https://en.wikipedia.org/wiki/Penang_Hokkien (accessed July 2021)

Instruction in Chinese schools in the Philippines was de facto in Hokkien until around the start of the 21st century.⁸¹

There is Hokkien-language broadcast media in Fujian and in Malaysia and Singapore. Hokkien speakers in these regions are nevertheless chiefly served in the more dominant languages (Mandarin, English, and Cantonese, as the case may be). Singapore banned Hokkien-language broadcasting from the late 1970s through the mid 2010s.⁸² There is now some entertainment and public service programming in Hokkien. In Malaysia, Astro Hua Hee Dai is a full-time Hokkien entertainment channel.

Hokkien has vernacular literature going at least as far back as the 1500s.⁸³ The Manila Incunabula—literature in and on Hokkien published in Manila in the early 1600s—indicate that literacy in Hokkien (and Classical Chinese) was widespread in the Hokkien-speaking homeland at the time,⁸⁴ although this may no longer have been the case by the 19th century.

The oldest vernacular Hokkien script is a Sino-script parallel to the Vietnamese *Hán Nôm* system: a few thousand common Sinographs from Classical Chinese form the core, supplemented by a number of regional graphs. This script seems to have become moribund at some point in the 20th century, although the mass-run rhyming songbooks (歌仔冊) of the 1910s used the script. Its radical cousin has survived marginally in Taiwan.

A romanized script for Hokkien evolved in the mid 19th century,⁸⁵ first in Malacca and later by way of Amoy.⁸⁶ It started out as a missionary script but spread beyond the Church.⁸⁷ Lee Kuan Yew, who did not speak Hokkien natively—and would go on to

81 <https://www.pinoyexchange.com/discussion/450202/filipino-chinese-singles/p157> (last accessed July 2021)

82 <https://www.nytimes.com/2017/08/26/world/asia/singapore-language-hokkien-mandarin.html> (last accessed July 2021)

83 A revised 1566 edition of the *Lychee Mirror* play (重刊五色潮泉插科增入詩詞北曲勾欄荔鏡記戲文全集) is the earliest extant Hokkien text. Much of the play takes place in the city of Teochew, and the female lead’s lines are in Teochew.

84 Piet van der Loon (1966), “The Manila Incunabula and Early Hokkien Studies,” Part I

85 The Teochew and Hainanese romanizations were probably derived from the Hokkien romanization.

86 <https://en.wikipedia.org/wiki/Pe%CC%8Dh-%C5%8Dej%C4%AB> (last accessed July 2021)

87 See “A legacy of literacy: commemorating old Peh-oe-ji system” (http://www.chinadaily.com.cn/m/fujian/2017-07/28/content_30278827.htm, last accessed Aug 2021). The way Hokkien words and names are romanized in Singapore and Malaysia (in both English and Malay) suggests general influence early on from the missionary script, which in fact began its gestation in Malacca. For instance, the Hokkien word *sin-kheh* (新客, “fresh off the boat”) has been borrowed into Malay as “Sinkheh” or “Sinkhek.” The word ends with

ban Hokkien from the public sphere—had Hokkien-language campaign speeches written for him in this script along with Mandarin glosses.⁸⁸ Hokkien Wikipedia, established in 2004, would be written in the same script. Use in Church circles persisted into the 2010s in and around Amoy.⁸⁹

Hokkien speakers now overwhelmingly write and read in languages other than Hokkien.

Indonesian Hokkien speakers write Hokkien informally using the modern Indonesian Latin script.

There have been attempts to devise a Mandarin-compatible Sino-script for Hokkien⁹⁰ since the last decades of the Cold War. One scheme can be seen on the website⁹¹ and Facebook page⁹² of the Penang-centered Speak Hokkien Campaign. However, the iron presumption among the Mandarin-educated is that Hokkien should be heard and not seen. Hemmed in by such norms, even the Speak Hokkien Campaign falls back on Mandarin and English for written communications longer than about three sentences.

Hokkien is on the short end of language shift to Mandarin throughout much of its habitat. A well-attested corollary phenomenon is that fluent, grown Hokkien speakers—especially Amoy natives—often feel compelled to speak to each other in Mandarin.⁹³

Within China, Amoy natives born after 2000 can rarely speak Hokkien. Urban Choanchew City and Chinkang and Chiangchew City are about a dozen years “behind” Amoy.⁹⁴ Shift has begun in the Hokkien-speaking districts of Zhejiang as well.⁹⁵ Hokkien has its rural strongholds, but the cities have money and time on their side.

a glottal stop, not [h]. A “straight” transcription into Malay would be “*Sinkek.”

88 See the ironically titled *Keeping My Mandarin Alive: Lee Kuan Yew’s Language Learning Experience*.

89 This is according to a correspondent who was in Amoy in 2011, and another who was in Chinkang 晉江~晉江 (Mand. *jìnjiāng*) around 2010.

90 The vernacular Sino-script evolved independently and is at odds with modern Chinese notions of how a “dialect script” should behave.

91 <https://www.speakhokkien.org/hokkien> (last accessed July 2021)

92 <https://www.facebook.com/SpeakHokkienCampaign> (last accessed July 2021)

93 <http://www.antimoon.com/forum/t15665-15.htm> (last accessed July 2021; relevant portion in Mandarin); <https://ppfocus.com/o/cu1fe0593.html> (last accessed July 2021; in Mandarin); https://www.facebook.com/uegugu/posts/3967764086584151?comment_id=3967840973243129 (last accessed July 2021; in Mandarin)

94 <https://www.zhihu.com/question/278258575> (last accessed July 2021; in Mandarin)

95 <https://www.zhihu.com/question/278258575> (last accessed July 2021; in Mandarin)

On Quemoy, adult fluency in Hokkien is so faded that elementary schools struggle to find teachers for the weekly elective Hokkien lessons.⁹⁶

In Singapore—one of the capitals of the Hokkien-speaking world not so long ago—the contrived exodus from Hokkien to Mandarin is nearly complete among people under 50. Up the coast, Klang has been sheltered by the “Cantonese layer” in Kuala Lumpur and environs, but the generation under 30 is shifting. Indonesia—including Batam, in plain view of the Singapore skyline—and the region surrounding Penang seem to be exceptions to the trend.

Penang Hokkien has lost ground to Mandarin, but there is a good bit of anecdotal evidence that teenagers and young adults are going back to Hokkien in significant numbers.⁹⁷

Meanwhile, Hokkien speakers in the Philippines continue to shift to national and regional languages, as they have for four centuries.

The situation of the Hokkien language has been singular, and rewarding for ethno- or sociolinguistic connoisseurs and connectors of dots.

Forty years ago Hokkien was—in a worldly or shallow sense—one of “the great languages of the world,” spoken by far-flung millions at the most storied, most moneyed ports between Rangoon and Nagasaki. Its speakers had gained fabulous wealth through centuries of all-in, all-out enterprise. Their genes were embedded in the aristocracies of Luzon and the Visayas; their tongue was woven into the street creoles of Java. Hokkien had gone international with zero support from any nation, belying a powerful nationalistic (Chinese) paradigm that systematically overlooked and undercut its existence.

The Hokkien homeland in southeastern China is cut off from the Chinese heartland and pinned against the sea by walls of steep terrain. In pre-modern times, the region was effectively an archipelago in relation to the higher seats of power. Hokkien speakers themselves dominated the sea routes to and from the mouths of the Yangtze, or Japan, or Canton and beyond. With the long decline of China after the 1200s, and the arrival of European traders and empire builders in the 1500s, the stage was set: Hokkien-speaking society dedicated itself radically to maritime trade—and to extracting profit from the European expansion across Southeast Asia.

After the Opium War, Amoy was declared a treaty port and opened to large-scale international trade.

96 <https://opinion.cw.com.tw/blog/profile/266/article/9421> (last accessed Aug 2021; in Mandarin)

97 <https://www.zhihu.com/question/278258575/answer/861747683> (last accessed July 2021; in Mandarin)

Amoy soon eclipsed the cities of Choanchew and Chiangchew, taking their collective place as center of the Hokkien-speaking region and world. But Amoy was a center of shipping, trade, and ideas, not administration and examinations.

Amoy was more than a city. Amoy was an epoch, a moment in time. That moment ended with the Cold War. A sterile, binding isolation fell upon Amoy and its hinterland, walling the region off from even nearby Formosa. By the time Voyager 1 exited the heliosphere (in 2012), Amoy Hokkien was a wilted municipal dialect. Kids didn't learn it. Adults were ashamed to speak it.

When its time came, the accidental, exquisite greatness of Hokkien collapsed quickly, first into columns of deceptively impressive statistics, then into a sporadic, nostalgic enthusiasm for “making Hokkien great again” under the new overlords—perhaps if Taiwanese could be captured and appointed dialect-in-chief.

Millions of Hokkien speakers have hit the crossroads and gone their separate ways—some evaporating into the global Anglophone ether; many “remembering” their Neo-Chinese roots, and Mandarin; others melting into their locale like so many before them; yet others perhaps evolving into a modest Hokkien-speaking multi-tribe no more or less worthy than speakers of Bugis or Korean or Mien.

There are many reports of lack of intelligibility between various dialects of Hokkien even within southern Fujian, and even within Choanchew.⁹⁸ Speakers of Chiangchew dialects reportedly have limited comprehension of Choanchew dialects.⁹⁹ But the unity of Hokkien as a language is not questioned. Hokkien is based on ethnolinguistic identity, not m.i.

Hokkien's linguistic next of kin are Chawan (not ethnolinguistically distinct) or Haklau or Taiwanese (ethnolinguistically distinct), depending on which dialect of Hokkien we are using as our point of reference; Hokkien is *linguistically* diverse. We suspect Datian Hokkien is closer to the Datian Frontlect than it is to Chawan, Haklau, or Taiwanese. We address Hokkien's relationships with each of these varieties further ahead.

98 One netizen from Eng Chun (“Yongchun” in Mandarin anglicization)—traditionally and culturally part of Choanchew—cannot understand the dialect of the Quangang district of Choanchew (<http://www.lalasky.com/thread-24849.htm>, last accessed July 2021; in Mandarin).

99 <https://www.zhihu.com/question/278258575/answer/1043395799> (last accessed July 2021; in Mandarin); <https://www.bilibili.com/video/BV1MD4y1m7VJ/> (last accessed Aug 2021; see comments, in Mandarin)

2.7—Introducing Chawan

Coastal Chawan 詔安~詔安 (Mand. *Zhào'ān*) County sits at the southern tip of Fujian, right up against the Guangdong province line. We follow the spelling used by the Chawan Associations of Miri and Kuching (Sarawak, Malaysia). An alternate spelling is “Chao Ann.”

The inland end of Chawan County is Hakka-speaking¹⁰⁰; the districts of Sidu 四都 and Meizhou 梅洲 speak a fairly unremarkable Hokkien¹⁰¹; the rest of the county speaks something that resembles Teochew and Hokkien. The population of Chawan County is about 560,000. Of these, about a third are native Hakka speakers¹⁰²; Sidu and Meizhou total roughly 80,000 souls.¹⁰³ The rest of the county, conservatively 270,000 people, speaks what we might call “Chawan.”

Chawan was traditionally part of Chiangchew (and Fujian), but Chawan speakers apparently did not see themselves as full-fledged Hokkien speakers in the past. This is evident from the Chawan (or Chao Ann) associations in Singapore and Sarawak: they appear to be “dialect group” associations rather than county associations, which are much less prominent.¹⁰⁴

In modern Sarawak, Chawan is clearly regarded as a separate ethnolinguistic identity, apart from and parallel to the Hokkien.¹⁰⁵ In China, though, Chawan identity is weak to nonexistent. One netizen

100 The Chiangchew Hakka spoken here is very different from “mainstream” Hakka, i.e. what is meant by “Hakka” in the Hong Kong context, or in Guangdong or ASEAN lands, or in Taiwan generally. Chiangchew Hakka is not m.i. with mainstream Hakka, and was not thought of as Hakka till well into the 20th century. Chiangchew Hakka has generated considerable scholarly interest in Taiwan, where scholars call it 詔安客語 (*Cheu-ôn Hak-ngî* in mainstream Hakka; Mand. *Zhào'ān Kèyǔ*). The lowland Chawan tongue, on the other hand, is rarely studied in depth by scholars in Taiwan.

101 洪惟仁 (2011), “漳州詔安縣的語言分佈”

102 <https://zh.wikipedia.org/wiki/詔安客語#分佈> (last accessed Aug 2021; in Mandarin)

103 <https://baike.baidu.com/item/四都鎮/6753665> (last accessed Aug 2021; in Mandarin); <https://baike.baidu.com/item/梅洲鄉> (last accessed Aug 2021; in Mandarin)

104 We have heard it said in the past, in passing, that Chawan migrants to Singapore did not feel comfortable at the Hokkien associations. We would be surprised if the Hokkien associations did not have some kind of open-arms policy towards them, though.

105 <https://www.theborneopost.com/2016/10/09/abg-johari-chawans-have-contributed-much-to-sarawak/> (last accessed Aug 2021); <https://www.theborneopost.com/2015/02/28/in-honour-of-the-deities/> (last accessed Aug 2021); <https://www.newsarawaktribune.com.my/the-oldest-chinese-temple-in-sarawak/> (last accessed Aug 2021); <https://www.guaishushu1.com/authentic-chawan-seafood-and-meat-porridge-aka-cat-porridge-%E8%AF%8F%E5%AE%89-%E7%8C%AB%E4%BB%94%E7%B3%9C%EF%BC%89/> (last accessed Aug 2021)

describes his mother tongue as a kind of “Hokkien (閩南語) ... intermediate between the speech of Yunxiao 云霄~雲霄 in Chiangchew and that of Swatow” (介于漳州云霄話與汕頭話之間), pinpointing Chawan without naming it.¹⁰⁶ Singaporean Chawan speakers also refer to Chawan as a type of Hokkien at least some of the time.¹⁰⁷

Despite Chawan being in Fujian, the closest cities are Teochew City and Swatow; Chiangchew City and Amoy are nearly twice the distance. In the past, analog radio stations out of Swatow were more accessible (if Chiangchew City or Amoy stations were accessible at all).

Significantly, local theater in Chawan is Teochew theater. Chawan society must have been Teochew-oriented during some key past formative phase.

Linguistically, Chawan is clearly a transition between Teochew and Hokkien. Provocatively, one netizen says Chawan speakers can understand both Hokkien and Teochew, but Hokkien and Teochew speakers cannot understand Chawan.¹⁰⁸ This might be a general wisecrack in Chawan, and it must have grains of truth in it, although we doubt it is literally true. We discuss this in §4.6 on Hokkien-Chawan m.i.

Chawan is notable for having central vowels—like Teochew, and utterly unlike any other dialect of Hokkien spoken in Chiangchew.

It seems clear that Chawan is not a separate language; it does not merit an ISO 639-3 code. What is less clear is whether Chawan is Teochew or Hokkien. We discuss this in §4.4 and §4.6.

Chawan is fairly uniform.¹⁰⁹

2.8—Introducing the Datian Frontlect

The county of Datian 大田 lies near the heart of Fujian, northwest and just outside of Choanchew, in both the modern sense and the traditional.

According to *Sānmíngshì Fāngyánzhì* (三明市方言志, the local speech gazetteer), three speech varieties are found there. One is plain Choanchew Hokkien. Two others lie within a Bermuda triangle between three “macrolanguages within a macrolanguage”: what Chinese dialectologists call “Central Min” (currently [czo] in ISO 639-3), “Eastern Min” [cdo], and “Southern Min” [nan]. The “Backlect” (后路话, our translation) resembles the former, while the

“Frontlect” (前路话; our translation) takes after the latter; the two are not m.i.

Another account¹¹⁰ adds two more varieties to the count: Hakka, and Taoyuan (桃源). Interestingly, this account groups the Backlect (as well as Taoyuan) under Southern Min. Taoyuan is apparently a form of Frontlect,¹¹¹ but one source has it as a kind of Central Min.¹¹² We follow the gazetteer and do not deal with “Taoyuan” separately.

Confusingly, the Baidu Baike entry for Datian¹¹³ flatly states that the county speaks “Datian, a kind of Southern Min” (閩南語大田話, our translation). English Wikipedia¹¹⁴ implicitly lumps the Frontlect and the Backlect together as “Datian Min”; the ensuing description corresponds to the description of the Frontlect in the gazetteer, but the population count corresponds to that of the Frontlect and Backlect together. English Wikipedia has 250,000 “Datian Min” speakers as of 2012. The gazetteer has 120,000 Frontlect speakers as of 2001, all in Datian; and 150,000 Backlect speakers across three counties.

The confusion is telling. The county seat of Datian—center of administration and examinations—is Frontlect-speaking. Hence, the Frontlect is also known as “Datiansprache” (大田話, our casual translation), and less meticulous accounts disappear the Backlect qualitatively but not quantitatively.

This is a general problem in researching the languages of Datian. Not only is it hard to know what each commentator means by “Datiansprache,” many commentators themselves may not be fully aware of the linguistic lay of Datian.

This much seems clear: excluding Hakka, three variants with limited m.i. are spoken in Datian: the Frontlect, the Backlect, and Hokkien.

While Ethnologue does not list the Datian lects, had they done so, we think it is likely they would have listed the Frontlect under [nan].

There is some m.i. between the Frontlect and Hokkien, although such efforts have now been abandoned in favor of Mandarin.¹¹⁵ There is no functional

106 <https://www.zhihu.com/question/29296454/answer/154184811> (last accessed Aug 2021; in Mandarin)

107 <https://mysingaporenews.blogspot.com/2011/02/chinese-was-worst-affected-in-mtl.html?showComment=1297993035215#c6166931577200411194> (last accessed Aug 2021)

108 <https://www.zhihu.com/question/68764110/answer/722089862> (last accessed Aug 2021; in Mandarin)

109 洪惟仁 (2011), “漳州詔安縣的語言分佈”

110 “閩語小稱詞語法化研究——語意與語音形式的對應性” by 曹逢甫 and 劉秀雪 (citing 陳章太’s 閩語研究; http://www.ling.sinica.edu.tw/Files/LL/Documents/Journals/9.3/j2008_3_06_1557.pdf, last accessed July 2021; in Mandarin).

111 <https://zh.wikipedia.org/wiki/%E5%89%8D%E8%B7%AF%E8%AF%9D> (last accessed July 2021; in Mandarin)

112 https://twgreatdaily.com/GfGvIm4BMH2_cNUGr_D9.html (last accessed July 2021; in Mandarin)

113 <https://baike.baidu.com/item/大田县/3450481> (last accessed July 2021; in Mandarin)

114 https://en.wikipedia.org/wiki/Datian_Min (last accessed July 2021)

115 <https://www.zhihu.com/question/278258575/answer/721203163> (last accessed July 2021)

intelligibility between the Frontlect and the Backlect.¹¹⁶

There has been massive language shift to Mandarin in recent times. The Datian Frontlect has its toes in the grave.

The Frontlect is not officially recognized, and has never had a written form as far as we know. Nor is there TV broadcasting or programming in the language, to the best of our knowledge. We are not sure about radio.

Mass media in the Frontlect barely exists, or doesn't, but in 2019 “*Dàtián Hòushēngzǎi*”¹¹⁷ (大田後生仔~大田后生仔, “DTHSZ”) happened. This began as a Lín Qǐdé (林启得) folk rock song sung partly in Mandarin and partly in the Datian Frontlect. Apparently, it became a huge hit at the national level in China, with 8.7 million views on Youtube¹¹⁸ (banned in China). One of the remakes has even more views. This song and its remakes put Datian on the map.¹¹⁹

In an ultra-lucid moment, the (Mandarin) lyrics to DTHSZ cite the inability of people born after 2000 to speak the vernacular languages.

We note that DTHSZ is consistently referred to on the Mandarin internet as a *Mǐnnán* (闽南语, “Southern Min”) song. While scholars use “Minnan” to refer to both Hokkien and the greater bundle that is [nan], only the former meaning has currency in mainstream usage and pop culture. In other words, DTHSZ is thought of as a “Hokkien” song.¹²⁰

On one hand, locally knowledgeable people list the Datian Frontlect separately from Hokkien. This implies that the Frontlect is *not* Hokkien. Yet there seems to be an implicit consensus that the Frontlect is Hokkien in a way that Teochew is not. Some Datian locals also identify as *Mǐnnánrén* 闽南人~闽南人,¹²¹ grouping themselves with the people of Chiangchew and Choanchew.

As we will see, “Hokkien” is fundamentally an ethnolinguistic concept. M.i. is not required. It is reasonable to consider the Datian Frontlect a dialect of Hokkien despite the “linguistic” distance.

116 Id.

117 In Mandarin. We're not sure how this title would be romanized in the Datian Frontlect.

118 <https://youtu.be/JpvrX-Nkz0Y> (last accessed Aug 2021)

119 The song dwarfs its vernacular on the internet. Searching for the language leads inexorably to the song. Or check out 后生仔 (福建) 实业有限公司 (<http://houshengzai.com>), for example—a Datian-based big data venture capital (among other specialties) firm founded in 2020.

120 We find that Chinese netizens consistently challenge anything they feel is a misnomer involving territory. Evidently Datianites and “metropolitan” Hokkien speakers alike have no problem with the idea of DTHSZ as a Hokkien song.

121 <https://www.zhihu.com/question/35923258/answer/74670453> (last accessed July 2021)

Two “special” issues should be borne in mind in this context.

First, as language shift to Mandarin broadens and deepens, the question of whether the Datian Frontlect is “a language,” or a dialect of Hokkien, becomes less and less empirical and more academic.

For reform thinkers 12 decades ago in China, a monolingual sort of modernity was a pipe dream. Today it is an emerging reality. As function leaches away from “the dialects,” it becomes more important to acknowledge the unique socio-political context of each variant.

The second issue is that, aside from its Hokkien- and Hakka-speaking minorities, Datianite society is ethnolinguistically formless. In this, Datian is atypical for regions where [nan]-associated languages are spoken, but typical for China. In China, identity lies with the county, the province and—in modern times—China. County, provincial, and national identity stand in when ethnolinguistic questions are posed; inquiries of “What language(s) do they speak?” are deflected with bureaucratic answers along the lines of “Well, here's the kind of Chinese speaker they are: ...” The question of whether a vernacular Sprache is a language or a dialect becomes rather unwanted. Any empirical approach would have to be copied and pasted across dozens of provinces and thousands of counties in an instant: unspeakably tedious! Hence the ISO 639-1 mono-code [zh], and later ISO 639-3 [yue] and [hsn] and [nan]—top-down, theoretical shortcuts, just numerous enough to seem like a plausible reflection of reality.

Seen from the bottom up, the matter is messier and simpler at the same time. We believe the empirical inquiry should proceed, as it has for the languages of Africa, the Americas, and most of Eurasia. The work we are doing here may leave much to be desired, but it should be engaged and surpassed on the empirical level, from the bottom up, rigorously, without any prejudice for or against putative pre-existing conclusions.

We are not certain, but m.i. may be strained even between certain varieties of the Datian Frontlect. Some accounts list “Taoyuan” separately from the Frontlect. This suggests that intelligibility is a bar short in at least one direction between that dialect and mainstream Frontlect.

The Datian Frontlect's linguistic next of kin is probably either Hokkien or Lengna. We suspect a Frontlect speaker and a Datian Hokkien speaker would have the best shot at fruitful communication, but in practice at least the Hokkien speaker will tend to have had exposure to the other language, unless s/he is but a wee tot.

2.9—Introducing Lengna

Lengna¹²² 龙岩~龍岩~龍巖 (Mand. *Lóngyán*) is spoken upriver from Chiangchew in the district of Xinluo 新羅~新羅, which used to be the county of Lengna. English Wikipedia calls Lengna “Longyan dialect.”

Lengna speakers have never been counted, to the best of our knowledge. The population of Xinluo District is 840,000.¹²³ The population of Lengna speakers would be much lower at this point.

As with Datian Frontlect and Amoy Hokkien, people born after 2000 do not speak or learn Lengna.¹²⁴ Young adults do not generally speak the language, although some have the ability. Older and middle-aged people speak Lengna and Mandarin in roughly equal measure; more Lengna is spoken in out-of-the-way corners of the district.¹²⁵ Transplants from neighboring Hakka-speaking districts do not learn Lengna, and they are many.¹²⁶

Lengna is not an official language. Local schools would have switched to teaching in Mandarin by 2000, and probably much earlier.

There is some TV programming in Lengna, including newscasts. Lengna has never been written, as far as we are aware.

We are not aware of non-m.i. varieties within the Lengna language.

There is some m.i. between Lengna and Hokkien, but Lengna is not considered a kind of Hokkien.¹²⁷ We examine the relationship between Lengna and Hokkien in §4.6.

Lengna’s linguistic next of kin is probably either Hokkien or the Datian Frontlect.

2.10—Introducing Lungtu & Namlong

Lungtu 隆都 (Mand. *Lóngdū*) and Namlong 南朗~南壟~南萌 (Mand. *Nánlǎng*) are spoken near the Cantonese city of Zhongshan 中山 (formerly Heungshan 香山, via Cantonese) in the districts of the same respective names, not far from Macau. We follow the spellings initially used by Nicholas Bodman when he was surveying these languages.¹²⁸

122 This is customary usage in English. See Gerald Francis De Jong, *The Reformed Church in China, 1842-1951*.

123 http://www.fjxinluo.gov.cn/zjxl/index_7291.htm (last accessed Aug 2021)

124 <https://www.douyin.com/video/6977174901190855973> (last accessed July 2021)

125 <https://www.zhihu.com/question/278258575/answer/416099129> (last accessed July 2021)

126 Id.

127 <https://www.zhihu.com/question/35923258/answer/351813173> (last accessed July 2021)

128 See Nicholas Bodman (1982), “The Namlong Dialect, a Northeastern Min Outlier in Zhongshan Xian and the Influence of Cantonese on Its Lexicon and Phonology,” in particular the bilingual map attached to the paper.

“Longdu” is named in Ethnologue as a “dialect” of [nan]. English Wikipedia lists “Longdu dialect,” “Nanlang dialect,” and “Sanxiang dialect” (Bodman’s “Samheung”) as dialects of “Zhongshan Min,” which it places within “Southern Min,” i.e. [nan]. We will discuss Samheung in the next section.

It was estimated that “Zhongshan Min” had about 150,000 speakers when the population of Greater Zhongshan was 1.2 million.¹²⁹ The population of Greater Zhongshan has ballooned to 4.4 million,¹³⁰ but it is doubtful that Lungtu, Namlong and Samheung have gained speakers accordingly. We suspect the total number of speakers of the three languages is well under 150,000 at this point.

Nicholas Bodman studied Lungtu and Namlong extensively in the 1970s and 80s. He concluded that the two were “Northeastern Min”—or “Eastern Min, i.e. ISO 639-3 [cdo], in modern terms.¹³¹ Apparently Jerry Norman also classified Namlong with Eastern Min in the 1970s.¹³²

It is clear from reading Bodman’s arguments and data that Lungtu and Namlong do not resemble Taiwanese, Hainanese, Teochew, Hokkien, Lengna, etc. the way those languages resemble each other.¹³³ Nor does the evidence indicate that Lungtu and Namlong are descended from [nan]-type tongues that [cdo]-ized. In other words, the placement of Lungtu and Namlong within Southern Min was erroneous, even under Southern Min’s own top-down terms.¹³⁴

We will not comment on whether Lungtu and Namlong should be absorbed into [cdo]. It is implied in all studies of these languages that they are not m.i. with Fuzhou 福州 or any other [cdo]-associated language.

129 高然, “中山閩語的聲調與閩、粵語聲調的關係,” found in 丁邦新 (ed., 2002), *閩語研究及其與周邊方言的關係...*

130 http://stats.zs.gov.cn/zwgk/tjxx/tjnj/content/post_1937766.html (last accessed Aug 2021)

131 Nicholas Bodman (1982), “The Namlong Dialect, a Northeastern Min Outlier in Zhongshan Xian and the Influence of Cantonese on Its Lexicon and Phonology”

132 Douglas D. L. Chong (2010), “Hawai’i’s Nam Long: Their Background and Identity as a Zhongshan Subgroup”

133 That is, if the reader has a working knowledge of one of those languages, as well as IPA and regular sound change. Otherwise there will be no clarity aside from Bodman’s own conclusions, which are certainly not to be ignored.

134 Some critical mass of “macro” scholars who lacked access to the relevant “micro” literature—and this would have been in the early days of the internet—may have concluded that since Teochew, Haklau and Luichew are “Southern Min,” any “Min” topolects found in the space between them would have to be Southern Min as well. Alternatively, they may have reviewed data for Samheung, which is bona fide Southern Min-looking, and assumed that Lungtu and Namlong could not be too different.

It is beyond dispute that Lungtu and Namlong are not m.i. with Luichew, Haklau, Teochew, etc., nor ethnolinguistically affiliated with the same except on the level of “Chinese.”

As for whether Lungtu and Namlong are m.i., Bodman wrote, “In an experiment conducted in Honolulu, I tried to see how well a speaker of [Longdu] and [Namlong] actually could understand each other, neither of them having previously been exposed to the other dialect. The understanding was quite minimal...”¹³⁵ Also: “The lexicons of the two dialects ... differ considerably.”¹³⁶

By all accounts, Samheung is much more different to Lungtu and Namlong than either is to each other. Loosely corroborating Bodman, Douglas Chong reported that:

[t]he three Min speech groups share traits that make their village subdialects mutually intelligible to some degree. Based on my fieldwork among informants in the three separate Min areas, I estimate verbal communication between a Nam Long and a Loong Doo native, each speaking his own vernacular, to be about 60 percent mutually comprehensible even if neither party has previously had much exposure to the other’s dialect. However, when either a Nam Long or a Loong Doo attempts to communicate with a Sam Heung without much exposure beforehand, the comprehension rate might be only 30 to 40 percent. Besides, the speakers may suffer rather intense frustration, for the sounds and tones they both use have familiar “Fujian” qualities, yet their phonemes and syntax do not correspond in exact patterns of speech.¹³⁷

Chong adds that “[t]he Nam Long people ... could, with practice, communicate intelligibly with the other Min speakers from Zhongshan, the Loong Doo and the Sam Heung.”¹³⁸

Namlong and Lungtu speakers have been bilingual in Cantonese for many generations. Thanks to self-segregation in settlement patterns and marriage matters, though, language maintenance has long been the rule. According to Chong, “[m]odern education ... has not led the Nam Long natives to forego their Min speech and switch totally to Cantonese

or the Chinese national language”¹³⁹—although in diaspora, in Hawai’i, the ancestral languages have naturally not been maintained.¹⁴⁰

Lungtu and Namlong have no official status. There is no mass media in these languages, nor literature.

We suspect large-scale language shift to Cantonese has begun in the home districts of Lungtu and Namlong, and in Samheung. We have not found the chatty first-hand accounts of local language use that we would expect to find on the Chinese internet. Instead, we find people commenting in Mandarin and Cantonese that their *parents and elders* speak Lungtu, etc.¹⁴¹ Lungtu, Namlong and Samheung seem to have petered out just shy of the internet generations.

To the best of our knowledge, there is no lack of m.i. internally within Lungtu and Namlong, respectively.

Lungtu’s linguistic next of kin is Namlong, and vice versa.

2.11—Introducing Samheung

This section continues the last section, on Lungtu and Namlong. We recommend reading that section before this one.

Samheung 三乡~三鄉 (Mand. *Sānxiāng*) is spoken in the district of the same name, toward the southern edge of (greater) Zhongshan, less than 20 km shy of Macau.

Unlike Lungtu and Namlong, Samheung looks “[nan]-ish” on paper, in spite of the deep, dense, ever-present layer of borrowings from Cantonese. To the naked eye, Cantonese layer aside, the greatest resemblance seems to be to Luichew, although lack of m.i. is implied.

There is some, limited m.i. with Lungtu and Namlong,¹⁴² most likely made possible by the thick layer of Cantonese that all three languages have in common. In our time, a Samheung speaker and a Lungtu or Namlong speaker would no doubt converse in Cantonese.

Samheung is not official. There is no mass media in Samheung, and the language has never been committed to writing.

As discussed in the last section, language shift to Cantonese may have been in effect for some time. We would not be surprised if heritage speakers under 50 were largely unable to speak Samheung.

135 Nicholas Bodman (1982), “The Namlong Dialect, a North-eastern Min Outlier in Zhongshan Xian and the Influence of Cantonese on Its Lexicon and Phonology”

136 Id.

137 Douglas D. L. Chong (2010), “Hawai’i’s Nam Long: Their Background and Identity as a Zhongshan Subgroup”

138 Id.

139 Id.

140 Id.

141 <https://www.facebook.com/CantoneseMuseum/photos/a.672316452778570/1347339775276231> (last accessed Aug 2021; in Cantonese and Mandarin)

142 See section §2.10.

M.i. is “almost perfect” within Samheung.¹⁴³

Samheung’s linguistic next of kin—none particularly close—would be Lungtu and Namlong or perhaps Luichew, depending on the criteria desired.

2.12—Introducing Kelantan Local Hokkien

Kelantan Local Hokkien is a mixed language blending Hokkien, Kelantan-Pattani Siamese, Kelantan Malay, and—implicitly—Standard Malay,¹⁴⁴ spoken by approximately 20,000¹⁴⁵ *Cina Kampung* in Kelantan, Malaysia. It is spoken non-natively in addition by some of the Kelantan Siamese.

English Wikipedia does not have an article on this language. Nor does Ethnologue specifically mention it.

The *Cina Kampung*—“Country Chinese”—are descended from a mix of pre-modern Chinese settler farmers and Pattani- and Kelantan-based Siamese.¹⁴⁶ They are nearly all Theravada Buddhists, although they maintain Chinese ancestor worship.¹⁴⁷

Locally, the mixed language of the *Cina Kampung* is referred to using a wide range of intensely context-based wording. On the internet, “Hokkien Kelate” (Kelantan Malay), “Hokkien Kelantan” (Standard Malay),¹⁴⁸ and “Kelantan Hokkien” (English) typically refer to this language. There are two kinds of Hokkien spoken in Kelantan, though. A different, more mainstream kind of Hokkien, similar to and m.i. with Penang Hokkien, is spoken in and around Kota Bharu.¹⁴⁹ Thus, for academic purposes, Teo Kok Seong—*Cina Kampung* himself—coined “Kelantan Peranakan Hokkien.” For this paper only, we shorten this to “Kelantan Local Hokkien” and sometimes “Local Hokkien.”

Modern *Cina Kampung* typically speak Local Hokkien,¹⁵⁰ Kelantan Malay, and Standard Malay; many speak Kelantan Siamese¹⁵¹ and English, and the children now learn Mandarin at school.¹⁵² It is striking that while the Country Chinese are uniformly and effortlessly polyglot, even taking Mandarin in stride, they do not learn mainstream Hokkien.¹⁵³ At least as of the 1990s, they preferred to deal with the local mainstream Hokkien speakers in *Kelantan Malay*.¹⁵⁴

The gulf between Local Hokkien and mainstream Hokkien is vast. For example, mainstream Hokkien *lán* means “we (inclusive),” and another word is used for “we (exclusive);” but Local Hokkien *lan* (tone not marked herein, although Local Hokkien is tonal) covers both inclusive and exclusive “we,” following the Siamese pattern. In mainstream Hokkien, the modifier generally precedes the head, with certain vestigial exceptions; in the Local, the modifier generally follows the head. The difference between Local Hokkien and the mainstream is so viscerally obvious that Malay speakers—outside of Kelantan, we figure—often take Kelantan Local Hokkien to be some kind of Malay.¹⁵⁵

Kelantan Local Hokkien has no official status. It does not have mass media, nor a discrete written form. In marked contrast to most—including Malaysian—speakers of all other [nan]-associated languages, Local Hokkien speakers’ go-to language for writing and reading is not Mandarin but *Standard Malay*; this includes speakers who attended *Mandarin-medium schools*.¹⁵⁶

In Kelantan, speakers of mainstream Hokkien hold the Local Hokkien to be impure and inferior.

143 Nicholas Bodman (1988), “Two Divergent Southern Min Dialects of the Sanxiang district, Zhongshan, Guangdong”

144 Teo Kok Seong (1993), “A Sociolinguistic Description of the Peranakan Chinese of Kelantan, Malaysia”

145 Pue Giok Hun, Ong Puay Liu and Loo Hong Chuang (2019), “Kelantan Peranakan Chinese Language and Marker of Group Identity”

146 The *Cina Kampung* identify as Kelantanese first and *Cina Kampung* second. They see themselves as related to, but not of, *orang Cina* (“the Chinese”). They consider it an insult to be called (the equivalent of) plain “Chinese” in Malay or Siamese, although in Local Hokkien they also call themselves the equivalent of Hokkien *Tng lâng* 唐人, which translates into English as “Chinese people.” See Pue, Ong and Loo (2019) and Teo (1993).

147 Pue, Ong and Loo (2019); Teo (1993)

148 Kelantan is called “Kelate” in Kelantan Malay. Local Hokkien may have borrowed “Kelantan” instead of “Kelate.” Pue, Ong and Loo (2019) have *lang Kelantan* (tones not marked) rather than **lang Kelate* in their data as the Local Hokkien word for “Kelantanese.”

149 Pue 2009; Pue, Ong and Loo (2019). We will comment later regarding Kelantan Local Hokkien intelligibility to Penang Hokkien speakers, or lack thereof.

150 *Cina Kampung* in certain villages speak mostly Kelantan Malay or Kelantan Siamese at home, switching to Local Hokkien when dealing with *Cina Kampung* from other villages (Teo 1993).

151 In one village, the Siamese speak a Standard Thai-like dialect of Siamese, which serves as lingua franca (https://en.wikipedia.org/wiki/Kampung_Balai, last accessed Aug 2021).

152 This was the case as of 2019 (see Pue, Ong and Loo), but not, apparently, 10 years earlier. See Pue Giok Hun (2009), “On Being Peranakan Chinese of Kelantan: Embodiment and Mistaken Ethnic Identity.”

153 As with all things *Hokkien Kelate*, there is a mind-boggling technical exception: at least as of the 1990s, the *Cina Kampung* communities farthest upriver—founded by a homogeneous group of Hokkienese settlers, then isolated for decades—spoke a very “pure” Hokkien among themselves, only acquiring Local Hokkien (and Kelantan Malay) through intermarriage and interactions with the world downriver (Teo 1993). Teo Kok Seong counts their in-group dialect as a form of Local Hokkien.

154 We have to wonder if Mandarin has been pressed into this line of duty in recent years.

155 Pue, Ong and Loo (2019).

156 Id.

The *Cina Kampung* are aware of this,¹⁵⁷ yet transmission of Local Hokkien is stable. While some *Cina Kampung* speak Siamese or Malay at home, even they continue to use and pass down Local Hokkien.

Hokkien and Kelantan Local Hokkien are not explicitly distinct languages. But this is implied with abundant clarity. Lack of m.i. is no bar, since Hokkien is an ethnolinguistic concept. But the Kelantan Local clearly lies outside that unit, ethnolinguistically.

Others among us, weaned on tables of “colloquial and literary readings” indexed by county, may feel that Kelantan Local Hokkien lies beyond the galaxy of [nan] and should be at most a footnote in this discussion.¹⁵⁸ We would remind them that the Kelantan Local likely shares innovations with Penang Hokkien,¹⁵⁹ which has won recent Sino-logical acceptance as a specimen of “Zhangzhou [Chiangchew] Southern Min.” Even the common ancestor of Hokkien and Teochew—assuming there was one—was likely a creole on the frontier of China in its time.¹⁶⁰

All varieties of Kelantan Local Hokkien are m.i.¹⁶¹ Linguistically, Local Hokkien has no close kin.

3—Mutual Intelligibility & Ethnolinguistic Contrast

In sections §3.1 through §3.10, we review several sets of closely related languages and dialects. We note m.i. and ethnolinguistic unity or lack thereof, as well as the existing pattern of ISO 639-3 code assignments.

In section §4, we examine the relationships between the various [nan]-associated varieties and their linguistic or ethnolinguistic next of kin. We review evidence of m.i. and ethnolinguistic unity or lack thereof, and argue for or against a separate ISO 639-3 code for the variety based on global precedent.

3.1—Arabic

The spoken Arabic vernaculars are m.i.—or not—to varying degrees, depending in no small measure on individual factors. The North African vernaculars, in particular, are difficult for Arabic speakers from Egypt and points east to comprehend. Dana Hooshmand describes the degree of m.i. between Tunisian Darija [aeb] and the Egyptian Masri [arz], and between Egyptian and the Lebanese (North Levantine) vernacular [apc] with refreshing nuance and finesse:

If you take two educated adults from Cairo and Tunisia, they will be able to understand each other by making an effort. The Egyptian will speak slowly and articulate, and the Tunisian will, being familiar with Egyptian dialect from the movies and TV coming out of Egypt constantly, know what’s going on. The Tunisian will then in turn consciously not code-switch to French, use any Egyptian words he can think of and speak slowly and more clearly.

If you, however, take a 10-year old kid off the streets of Cairo and a similar kid from Tunisia, they wouldn’t understand each other. They wouldn’t be able to simplify their vocabulary and grammar enough to make them more internationally accessible. But if they simplify to basics, like “milk? cookie? sweets? toys?” they will probably figure out that they have something in common, even if those words are different.

However, if you take two adults or children from Cairo and one from Beirut, they’ll be able to understand each other without making significant modifications to the way they speak. It’ll be weird pronunciation and vocabulary, but it won’t be totally alien. This is what we found speaking Egyptian Arabic in Palestine or Jordan. People just spoke back in their dialect, and it all went smoothly.¹⁶²

Thus Egyptian and Lebanese are functionally m.i.; Egyptian and Tunisian arguably are not. All three variants have ISO 639-3 codes, Egyptian and Lebanese evidently on the basis of the ethnolinguistic distinction, which dovetails with the intervening national boundaries and differences in national identity.

157 Teo (1993)

158 Thus Cantonized Samheung and the Siamized, Malay-fed Kelantan Local meet very different fates in their minds.

159 Take for example the [w] initial on the first-person singular pronoun. Local Hokkien’s Pattani-Kelantan upbringing may even gut the Teochew alibi sometimes given for Penang Hokkien.

160 See 潘悟雲 (2009), “吴语形成的历史背景——兼论汉语南部方言的形成模式。”

161 Except for the in-group dialect of the upriver enclave, as mentioned several footnotes back.

162 <https://discoverdiscomfort.com/arabic-dialects-maghrebi-egyptian-levantine-gulf-hejazi-msa> (last accessed July 2021)

3.2—Malagasy

The Malagasy “dialects” are often said to be m.i., perhaps with the exception of Antandroy. A correspondent of ours heard something along these lines more than once when in Madagascar. It could be that m.i. within Malagasy is systematically over-stated, for reasons having to do with what Alphonse Tsimilaza refers to as a fear of “tribalism” that “haunts the Malagasy intellectuals.”¹⁶³

We suspect there is considerable truth to the claims that most Malagasy dialects are m.i. We also speculate that the culture in Madagascar may be relatively conducive to fuzzy conversations that require repetition and slowing down, which is exactly what allows discourse to blossom between speakers of partially intelligible varieties. But this is speculation. We quote one Mr. Velontsoa, who answered thus when asked whether “the various Malagasy dialects [are] mutually intelligible”:

Being originally from Madagascar, I can answer your question. Various Malagasy dialects are mutually intelligible even those from the North (Antakarana) & from the South (Toliara). During secondary school we learn different dialects rules. Malagasy dialects (18) are mostly divided in 3 families (Eastern, Western & Southern families). At University we meet people from different regions of the island, and the unspoken rule is that you speak one another your dialect as long as [it’s] in the same “family”, otherwise you use the “official Malagasy” (80% of the capital’s dialect plus 20% of other dialects). Me for example as Betsimisaraka (from Tamatave / Toamasina), I speak my dialect to people from Diégo Suarez / Antsirana & Majunga / Mahajanga & they speak their dialect with me. However, with people from Tananarive / Antananarivo, Fianarantsoa & Tuléar / Toliara we speak the «official Malagasy».¹⁶⁴

Clearly, the speech of Tamatave is functionally m.i. with that of Diego Suarez, and also with that of Majunga. These are seaport towns in northern Madagascar. Antakarana [xmv] is spoken in Diego, and Sakalava [skg] in Majunga. A Betsimisaraka

tongue is spoken in Tamatave; going by Ethnologue, it should be Northern Betsimisaraka [bmm], which is spoken on the “[e]ast coast from Mahanoro north.”¹⁶⁵

Without addressing the m.i. of varieties spoken to the south, we can be confident that there is functional m.i. between Northern Betsimisaraka and Antakarana, and between Northern Betsimisaraka and Sakalava. Even university students—fluent speakers for sure of a mutually shared national tongue—choose to converse in their native varieties when among speakers of neighboring northern varieties. Yet each variety has its own ISO 639-3 code, evidently on the basis of longstanding ethnolinguistic distinctions.

3.3—Akan

Fante [fat] and Twi [twi] are two of several m.i.¹⁶⁶ Akan varieties which nevertheless carry separate ISO 639-3 codes on the basis of long-running orthographic independence and distinct literary traditions.

Akan writing and literature began with missionary writings in the 17th and 18th centuries.¹⁶⁷ A unified Akan written standard was created in the 1980s.

3.4—Bengali & Assamese

Whether these two are m.i. depends on the dialects involved. Eastern (i.e. Bangladeshi) Bengali [ben] speakers can understand Assamese [asm] with ease—indeed, more easily and effectively than certain dialects of Bengali.¹⁶⁸ Western Bengali speakers have a tougher time understanding Assamese.¹⁶⁹ Whether an Assamese speaker can understand Bengali is also contingent on his or her dialect or place of origin.¹⁷⁰ Assamese speakers seem to understand Bengali

163 See Leoni Bouwer (1998), “A Sociolinguistic Expedition into Madagascar: An Annotated Bibliography,” in *Language Matters: Studies in the Languages of Africa*, 29:1, 256–285.

164 <https://www.quora.com/Are-the-various-Malagasy-dialects-mutually-intelligible-If-so-why-is-French-still-the-official-language-of-the-government-when-there-really-isnt-a-need-for-a-neutral-lingua-franca/answer/Anthoniot-Velontsoa> (last accessed Aug 2021)

165 There is also a Southern Betsimisaraka [bzc].

166 <https://www.amesall.rutgers.edu/languages/our-languages> (last accessed Aug 2021)

167 <https://omniglot.com/writing/akan.htm> (last accessed Aug 2021)

168 See <https://www.quora.com/Can-Bengali-speakers-understand-Assamese/answer/Azwad-Ahsan> (last accessed Aug 2021); <https://www.quora.com/Can-Bengali-speakers-understand-Assamese/answer/Ritinkar-Dasbhaumik> (last accessed Aug 2021); <https://www.quora.com/Can-Bengali-speakers-understand-Assamese/answer/Touhidul-Haque-2> (last accessed Aug 2021).

169 Id.

170 https://www.reddit.com/r/india/comments/aokmd8/askindia_how_mutually_intelligible_are_indian/eg2d0an/ (last accessed Aug 2021)

better than Bengali speakers understand Assamese.¹⁷¹ Exposure is a key factor.¹⁷²

Bengali and Assamese (and Odia) form a dialect continuum, with certain (if not all) dialects of Bengali being closer to Assamese than to certain other dialects of Bengali. But, in keeping with the ethnolinguistic reality, there is one ISO 639-3 code for Bengali, and one for Assamese.

3.5—Malay-Indonesian

ISO 639-3 places 36 linguistic entities under the Malay [msa] macrolanguage.¹⁷³ These include Standard Malay [zsm] and Indonesian [ind], as well as Jambi Malay [jax], Minangkabau [min], Kedah Malay [meo], Negeri Sembilan Malay [zmi], Sabah Malay [msi],¹⁷⁴ Yawi (Pattani Malay) [mfa], and Malay [zlm]. Plain “Malay” seems to be a residual catch-all for otherwise unlisted non-standard Malay varieties, such as Kelantan Malay (baso Kelate) or Pontianak Malay. Malay [zlm] and Standard Malay [zsm] could be seen as colloquial and formal registers of one language,¹⁷⁵ although Malay includes varieties spoken in Indonesia that would interact with colloquial and formal Indonesian instead of formal Malay. Ethnologue does not bifurcate Indonesian into colloquial and formal registers.

The nine ISO 639-3 entities we have listed represent a collection of closely related varieties, functionally m.i. to a large extent. Minangkabau, for instance, is largely m.i. with Indonesian.¹⁷⁶ And the Ethnologue notes, under Malay [zlm]: “Negeri Sembilan Malay [zmi], Jakun [jak], Orang Kanaq [orn], Orang Seletar [ors], Temuan [tmw], Sabah Malay [msi], and Brunei [kxd] are so closely related that they may one day be included as dialects of Malay.” Meanwhile, m.i. between Indonesian and Standard Malay—and the limits thereof—is widely discussed and documented. The sharpest exceptions, to the best of our knowledge, are Kelantan Malay—listed under Malay [zlm]—and Yawi, the two of which are

practically one and the same.¹⁷⁷ Says one Malaysian on Quora, “[U]nderstanding Kelantanese is worse than understanding an Indonesian to be honest... I don’t understand why do when Indonesian speaking in the TV they got subtitle beneath them and Kelantanese don’t.”¹⁷⁸

Thus ISO 639-3 codes have been assigned to Malay varieties on the basis of ethnolinguistic, sociopolitical, and perhaps ethnohistorical distinctions. In most cases, these distinctions seem to have run much finer-grained than m.i. would dictate.

We note, also, that the idea of Indonesian as a linguistic entity was first disseminated—with an unborn nation in mind—in 1928, with the Sumpah Pemuda (Youth Pledge). Bahasa Indonesia as such has just nine decades of history. We also note that Malaysia and Indonesia jointly reformed the spelling of romanized Malay [zsm] and Indonesian [ind] in 1972. The ethnolinguistic distinction between Malay and Indonesian is well established without being ancient, and their orthographic co-evolution has not compromised their distinctness.

3.6—Bisaya

Although said to be ethnolinguistically distinct, Boholano is fully m.i. with Cebuano [ceb] as spoken throughout the rest of its range, on Cebu, Negros, and Mindanao.¹⁷⁹ Urban Boholano, especially as spoken in Tagbilaran, is even less linguistically distinct from mainstream Cebuano than rural Boholano¹⁸⁰; it may be safer to call it ethnically distinct. And there are no Boholano translations of the Bible, as far as we know.

Boholano does not have a separate ISO 639-3 code. Ethnologue lists it as a dialect of Cebuano.

ISO 639-3’s encoding of the Philippine languages—the more widely spoken ones, at least—tracks the popular linguistic boundaries recognized in Filipino society at large. The popular linguistic boundaries, in turn, track m.i. The implicit standard is appar-

171 <https://forum.wordreference.com/threads/assamese-pronunciation-characteristics.3065877/post-15865525> (last accessed Aug 2021)

172 <https://www.quora.com/Can-Bengali-speakers-understand-Assamese/answer/Debabrata-Saha-7> (last accessed Aug 2021)

173 There are a number of varieties called “Malay” that lie outside the Malay macrolanguage—typically Malay-lexified creoles like Manado Malay [xmm], Betawi (Betawi Malay) [bew], Papuan Malay [pmy] and Sri Lankan Malay [sci].

174 Also known as *cakap Sabah* or *bahasa Sabah*.

175 *Ethnologue* notes “[f]ew L1 speakers” under Standard Malay [zsm]; no figures are given.

176 Fajri Koto and Ikhwan Koto (2020), “Towards Computational Linguistics in Minangkabau Language: Studies on Sentiment Analysis and Machine Translation”

177 <https://www.quora.com/Are-there-any-differences-between-the-Pattani-Malay-and-Kelantan-Malay-dialects/answer/Zaid-Mohamad-1> (last accessed Aug 2021); <https://www.quora.com/Are-there-any-differences-between-the-Pattani-Malay-and-Kelantan-Malay-dialects/answer/Yusrin-Faidz-Yusoff> (last accessed Aug 2021, quoting [mfa] as the ISO 639-3 code for “Kelantan-Pattani Malay”)

178 <https://www.quora.com/Is-it-true-that-the-Kelantan-Malay-language-is-hard-for-other-Malaysians-to-understand/answer/Muhammad-Izz-Irfan> (last accessed Aug 2021)

179 <http://www.thelanguagejournal.com/2011/05/difference-between-language-and-dialect.html?showComment=1337123132725#c4099742679691167821> (last accessed Aug 2021)

180 <http://www.lowlands-l.net/anniversary/boholano-info.php> (last accessed Aug 2021); https://en.wikipedia.org/wiki/Boholano_people (last accessed Aug 2021)

ently closer to “full” m.i. than “functional” m.i. For example, there is enough inherent m.i. between Cebuano and Hiligaynon [hil] for untutored speakers of the two languages to converse gainfully, if painfully.¹⁸¹ But Cebuano and Hiligaynon have never shared an ISO 639-3 code.

3.7—Lao, Isan & Lanna

It is well known that Lao [lao] and Isan [tts] are m.i. The separate ISO 639-3 codes reflect the very real ethnolinguistic distinction between Lao and Isan, which in turn is purely a product of the latter-day split between the nations (and nationalisms) of Laos and Thailand.

What is less discussed, and systematically overlooked, is m.i. between Isan and Lanna (Northern Thai) [nod]. Martin Platt writes,

I was speaking to a native of [Lampang] who had spent all his life there and spoke the local language at home. He had traveled to Isan two or three times, he said, and, when I asked, stated that he had had no real trouble conversing with Isan people. He spoke Lampang dialect to them, and they spoke an Isan dialect back to him; except for a word here and there, they understood each other well. When I replied that the languages were virtually the same, he appeared to take offense and strongly disagreed. “No they’re not,” he maintained. “We speak Northern Thai, but they speak Isan.” The idea that the two could both be dialects of the same language, Lao, was simply not acceptable, or even to be considered. When I wrote an article in Thailand on this incident, claiming that Isan and Kam Meuang [Lanna] were dialects of Lao, it was received with amused interest by Isan readers and cool tolerance by others. One Bangkok friend commented, “People will not accept this.” A poet from Chiang Mai, when asked his opinion about the article, replied that he was glad that people from various places could express various ideas.

...

[S]ince “Northern Thai” and “Isan” are mutually comprehensible (a speaker of one can understand a speaker of the other), they can be considered dialects of the same language (although this might be disputed by some linguists as well). However, since Northern

language and culture enjoy a high relative status in Thailand, while Isan language and culture are generally looked down upon, an attempt to link or equate the two goes against social norms. [Volker] Grabowsky makes the observation that the political and cultural center of northern culture, Chiang Mai, lies within Thailand’s borders, while such centers for Lao and Khmer culture are situated outside of Thailand. This illusion of “foreign origin” may help explain why these cultures in Thailand are harder to claim as Thai, and thus are devalued and denigrated.¹⁸²

We especially note the asymmetry of the ethnolinguistic distinction between Isan and Lanna. Isan speakers are not intellectually or emotionally invested in the distinction, and might be open to the idea of Isan, Lanna, and Lao as a single language. Meanwhile, for Lanna speakers, and for all non-Isan Thais¹⁸³, the distinction between Lanna and Isan (and Lao) is set in stone. To blur it would be unthinkable, except when the lens is zoomed out and the focus is on the oneness of Thailand.

Academia and ISO 639-3 have resolved the matter of Lanna and Isan in favor of acknowledging the ethnolinguistic split. We suspect this is the customary outcome of such matters.

Lanna is m.i. with Lao as well, or at least with northern dialects of Lao.¹⁸⁴ This is not surprising if we consider that the Lanna- and Isan-speaking regions are not contiguous but, rather, mutually connected to the northern stretch of the Lao-speaking region of Laos.

3.8—Lacandon & Yucatec Maya

Yucatec Maya [yua] and Lacandon [lac] are m.i.¹⁸⁵ Christian Lehmann asserts that Yucatec Maya and Lacandon (and Itza’ [itz]) “do not differ more from each other than British and American English... They are mutually intelligible and should be regarded as dialects of one language rather than as distinct languages.”¹⁸⁶

182 Martin B. Platt (2013), *Isan Writers, Thai Literature: Writing and Regionalism in Modern Thailand*

183 Primarily Isan speakers, but most likely the ethnolinguistic minorities of the Isan region as well.

184 <https://www.quora.com/How-similar-are-northern-Thai-dialects-with-Lao/answer/Krudsada-Yantarawattana-1> (last accessed Aug 2021)

185 See James D. Nation (2010), *The Maya Tropical Forest: People, Parks, and Ancient Cities*.

186 Christian Lehmann (2017), “Grammaticalization of tense/aspect/mood marking in Yucatec Maya.” In Walter Bisang & Andrej Malchukov (eds.), *Unity and diversity in grammaticalization scenarios*, 173–237. Berlin: Language Science Press.

181 https://www.wikiwand.com/en/Talk:Karay-a_language (last accessed Aug 2021); https://www.reddit.com/r/linguistics/comments/klru8a/phil_languages_does_anyone_know_why_waraywaray_is/ (last accessed Aug 2021)

Lacandon has northern and southern varieties which are generally m.i., but “each Lacandón group considers the other’s dialect as deficient, and at times, unintelligible.”¹⁸⁷

ISO 639-3 has separate codes for Yucatec Maya and Lacandon, but not for northern and southern Lacandon.

3.9—Occitano-Romance

With the deprecation of ISO 639-3 code [gsc], Gascon and Occitan now share [oci]. Catalan [cat] has its own code. This reflects the sociopolitical reality of our time. In the Val d’Aran, both Catalan and Aranese—a variant of Gascon and of Occitan—are official,¹⁸⁸ along with Spanish. Catalan and Occitan are officially distinct.

Occitan (including Gascon, for the purposes of this paper, unless otherwise specified) and Catalan are substantially m.i.; mainstream Catalan and southerly dialects of Occitan—including Aranese—are close to fully m.i.¹⁸⁹

It is well and widely understood that till early medieval times (and arguably into early modern times), Occitan and Catalan or their precursors were indefinite parts of an immense continuum of Romance dialects, varying subtly from district to district across the landscape. What could today be understood as written Catalan did not emerge till the 11th and 12th centuries, and under the sway of Occitan¹⁹⁰; Catalan was written using the Occitan orthography till the early 13th century.¹⁹¹ In Catalonia, a diglossic situation prevailed from about 1200 to about 1400 where prose was written in Catalan, but poetry in Occitan.¹⁹²

The perceived boundaries between Occitano-Romance varieties have fluctuated over time; sometimes there are contemporary variations in perception, such as, in our time, the concept of Valencian as part of Catalan versus the concept of Valencian as a language apart. The name “Occitan” (or the

equivalent) did not emerge till the 19th century,¹⁹³ and what we call “Catalan” was historically referred to in different contexts—sometimes in broader or narrower conception—as *català* or *llemosí* or *romans* or *valencià*¹⁹⁴ or *mallorquí*.¹⁹⁵ The delicate nuance and changeability of the nomenclature can be a formidable threshold to understanding the history; efforts to simplify often simply add to the complexity. The historical view does not lend credence to the idea that the endeavor of classifying human speech can ever be apolitical.

3.10—Scots

Scottish society is conflicted as to what Scots is. Quoting from Hele Priimets (2017), “Non-Standard Language in Irvine Welsh’s *Trainspotting* and in Olavi Teppan’s *Translation of the Novel Into Estonian*”:¹⁹⁶

Adam J. Aitken (1979: 85), Scottish lexicographer and leading scholar of Scots, claims that many Scots speakers view Scots and Scottish English as separate registers and operate between them depending on social circumstances. It is therefore widely believed that while Scottish Standard English is at the one end of the linguistic continuum, Scots is at the other...

The idea that Scots is a language separate from English becomes perhaps most apparent when its history is observed...

On the one hand, Scottish English is the English language and all the varieties of English that are used in Scotland, including Scots, but on the other hand, Scottish English refers to the English language used in Scotland but not to Scots (Leith 1997: 133). When included, Scots is seen as a northern dialect of English that belongs to the category of English used in Scotland, but when excluded, Scots is regarded as a separate language that is to a certain extent

187 <https://dobes.mpi.nl/projects/lacandon/language> (last accessed Aug 2021)

188 This makes Aranese the only official dialect of Occitan anywhere.

189 <https://www.quora.com/Are-Occitan-Provençal-and-Catalan-mutually-intelligible> (last accessed Aug 2021); <http://www.trob-eu.net/en/catalan-and-occitan-one-diasystem-two-languag.html> (last accessed Aug 2021)

190 Kevin Stillwell & Lauren Hetrovicz (2013), “The standardizations of Catalan: Latin to present day,” in *Studies in the Linguistic Sciences: Illinois Working Papers* 2013: 68–86.

191 Id.

192 John S. Lucas (2006), “*Trovas lemosinas* or *Llengua catalana: Majaderos de Castilla* and the Many Names for the Catalan Language”; see also <http://www.trob-eu.net/en/catalan-and-occitan-one-diasystem-two-languag.html> (last accessed Aug 2021).

193 <http://www.trob-eu.net/en/the-name-for-the-language-of-the-troubadours.html> (last accessed Aug 2021)

194 John S. Lucas (2006), “*Trovas lemosinas* or *Llengua catalana: Majaderos de Castilla* and the Many Names for the Catalan Language”; see also <http://www.trob-eu.net/en/catalan-and-occitan-one-diasystem-two-languag.html> (last accessed Aug 2021).

195 <http://www.cardonavives.com/artdocumentos.asp?id=3018&tit=Andreu%20Sempere,%20Gramatica%20Valenciano-Llatina.%201546.%20Segle%20d%B4Or%20de%20la%20Llengua%20Valenciana:%20XV%20i%20XVI> (last accessed Aug 2021; in Valencian [cat]); <https://www.mallorcaapocrifa.com/language.html> (last accessed Aug 2021)

196 PDF available at <https://dspace.ut.ee/handle/10062/57592> (last accessed Aug 2021; pp. 13–16).

mixed with the English language originating from England (Aitken 1992b: 903).

Information on Scots is abundant but often extremely confusing. For example, Ethnologue 22 credits Broad Scots [sco] for 99,200 L1 speakers (including 55,800 in Scotland) and 1.5 million L2 speakers. These figures are free interpretations of the responses to two items on the 2011 census that in turn had been interpreted freely by the responding public.¹⁹⁷

The idea of Scots as a separate language seems to trigger resentment in many Scots¹⁹⁸—especially, perhaps, residents of the Central Belt who rarely hear Broad Scots spoken separately, but rather as fragments plugged into a matrix of Standard English or something like it.¹⁹⁹

Accordingly, written accounts of the intelligibility of “Scots” to English speakers²⁰⁰ are tricky to parse. The commentator may be referring to Broad Scots, to some form of Scottish (Standard) English, or to some mixture of the two. Broad Scots is somewhat intelligible to English speakers, but simple assertions that Scots is intelligible to English speakers would seem to be referring to some form of Scottish English that is not Broad Scots, or Scottish English with Broad Scots mixed in.²⁰¹

Broad Scots serves as an example of how the gravitational pull of a political and cultural center can make it, over time, mind-bendingly tricky to even talk about a subjugated neighboring language.

197 See <https://bellacaledonia.org.uk/2013/10/03/gaelic-scots-and-other-languages> (last accessed Aug 2021).

198 <https://www.thenational.scot/news/14861018.letters-to-the-national-ii-principled-objection-to-the-project-of-promoting-scots> (last accessed Aug 2021)

199 See Thomas Clark (2018), “‘Nobody actually talks like that’ – Why is everyone terrified of Scots?” (<https://www.thenational.scot/news/17262527.nobody-actually-talks-like-that--everyone-terrified-scots>, (last accessed Aug 2021).

200 The intelligibility of English to Broad Scots speakers is virtually untestable, since all Scots are educated in English and massively exposed to it.

201 We doubt most American English speakers could make heads or tails of the snippet of Broad Scots (Doric) in this video: <https://youtu.be/ehetPhuZQg> (last accessed Aug 2021). Heard together with the English translation in the comments, it suddenly becomes kind of comprehensible. We wonder if such a variety might nevertheless be inherently intelligible to people from the North of England. On the other hand, we find the Scots on this page functionally intelligible “in gestalt”: <https://www.gov.scot/policies/languages/scots/#scots> (last accessed Aug 2021). We had also noticed from some time back that Scots Wikipedia is comprehensible, but it turns that much of it is spurious (<https://www.theguardian.com/uk-news/2020/aug/26/shock-an-aw-us-teenager-wrote-huge-slice-of-scots-wikipedia>, last accessed Aug 2021).

4—The Varieties in Contrast

In this section we review the relationships between the languages or varieties subsumed under [nan]. But, first, let’s acknowledge the dominant underlying paradigm, which is that Chinese constitutes a single, unified language.

The paradigm of a single Chinese language has been discarded in meticulous linguistic study outside China. But it is still the dominant paradigm in global society. By habit, people the world over refer to the Chinese language: in casual speech, or when the focus is on stocks, or dogs, or anything but “the Chinese languages.”²⁰² There is nothing mean-spirited in this. If anything, it is perpetuated by how most Chinese—and many Taiwanese—present and explain their own languages.

In China, even linguists cultivate ambiguity. The word *Hànyǔ* 汉语~漢語 (“Chinese”) is used in place of *Zhōngwén* 中文 (“Chinese”) to refer to the uni-language, in the kind of contexts where an Anglophone official or professor might get grief for being vague. But *Hànyǔ* can also refer to *Hànyǔ* the language family (Sinitic), or to the national language, *Hànyǔ* (Mandarin). So besides sharing its name with Standard Mandarin, the *Hànyǔ* language family conveniently coincides with the beleaguered uni-language.²⁰³ Meanwhile, Hongkongers remind us that Cantonese is 中文 (Mand. *Zhōngwén*) too.²⁰⁴

To most Chinese in most contexts, Chinese is one language, just as most Tunisians see Tunisian and Standard Arabic as registers of one language. The Chinese belief is more acrobatic in linguistic terms; in terms of belief, this is irrelevant. In the orthodox view, Chinese is *the* language; breaking Chinese down into Eight Great Topolects (八大方言) and then some (seven more, to be exact) should be enough to satisfy idle intellectual curiosity. When did m.i. between this or that dialect ever matter in China?

To be fair, in some sense, the *Hàn* are a seamless ethnolinguistic unit. The *Hàn* supertribe as we know it is a modern invention, but we do live in modern times, and the *Hàn* people are a modern reality. If

202 What is a “Chinese” language? Victor Mair deduced that a Chinese (*Hàn* 汉~漢) language is any one of the languages “spoken by the so-called Han peoples.” See Victor Mair (2013), “The Classification of Sinitic Languages: What Is ‘Chinese’?” “Han peoples” seems to be implicitly defined in reality as any community descended from tax-paying subjects of the Qing empire who used Classical Chinese in a collective, traditional sense.

203 The *Hànyǔ* language family is nominally finding its way into mainstream discourse, but Chinese intellectuals tend to be unaware that “language family” implies a history of divergence from a common ancestor language.

204 <https://disp.cc/b/654-alup> (last accessed Aug 2021; mostly in Mandarin)

Hokkien is a language based on ethnolinguistic identity despite limited intra-intelligibility, then why not Chinese, on a much larger scale, without the intra-intelligibility but with Classical Chinese and modern Mandarin as unifying written forms? In this sense, [nan] consists of just three languages: Kelantan Local Hokkien, a moribund Taiwanese,²⁰⁵ and Chinese; and none of the Eight Great Topolects should be represented at ISO 639-3.

We have not seen this argument made, but it is coherent. We believe an unspoken sense that Chinese is a single language underlies *Ethnologue* or the *International Encyclopedia of Linguistics*²⁰⁶ listing of *Leizhou* and *Taiwanese*, for example, as dialects of a single language. There is no m.i. between Luichew and Taiwanese, not even strained m.i. à la Cebuano [ceb] and Hiligaynon [hil], or spotty m.i. à la Bengali [ben] and Assamese [asm]. On the level of reality, there is at most a shared *Chinese* ethnolinguistic identity, a sense widely shared—even by some Taiwanese speakers—that Taiwanese, Luichew, Hakka, and Cantonese are a uni-language despite the lack of m.i.

Would-be defenders of a unified Chinese seem to approach discussions of m.i., even, with a unique set of proxies, shorthands, and work-arounds designed to minimize what they perhaps perceive as the fragmentation of the putative uni-language. There is, for example, a Youtube video entitled “广东雷州话防疫宣传, 据说台湾人也能听懂八成,” which translates into “Public announcement on the pandemic in Luichew; the Taiwanese can supposedly understand 80% of this.”²⁰⁷ This video and its comments yield at least two takeaways.

First, the announcement is not actually in Luichew. It is an announcer reading a script written in a formal register of Mandarin into Luichew.

Second, the Mandarin script itself is given in the subtitles; most of the commenters gauged their listening comprehension of the Mando-Luichew announcement while reading it at the same time, unaware that this could have an effect.

For the uninitiated, what is involved here is a sort of “guided linear translation” where cognate morphemes are used by default. It would be something like if Spanish were written in a Sino-script, and English speakers spoke English but mostly wrote and read in Sino-script Spanish; on seeing the writ-

ten form representing *Una tarjeta para ti*, they could either read it directly as Spanish, or insta-translate it on the fly to “A target for at thee,” with no variance in meaning. Where using cognates would result in gibberish in the target language, straight word-for-word translation is substituted, i.e. “A target for thee” or “A card for thee.”²⁰⁸

Luichew is so distant from Taiwanese that Taiwanese speakers are hard-pressed to comprehend even Mando-Luichew without the aid of the Mandarin subtitles. As @KZYQN988 puts it (our translation):

I am Taiwanese. Don't kid yourselves. With the subtitles covered up, we can catch about 30%—we can figure out it's about COVID. [If you're trying to gauge your comprehension, d]on't just read along with the subtitles. Whoever says they got 80%, I guarantee they were reading along. In conclusion, for us Taiwanese: no subtitles, no understanding.²⁰⁹

This comment rocked boats and came under fire, but was not actually refuted.

The mental lapses surrounding this video line up with the belief of the old-time literati in the primacy of the written character, with speech being merely a corruption; and with the Chinese nationalist and Western imagining of Chinese as a uni-language within which people talk different but write and think the same. Thus Luichew is assumed to not be syntactically and lexically unique the way Tamil or Mapuche or Portuguese are; it is merely one of many possible phonological instantiations of written Mandarin, which is seen not as Mandarin but as *Chinese*.

Under this worldview, m.i. between actual Luichew and actual Taiwanese is wholly irrelevant. The syntax and lexicon of Luichew and Taiwanese are overlooked, or studied as fossils; their phonology—specifically, their readings of Sinographs—are simply localizations of the same linguistic operating system: Chinese. In fact, it is hard to imagine the 80%-ers in the comments speaking anything but Mandarin if they found themselves stranded in a Luichew-speaking land. If they stayed for five years, they might learn Cantonese. It would be ironic indeed if they chose m.i. as a basis on which

205 A Taiwanese of a subset of Christians and indigenous Formosans who as a (dwindling) community do not write or read Chinese.

206 William Frawley (2003), *International Encyclopedia of Linguistics*, 2nd edition. New York: Oxford University Press.

207 https://youtu.be/OkS_VhqYyQg (last accessed July 2021)

208 Over decades or centuries, these guided translations trickle through the target language and chemically gut it; “target” replaces “card,” first in classy speech, then entirely; “for at” is resisted for decades but eventually becomes sound, classy English; “you” becomes vulgar usage; and so on.

209 https://youtu.be/OkS_VhqYyQg (last accessed Aug 2021; comments in Mandarin)

to defend Luichew and Taiwanese being classified as the same language.

We would be remiss if we did not mention the tables of cognates through which macro-Sino-topolectologists view Sino-topolects. Below, for example, is Table 5.30 from *Southern M̃n: Comparative Phonology and Subgrouping* by Bit-chee Kwok; we have taken the liberty of expanding the abbreviations, and we apologize for the lack of graphic design.

The last column represents a Teoyeo dialect of Teochew; the second-to-last column represents a Chiangchew dialect of Hokkien. From this angle and altitude, Hokkien and Teochew sure do look alike.²¹⁰ The utility of such charts is not to be underestimated, but—for a variety of reasons, and much more so than with Romance or Philippine specialists—modern comparative Sino-topolectologists often live and work at a relentless geographic, social, and linguistic remove from the languages they study. Unless anchored by some combination of social or personal knowledge of the languages s/he is studying, the topolectologist poring over these tables can't help but be lulled into thinking, *How different could they really be? You say /tsik/, I say /tsek/...* It is easy to forget that Hokkien and Teochew have their own syntax, their own diction, their own tones, their own rhythm; that *lék* (as it is written) may or may not be a common word for LIGHT GREEN in Hokkien, and likewise with *lék* (as it was also written) in Teochew...

The idea of Southern Min as an intra-intelligible language à la Telugu or Lithuanian—conclusively refuted by the evidence, as we will see—appears to be substantially based on how Hokkien, Teochew, Luichew and even Samheung look on paper, in cognate tables.

If the same approach were taken with the Philippine languages,²¹¹ we would surely conclude that Cebuano and Waray were dialects of the same language, and take things from there. But we do not approach the Philippine languages, nor hardly any other languages, that way.

A principled approach can be taken to determining whether Taiwanese, Luichew, and the other elements of [nan] are languages in their own right, and, if not, which languages they group with. M.i. and ethnolinguistic identity should be approached empirically. Let us proceed with rigor and an aware-

ness of the biases that precede us, keeping in mind that:

By 'bottom up' criteria, there are no empirical bases for defining a given language (in the common sense)... The crucial factor is the distinct *identity* of the language and of its speakers *qua* speakers of that language...

The deciding factors are subjective and non-linguistic (i.e. are of political, ethnic, historical, and/or sentimental nature). In any case, 'languages' are not 'things' which we 'recognise' or 'discover' being then able to act upon (e.g. by working on 'maintenance' or 'preservation'). Rather, they are *created* by social action.²¹²

210 The Teoyeo form for "father's younger brother" is a typo—inconsequential in this context—and should be *tsek*7.

211 Tables from "Kapampáŋan Cognates with Other Philippine Languages," found on the UP Aguman Facebook page (<https://www.facebook.com/upaguman/photos/05-of-13/10159195277655405>, last accessed Aug 2021).

212 N.J. Enfield (2002), "How to define 'Lao', 'Thai', and 'Isan' language? A view from linguistic science"

213 This Sinograph is conjectural. Comparative macro-Sino-topolectology fears the worst when an etymon doesn't have an etymological Sinograph to escort it, even when no etymological Sinograph can be ascertained.


	Proto-Mǐn	Fú-ān (E. Mǐn)	Jiànyáng (N. Mǐn)	Yǒng'ān (C. Mǐn)	Zhāngzhōu (S. Mǐn)	Cháoyáng (S. Mǐn)
竹 'bamboo'	**_yk	tək ⁷	ty ⁷	ty ⁷	tik ⁷	tek ⁷
熟 'cooked'	**_yk	sək ⁸	sy ⁸	ʃy ⁴	sik ⁸	sek ⁸
叔 'father's younger brother'	**_yk	tsək ⁷	sy ⁷	ʃy ⁷	tsik ⁷	sek ⁷
燭 'candle'	**_yok	tsuk ⁷	tsy ⁷	tʃy ⁷	tsik ⁷	tsek ⁷
粟 'grain'	**_yok	souk ⁷	sy ⁷	tʃhy ⁷	tshik ⁷	tshek ⁷
綠 'light green'	**_yok	luk ⁸	ly ⁸	ny ⁴	lik ⁸	lek ⁸
蜀 ²¹² 'one'	?	si ⁷ ⁸	tsi ⁸	–	tsit ⁸	tsek ⁸

Table 5.30 from *Southern Mǐn: Comparative Phonology and Subgrouping* by Bit-chee Kwok

NUMBERS IN SOME PHILIPPINE LANGUAGES

	1	2	3	4	5	6	7	8	9	10
KAPAMPÁNGAN	isa	addua	atlu	apat	lima	anam	pitu	walu	siam	apulu
AKLANON	isaea	daywa	tatlo	ap-at	lima	an-om	pito	waeo	siyam	napueo
CEBUANO	usa	duha	tulo	upat	lima	unom	pito	walo	siyam	napulo
HILIGAYNON	isa	duha	tatlo	apat	lima	anum	pito	walo	siyam	napulo
ILOCANO	maysa	dua	tallo	uppat	lima	innem	pito	walo	siyam	sangapulo
KINIRAY-A	isara	darwa	tatlo	apat	lima	anem	pito	walo	siyam	pulu
PANGASINAN	sakey	dua	talo	apat	lima	anem	pito	walo	siam	sampolo
TAGALOG	isa	dalawa	tatlo	apat	lima	anim	pito	walo	siyam	sampu
TAUSUG	isa	duwa	tu	upat	lima	unum	pitu	walu	siyam	hangpu
WARAY	usa	duha	tulo	upat	lima	unom	pito	walo	siyam	napulo

SOME KAPAMPÁNGAN COGNATES (LOPEZ, 1974; BLUST & TRUSSEL, 2017)

PP	*ha(n)daw 'day; sun'	
KAP	aldó 'day; sun'	IVT araw 'day; sun'
AKL	adlaw 'day; sun'	MGD adlaw 'day; sun'
CEB	adlaw 'day; sun'	PNG agew 'day; sun'
HIL	adlaw 'day; sun'	SBL awló 'day; sun'
IBG	aggaw 'day; sun'	TAG araw 'day; sun'
ILC	aldáw 'day; sun'	WRY adlaw 'day; sun'
PP	*hanak 'child'	
KAP	anáq 'child'	IVT anák 'child'
CEB	anáq 'child'	MGD anák 'child'
HIL	anáq 'child'	PNG anák 'child'
IBG	ana? 'child'	SBL anák 'child'
ILC	anáq 'child'	TAG anák 'child'
ISG	annak 'child'	WRY anák 'child'
PP	*halun 'swell of the sea; wave'	
KAP	alon 'wave'	PNG alon 'small waves'
CEB	alon 'wave'	SBL alon 'wave'
HIL	alon 'small waves'	TAG alon 'wave'
ILC	alon 'wave'	WRY alon 'move in the water'
IVT	alnon 'restless sea'	
MGD	alon 'wave'	
PP	*hapuy 'fire'	
KAP	apí? 'fire'	MGD apoy 'fire'
IBG	afi 'fire'	PNG apóy 'fire'
ILC	apúy 'fire'	SBL apóy 'fire'
ISG	apúy 'fire'	TAG apóy 'fire'
ITB	hapoy 'fire'	BAL api 'fire'
IVT	apuy 'fire'	MAL api 'fire'

4.1—Taiwanese in Contrast

Linguistically, the closest varieties to Taiwanese are Hokkien, Haklau, and Teochew.

Vs Hokkien

By all good-faith accounts, Taiwanese is m.i. with Amoy Hokkien.²¹⁴ The Hokkien spoken at the mouth of the Jiulong 九龍 River, immediately to the west of Amoy, is also highly m.i. with Taiwanese, as is Philippine Hokkien. These dialects account for up to three million speakers (overwhelmingly native), or roughly 12 to 15% of native Hokkien speakers worldwide.²¹⁵ Comprehension of other dialects of Hokkien tends to be strained or limited for Taiwanese speakers²¹⁶; this is expected, since Chiangchew and Amoy Hokkien speakers have trouble understanding large swathes of Hokkien dialects themselves. See §2.6.

Comprehension of Taiwanese by Hokkien speakers tends to be much higher, for two reasons. First, many if not most Hokkien speakers have had exposure to Taiwanese mass media. Second, Amoy Hokkien and Chiangchew Hokkien generally are highly intelligible for most Hokkien speakers,²¹⁷ and mainstream Taiwanese is phonologically intermediate to the two.

Non-learned Japanese loanwords in Taiwanese come up frequently in discussions of m.i., but they occur modestly in general conversation. They are dominant in some technical fields. A Hokkien speaker probably could not follow a lecture on motorcycle engineering in Taiwanese.²¹⁸

214 Taiwanese speakers who put their minds to not understanding Hokkien will claim that Amoy Hokkien is incomprehensible. See <https://pttweb.tw/s/2Naex> (last accessed July 2021; in Mandarin).

215 Amoy Hokkien is the local dialect of most (but not all) of Amoy Island and the islet of Kulangsu, corresponding to the districts of Siming 思明 and Huli 湖里, home to about 2.1 million total (<https://zh.wikipedia.org/wiki/厦门市#行政区划>, last accessed Aug 2021; in Mandarin). If half the population are local (generous), and two thirds of the local population speak Amoy Hokkien—keeping in mind that locals under 20, and many young adults as well, do not know Hokkien—we (generously) estimate 700,000 Amoy Hokkien speakers. Meanwhile, the Haicang 海沧 and Longhai 龙海 districts west of Amoy total about 1.3 million inhabitants, of whom a relatively high percentage would be Hokkien-speaking locals. The *Ethnologue* estimate of one million Hokkien speakers in the Philippines is probably still roughly in the money, although at this point at least a large minority are probably unable to speak Hokkien without constant code-switching to English and Philippine languages, meaning Taiwanese speakers would not be able to understand them. Thus these highly intelligible dialects (to Taiwanese speakers) are spoken by somewhere shy of three million speakers, and possibly closer to two million.

216 See <https://twitter.com/laichinan/status/1262569135461945344> (last accessed July 2021; in Mandarin).

217 See §3.6.

218 Most Japanese loanwords in Taiwanese are learned, and shared with Korean, Mandarin, Cantonese, Vietnamese,

Hokkien and Taiwanese were arguably a single language till the late 19th or early 20th century. To be exact, Taiwanese probably did not exist as a firm ethnolinguistic entity till the creation of the province of Taiwan in 1887.²¹⁹ Taiwan became a dependency of Japan in 1895, cementing its separateness from Fujian. Thus in 1929 Lián Hêng 連橫 would write 夫臺灣之語傳自漳泉而漳泉之語傳自中國 (roughly “The Taiwanese tongue came from Chiangchew and Choanchew; that of Chiangchew and Choanchew came from Middle China”), in a series of essays about the Taiwanese language that would be published in book form as the *Tâi-oân Gú-tián* 臺灣語典 (Mand. *Táiwán yǔdiǎn*).

In 1885 when Formosa’s first periodical—the romanized Hokkien-Taiwanese *Church News* (*Kàu-hōe Kong-pò* 教會公報)—began publication in the city of Taiwan (later Tainan 台南; Taiwanese *Tâi-lâm*), the paper circulated on the continent as well; contributors hailed from both sides of the straits, and the paper actively reported on continental happenings. A similar periodical debuted on the continent in 1888; by some point around 1900, *Church News* had become strictly a Formosan paper.

The first Hokkien-Taiwanese translation of the Bible, completed in 1884, was spearheaded by James Laidlaw Maxwell, missionary to Formosa. Thomas Barclay, working in both Formosa and Fujian, completed another translation in the 1930s. These early Hokkien-Taiwanese Bibles skewed Amoy and were heavy on the translationese. The *Âng-phôe* 紅皮 New Testament—fully Taiwanese and a Catholic-Protestant joint effort—came out in 1973; in 1975, romanized Taiwanese Bibles were confiscated by Republic of China authorities, who feared subversion in the use of romanized Taiwanese. Work on the *Âng-phôe* Old Testament was halted and never resumed. In the post-martial law era, the Barclay Bible resumed its reign as the go-to Bible for Taiwanese-language worship, but the archaic Amoy Hokkien had become formidably and, for many, intolerably distant from the everyday Taiwanese vernacular. The *Choân-bîn* 全民 Bible, published in 2015, came about as a thorough Taiwanization of the Barclay translation.

There is today what looks like an asymmetry of ethnolinguistic identity. Taiwanese speakers, unless

mainstream Hakka, and mainstream Hokkien, but not Penang-Medan Hokkien. These words are usually not thought of as Japanese loanwords—outside of Korea, at least.

219 Ethnolinguistic identity among the Chinese is strongly under the influence of provincial identity. To add a fresh example to the pile, in Malaysia, the Kwongsai 廣西~广西 (Mand. *Guǎngxī*) are ethnolinguistically separate from the Cantonese despite sharing a language.

employed in certain segments of the Republic of China administration, do not recognize Hokkien and Taiwanese as a single language. Hokkien speakers, on the other hand, encounter what sounds like a Hokkien dialect in Taiwanese telenovelas and pop music, while state propaganda out of both China and the Republic of China stresses that Taiwanese is Hokkien. Many Hokkien speakers, especially from China and Malaysia,²²⁰ claim Taiwanese as a dialect of Hokkien.

The anatomy of the asymmetry is complex. Superficially, we are reminded of Malay and Indonesian. A slightly better analogy might be Isan [tts] and Lanna [nod]. Lanna speakers are more sensitive than Isan speakers to official viewpoints emphasizing the distinctness of Isan, while Hokkien speakers are more sensitive than Taiwanese speakers to official viewpoints downplaying the distinctness of Taiwanese.

An exceptional but telling case concerns Hokkien instruction on Quemoy. As with most dialects of Hokkien, m.i. between Quemoy Hokkien and Taiwanese is strained or lacking.²²¹ The Republic of China provides grade school pupils with an hour of instruction a week in the “roots” language²²² of their choice. But the Ministry of Education does not recognize Hokkien and Taiwanese as separate languages: it calls Taiwanese *Minnán* 閩南~閩南 (~ Hokkien); publications and testing for what is nominally Hokkien are all based on Taiwanese. So would-be Hokkien instructors on Quemoy—where many young adults struggle to speak Hokkien as it is—must study Taiwanese to get certified to teach a Hokkien that they will not have been prepared to teach. In January 2021, the Kinmen Daily News reported (our translation):

[According to Education Bureau Chief Lô Tek-chúi 羅德水, a]t this point, with the weekly roots-language classes, some instructors are Taiwanese. They speak *Hokkien in Taiwanese-language dialect* [台語腔閩南話; emphasis added]. They cannot speak the Quemoy dialect. There are also uncertified instructors.

The Ministry of Education has based certification for [Hokkien] instructors on the Taiwanese-language dialect (台語腔). According to Lô Tek-súi, the Ministry has agreed that Quemoy

may set up its own system of certification, in collaboration with Taiwan Normal University and Tsing Hua University. In the future, Hokkien instructors at the elementary and secondary levels will be able to get certified through a system created by the research institutes at the behest of the county government of Quemoy, instead of through the certification system for Taiwan, alleviating the phenomenon of *Quemoy speech being taught in Taiwanese-language dialect* [以台語腔教金門話; emphasis added].²²³

Thanks to the Cold War and its after-effects, most Hokkien speakers encounter Taiwanese at a distance, typically as consumers of mass media. Quemoy Hokkien speakers have skin in the same games as the Taiwanese-speaking community. They have no illusions that Taiwanese and Hokkien are interchangeable.

Taiwanese is m.i. with some dialects of Hokkien but not others. Amoy Hokkien is close to Taiwanese, but Amoy Hokkien has 700,000 speakers at best, less than 5% of the (native) total for Hokkien; and speakers of other dialects do not learn to speak Amoy Hokkien the way they may have done in earlier days. Comprehension of Hokkien by Taiwanese speakers is genuinely limited for the most part. Anecdotally, Hokkien-speaking tourists speak Mandarin when in Taiwan, in contrast to Taiwanese-speaking tourists in Taiwan, who switch between Taiwanese and other languages using roughly the same calculus that they use around town at home.

On the surface, there is asymmetric ethnolinguistic unity between Taiwanese and Hokkien, where Hokkien speakers feel the two are one and the same, while Taiwanese speakers feel Taiwanese and Hokkien are distinct. The trouble with Hokkien instruction on Quemoy suggests that Hokkien speakers do observe a distinction versus Taiwanese when things get real. In any case, shared ethnolinguistic identity must by definition be mutual. We do not know any examples of “unilateral” ethnolinguistic unity, where speakers of two partially m.i. varieties are bundled together because of unifying sentiments on one side. If such a case exists, it would be exceptional and probably unstable.

Malay [zlm + zsm] speakers assert from time to time that Indonesian [ind] is a kind of Malay, but the two languages have separate codes and separate service on every electronic platform from Android to Apple, from Facebook to Google Translate, from

220 Philippine Hokkien speakers, as a control, do not seem to have an emotional stake in the issue.

221 <https://www.cna.com.tw/news/ahel/202007210273.aspx> (last accessed Aug 2021; in Mandarin)

222 These include any non-extinct language native to Taiwan, the Pescadores, or Botel Tobago as of 1945; the local languages of Quemoy and Matsu; and Vietnamese and Indonesian.

223 <https://www.kmdn.gov.tw/1117/1271/1272/328288> (last accessed Aug 2021; in Mandarin)

Wikipedia to Malaysia's own Air Asia. And while Isan [tts] speakers can see how Isan and Lanna [nod] might be the same language, it is hard to imagine the two being formally merged over the protests of Lanna speakers. Boholano and Valencian have been rolled into Cebuano [ceb] and Catalan [cat], respectively, despite some ethnolinguistic contrast, but in both cases there is high or full m.i. throughout the pairing, as well as a majority sense among Boholano (Valencian) speakers that what they speak is also Cebuano (Catalan).

M.i. between Taiwanese and Hokkien is strained, albeit asymmetrically. There is also no shared ethnolinguistic identity as far as most Taiwanese speakers are concerned. This is crucial because Hokkien itself, lacking functional intra-intelligibility, is based solely on shared ethnolinguistic identity.

Hokkien-Taiwanese was arguably a single language till the 1880s. The split began with the creation of the province of Taiwan in 1887. The idea of Taiwanese as a language unto itself was already taken for granted in mainstream Taiwanese society by the late 1920s, when the idea of an Indonesian language first came into the world. Hokkien and Taiwanese were clearly separate languages by the end of the Cold War, a century later.

Unsatisfied, some will still argue that Taiwanese and Amoy Hokkien are just too alike to be anything but one language. There is more to this than meets the modern eye.

Amoy did not come into its own until the second half of the 19th century, after the Opium War and Treaty of Nanking. It had not been much of a city. It was a small port on the cusp of big seas, a place for drop-outs and outsiders. It had more in common with the frontier prefecture of Taiwan than with the rest of southern Fujian.

In the 17th century, the Koxinga family's rebel kingdom was based on both Formosa and Amoy. Their capitals were Amoy till 1662, and Anping 安平 after. They did not lose Amoy and Quemoy for good until 1680, three dozen years after the Qing took Beijing. In 1684, when the Qing empire annexed part of Formosa, they set up the Taiwan-Amoy Circuit 福建分巡臺灣廈門道~福建分巡台湾厦門道 to govern the new territory, along with Amoy. This setup lasted till 1727. In the mid 19th century, six ports in Fujian were opened to international trade by treaty: Amoy and Fuzhou 福州 on the continent, and Takow 打狗, Anping, Tamsui 淡水 and Kelung (Keelung) 鷓籠 on Formosa. When Charles Le Gendre served as American consul at Amoy (1866-72), he presided over Amoy and the four Formosan ports.

The 19th century was a time of massive change, and migration, along the Straits of Formosa. As a frontier society (from a continental, settler perspective), Formosa was wide open to the flux of the epoch. The old cities and towns of Chiangchew, Choanchew and Teochew changed more slowly. But Amoy was a new town, flooded overnight with migrants and merchants from all over the hinterland. A new blended speech had been forming in Taiwan, and another arose in the island city of Amoy. The two were alike because they were made of the same stuff at the same time, on two sides of the same lab with people going back and forth.²²⁴

Amoy Hokkien does not prove that Taiwanese is Hokkien. If anything, Amoy Hokkien and the dialects just to the west of it, with somewhere shy of two million speakers all told, are *continental Taiwanese*. They are phonologically unique relative to the rest of Hokkien.

On the ethnolinguistic level, it seems obvious that Amoy Hokkien casts its lot with Hokkien, not Taiwanese. But the question has probably never been posed.

Of course, it is worth pointing out that many sets of generally m.i. tongues—Lacandon [lac] and Yucatec Maya [yua], for one²²⁵—map to separate codes based on ethnolinguistic factors.

As discussed in §2.6, Hokkien is becoming a rural language in China. Every city and administrative seat in Hokkien-speaking China has become a center of language shift to Mandarin, Amoy first and foremost. Barring the unforeseen, the leading cities of the Hokkien-speaking world in 50 years' time will be Penang and Medan.

Taiwanese and Hokkien are ethnolinguistically distinct and generally not m.i. Taiwanese warrants a separate code from Hokkien. Amoy Hokkien is linguistically intermediate between the rest of Hokkien on one hand and Taiwanese on the other. In a linguistic and historical sense, Amoy Hokkien arguably belongs with Taiwanese; speaking from social experience, we have little doubt that Amoy Hokkien goes with Hokkien.

²²⁴ Much of the co-evolution was apparently by subtraction: rare or arcane expressions were culled from the mix; common elements and simple structures won out in the maritime koine. The intelligibility of Philippine Hokkien for Taiwanese speakers seems to flow from the same maritime process of simplification.

²²⁵ Lacandon speakers are descended from Maya-speaking natives of the Yucatán who fled into the jungle to escape subjugation at the hands of the Spanish; working-class and rural Taiwanese speakers claim cultural and in part genetic descent from 17th century desperadoes fleeing the baleful feudal regimes on the continent.

*Vs Haklau*²²⁶

Taiwanese is to some extent m.i. with Haklau. Comprehension of Haklau for Taiwanese speakers seems to be limited to direct conversation.²²⁷ Despite the familiar cultural context, we as Taiwanese speakers—without leaning on the Mandarin subtitles—are only able to catch a word here or a half-sentence there of this²²⁸ well-produced TV episode in Haklau. As for this²²⁹ 26-second snippet of *Crayon Shin-chan* dubbed in Haklau, the gist of the first 20 seconds is easily guessed in context, but the punchlines are a mystery.

In our limited experience, Haklau is less comprehensible than Hokkien in general (and not comparable to Amoy or Philippine Hokkien). Some Haklau speakers seem to have better comprehension of Taiwanese than vice versa, possibly through exposure to Taiwanese mass media.

One of our correspondents rented a room from a Haklau speaker in the city of Canton. They could not converse heterolingually at normal speeds. The lessor often spoke loudly on the phone in Haklau on the premises; our correspondent says he generally did not understand, but from time to time a fully-formed Taiwanese-sounding sentence would jump out. We are reminded of Afrikaans sentences like *My pen is in my hand...*

Taiwanese and Haklau are even less m.i. than Taiwanese and non-Amoy Hokkien. Taiwanese and Haklau are ethnolinguistically distinct, and did not undergo a period of shared ethnolinguistic gestation as Taiwanese did with Hokkien. Taiwanese warrants a separate code.

We compare Haklau to Hokkien in §4.5.

Vs Teochew

By all accounts, Taiwanese and Teochew are even less m.i. than Taiwanese and Haklau.²³⁰ Taiwanese and Teochew are ethnolinguistically distinct. This is heightened by the strength of the Teochew identity; it is arguably the strongest ethnolinguistic identity under the [nan] umbrella. Taiwanese warrants a separate code from Teochew.

We compare Teochew to Hokkien in §4.4.

4.2—Hainanese in Contrast

Linguistically, the closest variety to Hainanese is Luichew.

Vs Luichew

Linguistically, Hainanese and Luichew are closest to each other. There are reports of functional m.i. between Luichew and the Hoihow 海口 (Mand. *Hāikǒu*) dialect of Hainanese.²³¹ There is a concrete question of which dialect of Luichew is involved as well.²³² Hainanese-Luichew m.i. seems to be restricted to the Luichew spoken toward the tip of the peninsula. As mentioned in §2.3, southern Luichew, like mainstream Hainanese and Vietnamese, took part in a regional implosivizing consonant shift; Luichew City Luichew did not. Even at the tip of Luichew, some claim Hainanese is incomprehensible.²³³ Note that while Hoihow is the capital of Hainan, the Bunsio (Mand. *Wénchāng*) 文昌 dialect is (or was) the koine. There is some linguistic distance

226 Taiwanese-Haklau and Taiwanese-Teochew m.i., although limited, is not necessarily “transitive” through Hokkien. If nothing else, 18th century Teochew and Haklau settlers and their descendants played a part in the making of the Taiwanese language and Taiwanese-speaking society. In the middle-west of Taiwan, especially, there are a number of villages named *Hái-hong* 海豐 (Mand. *Hāifēng*) after the Haklau-speaking county; and the villages named *Tiô-iû* 潮洋 (Mand. *Cháoyáng*) must have been founded by folk from Teoyeo 潮阳~潮陽. There is also a village in northern Taiwan where the speech of the elderly shows clear Haklau characteristics. The town of *Tiô-chiu* 潮州 was obviously named after the original Teochew, and much ink has been spilled on the relationship between the Lords of the Three Hills (*Sam-san Kok-ông* 三山國王) cult and a Teochew minority that is no longer distinct. Linguistically, in etyma where Choanchew Hokkien, Teochew, and mainstream Taiwanese *-ng* correspond to Chiangchew Hokkien *-ui*, the mainstream Taiwanese form is assumed to be a Choanchewism; it might be more realistic to acknowledge secondary influence from Teochew.

227 <https://www.zhihu.com/question/22463095> (last accessed July 2021; in Mandarin)

228 The series is entitled 暹罗猪槽 (<https://youtu.be/r9UsOFSZCeQ>; last accessed July 2021).

229 https://www.bilibili.com/video/BV1vT4y1c7Pd?spm_id_from=333.905.b_72656c61746564.2 (last accessed July 2021)

230 <https://www.getit01.com/p20180122868764110/2> (last accessed Jul 2021; in Mandarin); <https://www.zhihu.com/question/68764110/answer/913256432> (last accessed Jul 2021; in Mandarin)

231 See “海口過去比湛江落後，現在卻超過了湛江，說明了什麼問題” (<https://kknews.cc/history/ra4xa9n.html>, last accessed July 2021). See also Luichew speaker 藍色天空’s reply to XH H’s comment on the video “广东雷州话防疫宣传，据说台湾人也能听懂八成” (https://youtu.be/OkS_VhqYyQg, last accessed July 2021): s/he and a Hainanese-speaking friend converse in each their own language; people who know the friend (and don’t speak Hainanese) will ask if s/he is Hainanese too. It is not clear whether such a scene could unfold given a Hainanese speaker and a Luichew speaker with negligible prior exposure to the other variety.

232 See “同一片火山 同一種民俗” (<https://kknews.cc/other/kn5nzjq.html>, last accessed July 2021), according to which the respective Hainanese and Luichew dialects adjacent to the strait that divides the two “identities” are m.i.; however, at Luichew City, the Hainanese reporters were able to understand the local dialect, while the locals could not understand Hainanese. See https://zhidao.baidu.com/question/151031558.html?qbl=relate_question_4 (last accessed July 2021).

233 <https://zhidao.baidu.com/question/413089682.html> (last accessed July 2021; in Mandarin)

between the two dialects.²³⁴ Hainanese intra-intelligibility is marginal (see §2.2).

Regardless of m.i., Hainanese and Luichew are ethnolinguistically distinct. The Luichew tongue is never implied to be a type of Hainanese, nor vice versa. “Luichew–Hainanese,” à la Serbo-Croatian, is unheard of. In Muar, Malaysia, there is a Luichew association (Persatuan Lui Chew Johor) as well as a Hainanese association. In Malacca, there is a Luichew association and several Hainanese associations. As suggested in the story of the founding of Persatuan Lui Chew Johor,²³⁵ if a Luichew speaker wound up in a town with no Luichew community, s/he would probably join the Canton 廣東~广东 association rather than the Hainanese association.

To the best of our knowledge, there has never been a Luichew translation of the Bible. As for whether Luichew speakers ever made use of the Hainanese Bible, it is plausible that it was at least attempted—in the districts closest to Hainan—since the gospel seems to have arrived in the Luichew region via Hainan.^{236 237}

Hainanese and Luichew are ethnolinguistically distinct and generally not m.i. Hainanese warrants a separate code. While Hainanese and Luichew may form an intelligibility continuum, so do Bengali [ben] and Assamese [asm], or Occitan [oci] and Catalan [cat]. It is hard to justify assigning Hainanese and Luichew to the same code.

Since Hainanese is separate from Luichew, it is obviously separate from Teochew, Haklau and Samheung. We compare Luichew to Teochew and to Samheung in §4.3.

4.3—Luichew

Luichew’s next of kin is Hainanese. As discussed in §4.2, Luichew and Hainanese are ethnolinguistically distinct and generally not m.i. Luichew’s next closest linguistic relative is Teochew. We also compare Luichew to Samheung because Luichew may be Samheung’s next of kin.

Vs Teochew

Luichew—at least the non-implosivized varieties—appears to have some (non-functional) m.i. with Teochew,²³⁸ perhaps to the same extent as Spanish [spa] and, say, Venetian [vec],²³⁹ but not to the degree to which the Tunisian [aeb] and Egyptian [arz] vernaculars are m.i. (see §3.1). One Luichew speaker’s assessment of Teochew intelligibility (in Mandarin; we translate; emphasis added) is revealing, and not just of Teochew–Luichew m.i.:

I am a Luichewite. I have not known any [Hokkien speakers], but I have known quite a few [Teochew speakers]. I can catch some of what’s being said in Teochew. From listening to Hokkien songs *while reading the lyrics*, though, Hokkien seems to be easier to understand.²⁴⁰ Sometimes when people are speaking Teochew, it sounds like they’re speaking Thai. What I mean is, I watch Thai horror movies sometimes, and the feeling I get from hearing Thai—which I don’t understand—is kind of like the feeling I get hearing Teochew when I’m not understanding it. Now ... if you’re acquainted with many dialects [in your own language], and you have a good ear... An older Luichew speaker, they’ll catch more of what’s being said in Teochew, in context. Now ... someone I know from my hometown, s/he thinks Luichew and Teochew are totally different. But I can under-

234 <https://zh.wikivoyage.org/zh-mo/海南話會話手冊> (last accessed Aug 2021)

235 <http://www.luichew.com> (last accessed Aug 2021; in Mandarin)

236 See 陳國威 (2019), “廣東雷州半島基督教發源地的文獻與田野調研——兼論湛江基督教部份史實” (http://nansha.schina.ust.hk/Article_DB/sites/default/files/pubs/news-094.02.pdf, last accessed Aug 2021).

237 The “雷州話聖經” (Luichew Bible) app on Android (<https://play.google.com/store/apps/details?id=org.share35.AudioBibleLeiZhou>, last accessed Aug 2021) provides recordings of a Mandarin translation of the Bible either read into Luichew (as described in §4) or, more likely, read untranslated using Luichew readings.

238 <https://www.zhihu.com/question/24516606> (last accessed July 2021; in Mandarin); <https://www.zhihu.com/question/68764110/answer/913256432> (last accessed July 2021; in Mandarin)

239 <https://youtu.be/yfxiSpQrmu8> (last accessed Aug 2021)

240 Hokkien songs are poor subject matter for gauging intelligibility. Fully vernacular lyrics are rare, for complex reasons. Most lyrics are written in a stylized inter-language bridging Hokkien, Mandarin (the de facto written koine of the ethnos) and Classical Chinese (the former written koine). Additionally, most songs are slower than natural speech. If the lyrics are in the field of vision even peripherally, the futility of drawing conclusions about intelligibility is complete. Despite their relative sophistication, our now-anonymous commentator failed to take these factors into account; s/he is certainly far from alone in having drawn spurious inferences on intelligibility from music videos online. It may be helpful to consider this example together with the COVID announcement described in §4.

stand Teochew some. I can get the gist, if nothing else... Maybe it's because I [am very familiar with two dialects of Luichew], plus I'm interested in this stuff—that's why I can catch quite a bit.^{241 242}

Luichew-Teochew m.i. falls far short of Luichew-Hainanese m.i., and Luichew speakers and Teochew speakers do not share a common ethnolinguistic identity (besides *Chinese*), although—for speakers in or from China—there is a shared provincial identity. It is almost certain that if a Luichew speaker emigrated to a town with no Luichew association, they would not join the Teochew association but rather the Cantonese association (see §4.2).

Luichew and Teochew are ethnolinguistically distinct and not m.i. Luichew warrants a separate code from Teochew.

We compare Teochew to Haklau, Chawan and Hokkien in §4.4.

Vs *Samheung*

There are no reports to be found online on m.i. between Luichew and Samheung. Most Luichew speakers probably do not know the Samheung speech variety exists; it is safe to say that if there was a lost dialect of Luichew spoken on the outskirts of Macau, generations of Luichew migrants to the Pearl Delta would have discovered it by now.

We could not help noticing, in Bodman's²⁴³ data, that this is how you say A BOWL OF RICE (one-bowl-rice) in two (m.i.) dialects of Samheung (tones stripped by us for simplicity):

/tit oa pui/
/tit ə pui/

Translating word for word into Luichew City Luichew,²⁴⁴ again with the tones stripped, we get:

/*ziak ua pui/

Lacking personal and social knowledge of Luichew, we are not 100% sure this is well formed, idiomatic

241 Yet another sophisticated observation. Anecdotally, it seems obvious that many affirmations of m.i. between [nan]-associated varieties are powered by individuals who take an unusual interest in the matter. This doesn't seem to be the case in Europe, for example, where commentators tend to be candidly matter-of-fact about whether *social consensus* deems varieties A and B to be m.i.

242 <https://www.zhihu.com/question/24516606/answer/110955625> (last accessed Jul 2021; in Mandarin)

243 Nicholas Bodman (1988), "Two Divergent Southern Min Dialects of the Sanxiang district, Zhongshan, Guangdong"

244 Using the 雷州方言詞典, published by 江蘇教育出版社.

Luichew. But since Hokloid languages often look more alike on paper than they really are, this does not bode well for Luichew-Samheung m.i.

As a practical matter, m.i. is inherently a social matter. It cannot be reliably determined without input from actual speakers of the languages involved. M.i. cannot be macro-determined "centrally," in bulk.

By all accounts, Samheung is heavily Cantonized.²⁴⁵ Luichew and Samheung would probably lack m.i. due to this alone.

It can be inferred that Luichew and Samheung are ethnolinguistically distinct and not inherently m.i. Luichew warrants a separate code from Samheung.

4.4—Teochew in Contrast

Teochew's linguistic next of kin are Haklau, Hokkien, and Chawan. As discussed in §4.3, Teochew and Luichew are ethnolinguistically distinct and functionally not m.i.

Vs *Haklau*

There are reports that Teochew and Haklau are highly m.i.²⁴⁶ As discussed in §2.5, though, a 2001 Hong Kong trial was derailed for a day because the Haklau-speaking defendants were not able to understand the Cantonese-Teochew interpreter on a functional level. Since both varieties are sometimes called HOKLO 學老 in Cantonese, the court had assumed m.i. between Haklau and Teochew.

The anecdotal evidence is not hard to reconcile. Reports of m.i. would tend to come from individuals with consistent, repeated exposure to the other variety via low stakes situations, typically with a focus on topics familiar to the conversants. Individuals with limited exposure to the other variety, engaged in high stakes situations or on unfamiliar subject matter, would find the other variety functionally unintelligible.

With the last 40 years of social change in Guangdong, a Teochew speaker and a Haklau speaker unable to converse heterolingually in their own languages would switch to Cantonese or Mandarin without further ado.

The ethnolinguistic distinction between Teochew and Haklau is unambiguous. The Teochew linguistic tribe has never included Haklau speakers, who

245 See <https://www.bilibili.com/s/video/BV1Jt411a7Yu> (last accessed Aug 2021; interface in Mandarin), comparing Samheung vocabulary with Cantonese and Hokkien vocabulary.

246 <https://www.zhihu.com/question/68764110/answer/381762966> (last accessed Aug 2021; in Mandarin); <https://www.zhihu.com/question/68764110/answer/913256432> (last accessed Jul 2021; in Mandarin)

belonged to a different prefecture in late antiquity. The distinction was explicitly observed overseas in northern Sumatra. Haklau speakers do not learn Teochew as a matter of course, but rather Cantonese; Haklau speakers in diaspora may lose their Haklau, but they do not replace it with Teochew the way diaspora Hoisan 台山 (Mand. *Táishān*; sometimes romanized via Cantonese as *Toisan*) speakers sometimes replace their Hoisan with Cantonese (i.e. even in environments where Cantonese is not dominant).

Teochew and Haklau are partially m.i. but ethnolinguistically distinct, à la Minangkabau [min] and Indonesian [ind], or Hiligaynon [hil] and Cebuano [ceb] at best. Teochew warrants a separate code from Haklau.

We compare Haklau to Hokkien in §4.5.

Vs Hokkien

Numbers are subjective, but the consensus in the China context seems to be that Teochew and coastal Hokkien are roughly half to three quarters m.i. with some exposure, depending on the dialect of Hokkien.²⁴⁷ Subjectively, again, this puts Teochew-Hokkien m.i. on the borderline of low-stakes functionality. Any individual Teochew or Hokkien speaker with the need or desire can make things work by paying extra attention for a few weeks or months. A realist observation is that actual Teochew-Hokkien m.i. has declined with the last generation or two.²⁴⁸

Again, an anecdote from the Hong Kong justice system goes a long way toward pinpointing where “Hoklo-Hoklo” m.i. begins and ends. As one Mr. Chiu related in 2007:

In Hong Kong, the official term for Teochew is Chiuchow. I think I can speak authentic Chiuchow because my late parents were immigrants to Hong Kong from Chaoyang city (潮陽市) [Teoyeo].

I still remember how embarrassed I was during an Immigration Tribunal meeting held inside the Victoria Prison (now closed) sometime in the early 1980s. I was the interpreter responsible for doing interpretation from English (for the benefit of two expatriate lawyers) to Chiuchow and vice versa. It turned out that the teenaged defendant (illegal immigrant) was a Fujianese who could speak only Hokkien. I could understand only about 50% of what he said. By then, it was too

late to find a Hokkien interpreter and I went through the meeting on tenterhooks. Fortunately, the defendant could also understand a bit of Chiuchow and the meeting could manage to carry through to the end somewhat unsmoothly.

Today, I am living in a district populated mainly by Fujianese. And I still can understand only about half of what they are speaking.²⁴⁹

Spanish [spa] and Portuguese [por] are excellent analogs for Hokkien and Teochew, right down to the wildly restructured overseas dialects, arguably creoles; Chawan plays the part of Galician [glg].

As far back as we can see at this point, Teochew and Hokkien orthography and literature have never been undifferentiated. However, the traditional Teochew Sino-script and the traditional Hokkien Sino-script work(ed) the same way and are, or were, very similar. The 1566 edition of *Lychee Mirror* (重刊五色潮泉插科增入詩詞北曲勾欄荔鏡記戲文全集)—to date the oldest extant Hokkien text—is partly in Teochew.²⁵⁰

In the 19th century, separate missionary scripts were devised for Hokkien and Teochew, and translations of the Bible proceeded independently.

In Singapore and Malaysia, the news is reported separately in both Teochew and Hokkien. See §2.4 and §2.6.

Teochew and Hokkien are ethnolinguistically very distinct, although there are linguistic transition zones between the two. On the island of Namoa, which was bifurcated between Guangdong and Fujian during late antiquity, the ex-Guangdong side speaks Teochew while the ex-Fujian side speaks Hokkien internally and Teochew—the lingua franca—externally. Namoa Hokkien has converged toward Teochew, but its persistence illustrates the distinctness of Hokkien versus Teochew.

Special mention must be made of the situation in Singapore, Riau (Indonesia), and southern Peninsular Malaysia. Here the Teochew is Hokkienized, the Hokkien is Teochewed, and both are widely spoken. Although it is taken for granted here that Teochew and Hokkien are different languages, it has been observed that no matter whether people are trying to speak the local Teochew or the local Hokkien, they wind up speaking what Lee Kuan Yew

247 <https://www.zhihu.com/question/68764110/answer/913256432> (last accessed Aug 2021; in Mandarin)

248 <https://www.zhihu.com/question/68764110/answer/718725191> (last accessed Aug 2021; in Mandarin)

249 <http://www.cantonese.sheik.co.uk/phorum/read.php?16,67226,page=1> (last accessed Aug 2021)

250 The female lead character is from Teochew City. Much of the story takes place there.

called “a Hokkien-Teochew hybrid.”²⁵¹ As with the formation of Amoy Hokkien and Taiwanese, much of the convergence must have been subtractive. The Hokkien spoken here is more comprehensible to Chinese Teochew speakers than Chinese Hokkien,²⁵² despite the Malay and English loans. And while Taiwanese speakers generally have trouble understanding non-Amoy Hokkien (and Teochew), the Singaporean *Money No Enough* movies were viewed and appreciated in Taiwan—with subtitles, obviously—and bits of wickedly crude, Teochew-filtered Hokkien banter made a lasting impression.

Teochew and Hokkien are partially m.i. and ethnolinguistically distinct, like Spanish [spa] and Portuguese [por]. They have distinct orthographies and literary traditions, like Fante [fat] and Twi [twi]. Like Twi from Fante, or Spanish from Portuguese, Teochew warrants a separate code from Hokkien.

Vs Chawan

Teochew is highly comprehensible for Chawan speakers.²⁵³ Intelligibility of Chawan by Teochew speakers should at least be at the high end for Chiangchew Hokkien, although first-hand reports are hard to come by.

Modern Teochew and Chawan speakers are not so much ethnolinguistically distinct as politically distinct. Even in Sarawak and Singapore, where Teochew and Chawan are ethnolinguistically distinct, provincial identity may have driven the distinction from the start. Early Chawan-speaking expats in Sarawak and Singapore may have seen themselves as culturally and linguistically distinct from Hokkien speakers, but politically distinct from Teochew speakers; hence the need for Chawan institutions. However, unlike Galician, Chawan has never had its own written form, and Chawan ethnolinguistic identity is nowadays indistinct in its original and most important locale (§2.7).

Teochew and Chawan are partially or largely m.i. and politically, more than ethnolinguistically, distinct. It is tempting to group Chawan with Teochew, but Chawan does not clearly skew Teochew, and Chawan speakers see Chawan more as a dialect of Hokkien. We will address this more thoroughly

in §4.6. Teochew warrants a separate code from Chawan.

4.5—Haklau in Contrast

Haklau’s next of kin are Teochew, Taiwanese, and Hokkien. As discussed in §4.1 and §4.4, respectively, Haklau has limited m.i. with Taiwanese and Teochew, but is ethnolinguistically distinct from both.

Vs Hokkien

Haklau and Hokkien are partly to substantially m.i.²⁵⁴ Some scholars and enthusiasts classify Haklau as Chiangchewish Teochew, while others see it as Teochewish Chiangchew Hokkien; we would expect Haklau-Hokkien m.i. to surpass Hokkien-Teochew m.i., and there are reports confirming this.²⁵⁵ There are also reports of non-intelligibility between Haklau and Hokkien.²⁵⁶ We would expect perceived intelligibility to vary based on the dialects of Hokkien and Haklau involved, whether the communication is direct, the depth of the listener’s knowledge of his or her own language, and expectations (or stakes). Through exposure to entertainment media, Haklau speakers probably understand Hokkien (at least Amoy and Chiangchew Hokkien) better than Hokkien speakers understand Haklau.

Overall, Haklau and Hokkien seem to be less m.i. than Haklau and Teochew. The ethnolinguistic distance between Haklau and Hokkien is also greater than that between Haklau and Teochew. Haklau warrants a separate code from Hokkien.

251 <https://www.quora.com/How-different-is-Singaporean-Teochew-from-Singaporean-Hokkien-language> (last accessed Aug 2021); <https://www.asiaone.com/News/The%2B-New%2BPaper/Story/A1Story20090319-129731.html> (last accessed Aug 2021)

252 <https://www.zhuhu.com/question/68764110/answer/913256432> (last accessed Jul 2021)

253 See <https://www.getit01.com/p20180122868764110> (last accessed Jul 2021) and §2.7.

254 There are surprisingly few first-hand reports of Haklau-Hokkien m.i. on the Chinese internet. Apparently, for economic reasons, Haklau speakers do not visit Hokkien-speaking regions, and Hokkien speakers do not visit the Haklau-speaking region. Comparisons on the Hong Kong internet would be confounded by the inconsistent, overlapping names of the two varieties. See <https://www.getit01.com/p20180122868764110/2> (last accessed Aug 2021; in Mandarin); <https://bbs.hupu.com/27384337-3.html> (last accessed Aug 2021; in Mandarin); <https://www.quora.com/How-mutually-intelligible-is-Hokkien-and-Hoklo-Is-there-a-comparison-one-between-the-two-European-languages> (last accessed Aug 2021).

255 <https://zhidao.baidu.com/question/616480965832712532.html> (last accessed Aug 2021; in Mandarin)

256 <https://zhidao.baidu.com/question/157394085855422300.html> (last accessed Aug 2021; in Mandarin)

4.6—Hokkien in Contrast

Hokkien is linguistically close to Taiwanese, Teochew, Haklau, and Chawan. Hokkien is probably linguistic next of kin for Lengna and the Datian Frontlect as well. As discussed in §4.1, §4.4, and §4.6, respectively, Hokkien is distinct from Taiwanese, Teochew, and Haklau.

We discuss m.i. between Hokkien and Kelantan Local Hokkien in §4.12.

Vs Chawan

Hokkien is linguistically diverse within itself, but Chawan speakers should have high to full comprehension of Chiangchew Hokkien (see §2.7). Chawan speakers cannot effectively comprehend certain Choanchew dialects of Hokkien,²⁵⁷ and Amoy Hokkien speakers may struggle to understand the Hokkien dialects neighboring Chawan, let alone Chawan.²⁵⁸ Chawan speakers would probably struggle to understand raw Penang-Medan Hokkien (see §2.6).

Hokkien and Chawan are or have been ethnolinguistically distinct in Sarawak and Singapore. The cultural uniqueness of Chawan speakers within Chiangchew suggests that their forebears may have been ethnolinguistically distinct versus Hokkien speakers during some period of antiquity. However, Chawan is today considered to be a variant of Hokkien in China (see §2.7), where it is mostly spoken. In Singapore, Chawan is also understood to be a special dialect of Hokkien (§2.7).

Hokkien at large and Chawan are somewhat to highly m.i., depending on the dialect of Hokkien, but Chawan is considered to be a dialect of Hokkien by nearly all its speakers, as well as those who interact with them. We tentatively recommend that Chawan be considered a dialect of Hokkien for the purposes of ISO 639-3.

Vs the Datian Frontlect

Inherent m.i. between Hokkien and the Datian Frontlect is stressed at best, and poor for many individuals, if not most.²⁵⁹ Datian Hokkien might be a marginal exception.

Ethnolinguistically, a well formed, stable Datian Frontlect-speaking identity does not exist. Some Datianese identify as *Mǐnnán* 閩南~閩南 (≈ Hokkien).

257 <https://www.getit01.com/p20180122868764110> (last accessed July 2021; in Mandarin)

258 <https://www.getit01.com/p20180122868764110/2> (last accessed July 2021; in Mandarin)

259 <http://www.manyanu.com/new/4e150d143bd14e928cb-c06a802b8f636> (last accessed July 2021; in Mandarin); <https://www.zhihu.com/question/278258575/answer/721203163> (last accessed July 2021; in Mandarin)

Dàtián Hòushēngzǎi (大田後生仔~大田后生仔), a huge hit sung partly in the Frontlect and partly in Mandarin, has been universally classified as a Hokkien song with no objection from either Frontlect speakers or Hokkien speakers generally.

Hokkien and the Datian Frontlect are not functionally m.i. for the most part, but a universal consensus is forming that the Frontlect is a dialect of Hokkien for social, political, and linguistic reasons. Hokkien is an ethnolinguistic concept anyway; Hokkien is not functionally intra-intelligible. We tentatively recommend that the Datian Frontlect be grouped with Hokkien for the purposes of ISO 639-3, the way Aranese [oci] is grouped with Occitan [oci].

Vs Lengna

Hokkien and Lengna are not functionally m.i.,²⁶⁰ although they are similar enough for the heterolingual listener to get the gist of what is being said.²⁶¹ Lengna is not functionally m.i. with the speech of Zhangping,²⁶² spoken just 70 km downriver but linguistically a Chiangchew dialect of Hokkien.

Lengna speakers do not consider themselves Hokkien speakers.²⁶³ The Lengna ethnolinguistic identity is well defined, as suggested by the existence of the local Lengna-language talk show 我是龙岩人 (“I am Lengneño”).

Hokkien and Lengna are ethnolinguistically distinct and not functionally m.i. Hokkien warrants a separate code from Lengna.

4.7—Chawan in Contrast

As discussed in §4.4 and §4.6, linguistically, Chawan is a transition between Teochew and Hokkien; ethnolinguistically, it is now a dialect of Hokkien. We tentatively recommend that Chawan be considered a dialect of Hokkien for the purposes of ISO 639-3.

4.8—the Datian Frontlect in Contrast

Within the [nan] sphere, the Datian Frontlect is linguistically closest to Hokkien and Lengna.²⁶⁴ As

260 <https://www.zhihu.com/question/35923258/answer/74670453> (last accessed July 2021; in Mandarin)

261 <https://www.zhihu.com/question/35923258/answer/351813173> (last accessed July 2021; in Mandarin)

262 <https://youtu.be/-TnWLhVuabc> (last accessed Aug 2021; in Mandarin)

263 <https://www.zhihu.com/question/35923258/answer/66040251> (last accessed Aug 2021; in Mandarin)

264 Based on phonological shared retentions, David Prager Branner grouped the Frontlect together with Lengna and the dialect of Zhangping into a subgroup called “Northern Miinnan” or “Inland Miinnan.” See David Prager Branner (1999), “The Classification of Longyan.” In the same historical or genetic vein, Bit-chee Kwok grouped the Frontlect and Lengna with Chiangchew Hokkien. See Bit-chee Kwok (2018), *Southern Mǐn: Comparative Phonology and Subgrouping*.

discussed in §4.6, the Datian Frontlect is ethnolinguistically indistinct from Hokkien. We tentatively recommend that the Frontlect be grouped with Hokkien for the purposes of ISO 639-3, at least the way that Aranese [oci] is grouped with Occitan [oci].

Vs Lengna

Given the graduated nature of inland settlement and how far it is from Datian to Lengna, we suspect the Frontlect and Lengna are not inherently m.i. to a functional degree. Nowadays Mandarin would bridge the gap, however small.

The publicly available evidence does not indicate m.i. between the Frontlect and Lengna. Given the ethnolinguistic reality as it appears, there is reason to group the Frontlect with Hokkien and not with Lengna.

4.9—Lengna in Contrast

Lengna's linguistic next of kin are Hokkien and perhaps the Datian Frontlect. As discussed in §4.6 and §4.8, Lengna is distinct from both and warrants a separate code.

4.10—Lungtu & Namlong in Contrast

Lungtu's linguistic next of kin is Namlong, and vice versa. As discussed in §2.10, the two varieties are not m.i.

Lungtu and Namlong, along with Samheung, are what Sino-topolectologists call “Zhongshan Min” (中山閩語~中山閩語). In fact, Lungtu and Namlong probably wound up under [nan] because a critical mass of linguists assumed they were something like what Samheung really is. “Zhongshan Min” does not seem to have a correlate in the social reality. While there is a shared county identity, a shared provincial identity (Guangdong), and possibly a shared ancestral-province identity (Fujian), Lungtu and Namlong (and Samheung) speakers do not seem to share an ethnolinguistic identity below the level of Chinese.

Lungtu and Namlong are ethnolinguistically distinct and not m.i. Each warrants a separate code.

Vs Samheung

Neither Lungtu nor Namlong is m.i. with Samheung. As discussed above, Samheung is ethnolinguistically distinct from both Lungtu and Namlong as well. Lungtu and Namlong warrant separate codes from Samheung.

4.11—Samheung in Contrast

Samheung has no close kin. As discussed in §4.3, Samheung is distinct from Luichew, perhaps the closest of its distant kin. As discussed in §4.10, Samheung is distinct from Lungtu and Namlong. Samheung warrants a separate code.

4.12—Kelantan Local Hokkien in Contrast

Kelantan Local Hokkien²⁶⁵ is a mix of Hokkien, Siamese, and two varieties of Malay. It does not have close kin.

Vs Hokkien

The Kelantan Local is not m.i. with Hokkien, including the in-town Hokkien of Kota Bharu, the capital of Kelantan. When asked, a Penang Hokkien-speaking correspondent said he could understand about half of the Kelantan Local Hokkien monologue in this²⁶⁶ Youtube video. A Chinese or Philippine Hokkien speaker would be able to pick out a few words at best, and most likely would not suspect that this language had a Hokkien element unless told beforehand. Local Hokkien speakers, for their part, do not learn or understand the “straight” Hokkien spoken in Kota Bharu and around Malaysia.

As discussed in §2.12, Kelantan Local Hokkien and Hokkien are ethnolinguistically distinct. Based on their words, actions, and interactions, Local Hokkien speakers arguably do not even share in the *Chinese* ethnolinguistic identity, although they call themselves *Tng lang* (“Chinese people”; tones not marked herein). Local Hokkien probably lies outside the Chinese macrolanguage entirely, à la Dungan [dng].

Kelantan Local Hokkien is ethnolinguistically distinct from Hokkien. It is not m.i. with any other language, including Penang Hokkien. It warrants a separate code.

265 Again, we exclude the in-group dialect of the upriver settlements. That dialect is sociolinguistically Local Hokkien, but linguistically Hokkien.

266 <https://youtu.be/ZeSTpWVzTwk&t=44s> (last accessed Aug 2021)

5—Conclusion

Based on the research and reasoning presented in the previous sections, it would be most logical to reanalyze [nan] into 11 ISO 639-3 codes spanning 12 languages, as follows:

English Name	Autonym	Name for Code	Code
Hainanese	Hái-nâm		hnm
Haklau	海陸丰	HAI LOK HONG	hlh
Hokkien	Bân-lâm		hbl
Chawan			
Datian Frontlect	大田前路话	—	
Kelantan Peranakan Hokkien ²⁶⁷	Hokkien Kelantan		hkl
Liongna ²⁶⁸	龙岩	LIONGNA	lnx
Longdu ²⁶⁹	隆都	LONGDU	zld
Luichew	雷州	LUICHEW	luh
Namlong	南朗	NAMLONG	znl
Samheung	三乡	SAHIU	zsh
Taiwanese	Tâi-oân		oan
Teochew	Tiô-Sòa ⁿ		tws

Of the 11 codes, [hnm], [luh], [tws], [hlh], [hbl], [lnx], [zld], [znl], and [zsh] fall within the Chinese [zho] macrolanguage. Codes [oan] and [hkl] arguably fall within [zho] as well; we would argue the opposite, since Dungan [dng] is not listed under [zho].²⁷⁰ The basis of [zho] is clearly not *genetic*, nor is it based on historical association with Literary Chinese [lzh].

Appendix A-1—Names of Taiwanese

In Taiwanese, Taiwanese is most commonly called *Tâi-oân-ōe* 台灣話 (“Taiwan speech”) or *Tâi-gí*²⁷¹ 台語 (“the Taiwan language”). The corresponding Mandarin names are *Táiwānhuà* and *Táiyǔ*. English *Taiwanese*, anglicized via Mandarin, corresponds to all of the above. For ease, we will refer to this set of names as TAI.

A more ancient but much less common name is *Hō-ló-ōe* (“Hoklo speech”; etymological spelling *Hòh-ló-ōe*), corresponding to Hakka *Hók-ló-fa* (“Hoklo

speech”), occurring in English as *Holo* or *Hoklo*, sometimes in combination with Taiwanese. We will refer to this cognate set as HOKLO.

The name *Hō-ló* is used, and not exclusively, by people from areas with Hakka settlements. People from areas without Hakka settlements may reach adulthood without having heard it. *Hō-ló* tends to be used when the speaker especially wishes to refer to Taiwanese speakers without implying that the Hakka are less Formosan.

An objection was raised to the TAI set of names starting in the 1960s, led by intellectuals associated with the Republic of China dictatorship. Fearing that the concept of a Taiwanese language would eventually enable Taiwanese independence (and in turn terminate the Republic of China), the state began to refer to Taiwanese as *Mǐnnán* 閩南 (“southern Fujianese”)—basically Hokkien. As for why the modern Chinese states call Hokkien *Mǐnnán*, we touch on that in [Appendix A-6](#).

A second objection to TAI was raised starting in the 1990s, led by intellectuals associated with the Republic of China democracy. They contended that the TAI set of names marginalized Hakka, Amis, Atayal, and the other roots languages of Formosa by implying that they were less *Taiwanese* than Taiwanese, and that TAI should refer to all the languages of Formosa in the aggregate.

A grassroots solution was to replace TAI with HOKLO, at least in writing. The problem was that HOKLO does not have a stable Sinographic written form,²⁷² nor—as a result—a stable Mandarin translation. The option of writing HOKLO (HOLO) in romanization and not translating it into Mandarin was yet another non-starter for many, not least those who objected to TAI.

Attempts were made to represent HOKLO as 福佬 or 河洛, in turn drawing objections that the Sinographs were derogatory (佬), Chinese nationalist (河洛), or spurious (which they probably are). Others objected to HOKLO on the grounds that they had never heard it growing up, or that the Hakka used

²⁷² *Hō-ló* and Hakka *Hók-ló* do not have etymological Sinographs. Traditionally, they were rarely written; if they had to be written, sound-borrowings would suffice. In Miaoli 苗栗, for example, there is a place called (in Hakka) *Hók-ló-phài* 學老排 (“Hoklo slope”). From the 1910s on, with the institutionalization of Chinese nationalism, the South Chinese and Formosan literati began to take an interest in etymological connections between Sinographs and vernacular words; manufacturing connections became a pastime in step with their newfound belief that their speech (and their genes) must be of pure Middle China extraction. Two popular etymologies eventually emerged for HOKLO in the form of 福佬 and 河洛—the first unproven (as far as linguistic science goes), the second unproven and implausible.

²⁶⁷ See [Appendix A-12](#).

²⁶⁸ See [Appendix A-9](#).

²⁶⁹ See [Appendix A-10](#).

²⁷⁰ <https://iso639-3.sil.org/code/zho>

²⁷¹ *Tâi-gú* in some dialects; *Tâi-gŭ* in yet others, if we take the liberty of borrowing a vowel from the Hakka missionary script.

the word much more. Over time, HOKLO became welded to 福佬 and 河洛 in the public consciousness, and the objections merged and hardened.²⁷³

Some argue that it is in the natural order of things for a majority language (and tribe) to carry the name of the locale. Others suggest that, historically, the tribe was not named after the island. The island was named after the tribe, mostly by colonizing authorities,²⁷⁴ which is why the *Tâi-oân* tribe in its entirety has never had any other name.²⁷⁵ Still, it might be wishful—and objectionable to many—to suppose that the island and territory could be renamed (to *Formosa*, for example). Instead, Tō' Kiàn Hong 杜建坊 and others have proposed switching the Sinographs used for *Tâi-oân* to 台員.²⁷⁶ This makes little difference within the logic of Taiwanese itself, but different Sinographs yield different readings in Mandarin and Hakka; from any point of view but a *Tâi-oân*-centric one, the *Tâi-oân* tribe will have sacrificed, renaming itself.²⁷⁷

ISO 639-3 codes [tai] and [tan] are already in use. Considering that switching Sinographs for *-oân* could deflate the naming dispute that surrounds Taiwanese, we find it serendipitous that ISO 639-3 code [oan] is available, and ask that it be adopted for Taiwanese.

Appendix A-2—Names of Hainanese

The Hainanese name for the Hainanese language is *Hái-nâm-oe*²⁷⁸ (“Hainan speech”).²⁷⁹ The corresponding Mandarin name is *Hǎinánhuà* 海南話~海南話. English *Hainam* and *Hailam* are derived from the

273 In a global context, HOKLO is ambiguous as well. See §2.5.

274 The *Tâi-oân* people began modestly in the 1600s as a seaside trading village named *Tâi-oân* in what is today the borough of Anping. *Tâi-oân* would broaden in scope as the tribe intermarried, multiplied and fanned out. Toward century's end, the Qing empire annexed the sections of Formosa inhabited by the *Tâi-oân* people and named the territory *Tâi-oân* 臺灣—Mandarin *Táiwān*. Yet the name *Tâi-oân* stuck to the slice of the island where it had started out. Till the end of the 19th century, *Taiwan* in English generally referred to the city of Taiwan, which later became *Tainan*. The island was mainly called *Formosa* in English till the 1960s, when the occupying Republic of China moved to change the name to *Taiwan*, around the same time that they began to call Taiwanese *Mínnán*.

275 A circular second name has gained currency in recent years: *Tâi-gí-lâng* 台語人 (Mand. *Táiyǔrén*, “people of the TAI language”). It is likely that this name would be contested if it became more popular.

276 *Tâi-oân* was variously written 大員 臺員 大灣 臺灣 during the 1600s. All were sound-borrowings.

277 In Hakka, for example, 台員 reads *Thòi-yàn*, distinct from *Thòi-vàn*, which is the Hakka name for Formosa.

278 See <https://en.wikipedia.org/wiki/Hainanese> (last accessed Aug 2021). This appears to be in the missionary script.

279 There might be dialect variation between N and L; the M is stable across all dialects of Hainanese, to the best of our knowledge.

Hainanese; *Hainan*, an anglicization of the Mandarin, has won out in English. Likewise, *Hainanese* is the common form in English; *Hainamese* and *Hailamese* are rarely seen.

Hainanese has a poetic name in Classical Chinese: 瓊語~琼语 (“the Hainan language”), read *Qióngyǔ* in Mandarin or *Quỳnh Ngữ* in Vietnamese. We are not sure exactly how it would be read in Hainanese. The HAINAN series of names are heavily favored in both Hainanese and Mandarin, as far as we can tell.

Reminiscent of the naming controversy surrounding Taiwanese, the name *Hǎinánhuà* 海南話 has begun to be disputed on the Chinese internet. There is an argument that *Hǎinánhuà* 海南話 should refer to all the languages of Hainan in the aggregate. We salute the egalitarian seed of the argument, but a real-world alternative has not arisen,²⁸⁰ and the names of any number of languages could be challenged on the same grounds, including *Zhōngwén* 中文 and *Huáyǔ* 华语, which currently refer to a mere subset of the languages of China, and at times just Mandarin.

ISO 639-3 code [hnm] seems like a perfect fit, and remains unused; [hnn] is taken. In any case, it makes sense that the autonym should take precedent over English usage or the Mandarin translation.

It is well advised that, while our suggestions follow from the preceding analysis, we do not have extensive personal experience of Hainanese-speaking society.

Appendix A-3—Names of Luichew

Luichew consists of two main speech varieties known colloquially as, roughly, LOI²⁸¹ and HAI.²⁸² We are not sure to what extent Luichew speakers refer to both together as LUICHEW²⁸³ (the language), or if LUICHEW only refers to LOI.

The Mandarin name corresponding to LUICHEW is *Léizhōuhuà* 雷州話~雷州話. English *Luichow* is an anglicization of the Cantonese equivalent. *Loi Chau*, less commonly seen, comes from Vietnamese *Lôi Châu*.

Since ISO 639-3 codes [lui] and [luy] are taken, we looked at other codes starting with L-U; [luh] seems

280 琼语 *Qióngyǔ* would have the exact same problem.

281 [loi] in Luichew City. See page 167 of the 雷州方言詞典 (江蘇教育出版社).

282 We are not sure how this would be pronounced in the Sea Speech itself. The Luichew City pronunciation would be [hai]. See page 161 of the 雷州方言詞典 (江蘇教育出版社). The Sea Speech pronunciation is almost certain to start with [h] or similar.

283 Luichew does not have a romanized script. We use all-caps forms in two ways to fill the gap: HAI indicates an approximate romanization. LUICHEW signifies “the equivalent of (English) Luichew.”

to be the next best option, keeping the L-U for LUI-, and bringing in the H for HAI.

It is well advised that, while our suggestions follow from the preceding analysis, we do not have extensive personal experience of Luichew-speaking society.

Appendix A-4—Names of Teochew

The traditional Teochew name for the Teochew language is *Tiô-chiu-ūe* 潮州話 (“speech of the prefecture of Teo”). The corresponding Mandarin name is *Cháozhōuhuà* 潮州话~潮州話. English *Teochew* is a direct anglicization of the native *Tiô-chiu*. *Chiu Chow*, often seen, was anglicized via Cantonese. *Trieu Chau*, less frequently seen, comes from Vietnamese *Triều Châu*.

The Classical Chinese abbreviation for *Tiô-chiu* is 潮語~潮语 (“the Teochew language”), read *Tiô-gŭ* in Teochew. This name is rarely used in speech or colloquial writing, as far as we can tell.

The old Teochew prefecture no longer exists under the current administrative hierarchy of China. There is a smaller *Tiô-chiu* or *Cháozhōu* 潮州 City centered on the city that had been the seat of Teochew prefecture. What had been the prefecture of Teochew remains a geographic and cultural unit named *Cháoshàn* 潮汕 in Mandarin, which translates to *Tiô-Sùaⁿ* in Teochew. This is a portmanteau of *Cháozhōu* 潮州 (Teochew) and *Shàntóu* 汕頭~汕頭 (Swatow), the names of the two major cities in the region.

The outlines of the war between the *Teochew dialect* and *Chaoshan Min* pages on English Wikipedia (mentioned in §2.4) should be clear from here for those with some acquaintance with modern Chinese society. First, *Tiô-chiu* as used in its traditional sense is based on the former administrative hierarchy. Continued use of *Teochew* may irk people who identify with the current administrative hierarchy. Second, the English *Teochew* was romanized directly from Teochew *Tiô-chiu* in pre-modern times. Continued use of *Teochew* annoys people who identify with the modern national language, Mandarin. Third, Teochew City is still called *Tiô-chiu*. Like many pairs of neighboring cities or districts of comparable stature in modern China, Teochew City and Swatow are locked in a never-ending competition to win face; using *Tiô-chiu* or *Teochew* in the traditional sense may be seen by some as giving face to Teochew City at the expense of Swatow.

Thus *Tiô-chiu* and its cognates continue to be used in their traditional sense outside China, but are being phased out in China. The more roots-minded Teochew-speaking individuals outside China under-

stand and generally sympathize with at least the first and third issues outlined above. Thus *Tiô-Sùaⁿ* is arguably the preferred autonym for Teochew in the 21st century, and few will find it objectionable.

Since ISO 639-3 [tio] and [teo] are taken, we begin by looking through the [t*s] and [ts*] options, where the third letter would be filler. We find [tfs], [tps], [tqs], and [tws] available. We suggest [tws] because the filler letter corresponds to the ending letter of the English *Teochew*. Should this actually be undesirable, we would suggest [tqs].

It is well advised that, while our suggestions follow from the preceding analysis, we do not have extensive personal experience of Teochew-speaking society.

Appendix A-5—Names of Haklau

In Haklau, the Haklau language or dialects thereof are referred to either as HAKLAU,^{284 285} or using place names—HAI HONG [SPEECH], LOK HONG, SWABUE 汕尾, or HAI LOK HONG. HAKLAU does not have a stable Mandarin equivalent; the rest line up with Mandarin *Hǎifēnghuà*, *Lùfēnghuà*, *Shànwěihuà*, and *Hǎilùfēnghuà*, respectively. HAI LOK HONG is a portmanteau place name that covers the other three, including SWABUE. Thus, HAI LOK HONG refers to the Haklau language as a whole.

ISO 639-3 codes [hlh] (for HAI LOK HONG) and [hkl] both seem ideal, and both are available, assuming [hkl] has not already been reserved for Kelantan Local Hokkien. We tentatively recommend [hlh], as Haklau speakers seem to prefer HAI LOK HONG over HAKLAU for formal usage.

It is well advised that, while our suggestions follow from the preceding analysis, we do not have extensive personal experience of Haklau-speaking society.

Appendix A-6—Names of Hokkien

English *Hokkien* corresponds to Hokkien *Hok-kiàn-ōe*²⁸⁶ (“Fujian speech”), as does Mandarin *Fújiànhuà* 福建话. These names are used throughout the diaspora, including in southern Zhejiang. English *Fukien* comes from Mandarin by way of a pre-Pinyin romanization.²⁸⁷

284 As with Luichew, Haklau does not have a romanized script.

We use small-caps forms to indicate an approximate romanization, or to signify “the equivalent of _____.”

285 HAKLAU is cognate with HOKLO and HOLO from Appendix A-1.

286 Or *Hok-kiàn-ōa*, depending on the dialect.

287 English *Fujianese* may refer to Hokkien, but more often refers to Fuzhou. This reflects the history of engagement between the English-speaking world and people from Fujian.

In southern Fujian, Hokkien is now called *Bân-lâm-gí*²⁸⁸ (“the language of southern Fujian”) or *Bân-lâm-ōe* (“speech of southern Fujian”). This corresponds to Mandarin *Mǐnnányǔ* 闽南语 and *Mǐnnánhuà* 闽南话, which is what Hokkien is called in Mandarin throughout China, except in southern Zhejiang. English *Minnan* and *Minnanese*, as well as the academic *Southern Min*, are derived from the Mandarin; *Banlamese*, less commonly seen, is derived from the Hokkien.

In rural districts of western Chiangchew, in proximity to Hakka settlements, Hokkien is also called *Hòh-ló-ōa* (“Hoklo speech”).

Hokkien *Lán-lâng-ōe* 咱人話 (“we [inclusive] folks’ speech”), translated into English as *Lannang*,²⁸⁹ is used mostly in the Philippines, where *Lán-lâng-ōe* seems to be used more than *Hok-kiàn-ōe*, and *Lannang* more than *Hokkien*. This autonym has not been translated into Mandarin.

English *Amoy*, once used in the sense of *Hokkien*, was derived from a typical Chiangchew pronunciation of AMOY.²⁹⁰ Colloquial Hokkien does not seem to have ever referred to *Hokkien* as AMOY, but the Classical Chinese collocation 厦語~厦语 (“the language of Amoy”), pronounced *Hāgú* in Amoy Hokkien and *Xiàyǔ* in Mandarin, corresponds to English *Amoy* in meaning, if not register. Capital 95.8 FM in Singapore uses this name for Hokkien (see §2.6).

The Hokkien language did not really have a name before it left home.²⁹¹ Even then, its new name—HOKKIEN—was only used away from home: in Indochina, in Nusantara, in Zhejiang. Not just the name, but the mere awareness that a language should have a name, were products of something like an out-of-body experience, collectively experienced for centuries through Hokkien-speaking expats at faraway ports. The modern Chinese states as well as Hokkien speakers in southern Fujian, strangers to that experience, see *Hok-kiàn* as defective: it is improper for a language not spoken in the capital of a province to take on the name of the province.²⁹² Hence the administratively correct *Bân-lâm*.

288 Or *Bân-lâm-gú* or *Bân-lâm-gú*, depending on the dialect.

289 This is a phonetic rendering of the pronunciation in Philippine Hokkien, where the second L is assimilated by the preceding nasal.

290 Romanized Hokkien does not generally allow or provide for Chiangchew Hokkien to be fully written. Borrowing minimally from sister scripts, we could represent the source form of *Amoy* as **Ē-mûi*.

291 <http://hokkienese.com/?p=923> (last accessed Aug 2021; in Hokkien)

292 Later, when other ethnolinguistic groups arrived from Fujian, they were named according to their prefecture or county: *Foochow* 福州, *Hokchia* 福清 (Mand. *Fúqīng*), *Henghua* 興化~興化 (Mand. *Xīnghuà*, modern 莆仙 *Pǔxiān*).

As with Teochew and *Tiô-Sūa*, diaspora Hokkien speakers—if they are aware of the problem—tend to be sympathetic to both the homeland and diaspora points of view. Thus *Bân-lâm-ōe*, or *Bân-lâm-ōa*, is probably the preferred autonym for Hokkien in the 21st century; few will find it objectionable.

ISO 639-3 codes [blm], [bal] and [bnl] are in use. Looking through the [bl*] and [*bl] codes, we find [ebl], [hbl], [jbl], [vbl] and [xbl] available. Code [hbl] seems suitable, with H representing the province of *Hok-kiàn*, or even *Hòh-ló*; and B-L as an abbreviation of *Bân-lâm*.

It is well advised that, while our suggestions follow from the preceding analysis, we do not have extensive personal experience of Hokkien-speaking society.

Appendix A-7—Names of Chawan

In Chawan and mainstream Hokkien, the Chawan variety is called *Chiàu-an-ōe* (“Chawan speech”).²⁹³ The corresponding Mandarin form is *Zhào’ānhuà* 詔安話~詔安話. English *Chawan* and *Chao Ann* are derived from Chawan-Hokkien *Chiàu-an*.

We suggest grouping Chawan with Hokkien for ISO 639-3 purposes.

It is well advised that, while our suggestions follow from the preceding analysis, we do not have extensive personal experience of Chawan-speaking society.

Appendix A-8—Names of the Datian Frontlect

In Mandarin, the Datian Frontlect is called *Qiánlùhuà* 前路話 (“front way speech”) or *Dàtián Qiánlùhuà* 大田前路話 (“Datian front way speech”). We have every reason to believe *Qiánlùhuà* is a direct translation from the Frontlect; the phrasing is not really native to Mandarin. Mandarin *Dàtiánhuà* 大田話 typically refers to the Frontlect as well, if used by someone who knows the linguistic situation in Datian. Since ISO 639-3 and *Ethnologue* are important sources for people trying to understand Datian linguistically, we recommend listing 前路話²⁹⁴ as an autonym and *Datian Frontlect* or the like as an English translation of the autonym.

We suggest grouping the Datian Frontlect with Hokkien for ISO 639-3 purposes, the way Aranese [oci] is grouped with Occitan [oci].

It is well advised that, while our suggestions follow from the preceding analysis, we do not have

293 The Chawan pronunciation of -ōē is [ue] (tone aside), according to 洪惟仁 (Ang Uijin, 2014) “十六、七世紀之間呂宋的漳州方言。”

294 There does not seem to be any way to write the Frontlect except using Standard Chinese Sinographs.

extensive personal experience of Frontlect-speaking society.

Appendix A-9—Names of Lengna

English *Lengna* was derived from Hokkien *Lêng-nâ*. In Lengna, at least as it is spoken in town, the pronunciation is roughly LIONGNA. The corresponding Mandarin name for the language is *Lóngyánhuà* 龙岩话~龍巖話.²⁹⁵

The ISO 639-3 codes [lna] is already in use, as are [lgn] and [lnn], and [len]. Code [lnx] is available and appears to be the most suitable option.

It is well advised that, while our suggestions follow from the preceding analysis, we do not have extensive personal experience of Lengna-speaking society.

Appendix A-10—Names of Lungtu & Namlong

Lungtu and *Namlong* are apparently approximations of the respective autonyms. The Lungtu language is called *Lóngdūhuà* 隆都话 in Mandarin; the Mandarin for Namlong varies depending on which of several Sinographs is used to write the back syllable.

Bodman consistently wrote “Namlong,” but sometimes wrote “Longdu” instead of “Lungtu.” It is likely that the native pronunciation of LUNGTU is so close to the Mandarin pronunciation that Bodman figured he might as well romanize it by way of Mandarin and Pinyin, although Egerod had used “Lungtu.”

We have recommended the ISO 639-3 code [zsh] for Samheung (see [Appendix A-11](#)). By the same logic, we recommend [zlt] or [zld] for Lungtu, and [znl] for Namlong. Between [zlt] and [zld], we would choose [zld] as a nod to the prevalence of *Longdu* in recent scholarship.

It is well advised that, while our suggestions follow from the preceding analysis, we do not have extensive personal experience of Lungtu- and Namlong-speaking society.

Appendix A-11—Names of Samheung

Samheung is romanized from the Cantonese. According to Bodman, the region is “called Sa-hiu by its natives.”²⁹⁶ The corresponding Mandarin name for the language is *Sānxiānghuà* 三乡话~三鄉話.

²⁹⁵ As with Hainanese, the Mandarin term *Lóngyánhuà* 龙岩话 is contested on the Chinese internet by locals—ethnic Hakka, in this case—who believe the term should refer in the aggregate to all the languages of *Lóngyán* 龙岩 as defined under the current administrative hierarchy. We are not sure if there is any other name for the Lengna tongue, but perhaps justice would be served and controversy quelled if the native construction LIONGNA is used instead of the Mandarin *Lóngyán*, much as the Yami autonym TAO (“people”) is not thought to be prejudicial to other humans.

²⁹⁶ Bodman (1988)

ISO 639-3 codes [sah] and [shx] are in use. Code [zsh] is available, and may be ideal given the circumstances that the language finds itself in; z could also be read as representing Zhongshan 中山, the modern name of the county where Samheung is spoken.

It is well advised that, while our suggestions follow from the preceding analysis, we do not have extensive personal experience of Samheung-speaking society.

Appendix A-12—Names of Kelantan Local Hokkien

Kelantan Local Hokkien is a phrase that we adapted from *Kelantan Peranakan Hokkien* solely to use in this paper. *Kelantan Peranakan Hokkien*, in turn, was coined by scholars. Local Hokkien-speaking (*Cina Kampung*) scholar Teo Kok Seong uses the term himself, but states that it is unique to academic usage; nor do the *Cina Kampung* (“country Chinese”) call themselves *Peranakan*.

The *Cina Kampung* have many words for themselves, including TNG LANG²⁹⁷ as well as COUNTRY CHINESE, THAI CHINESE, and SIAMESE HOKKIEN.²⁹⁸ We suspect most of these autonyms can be applied to the language. But these terms more or less assume a Kelantan context. Mek Yun’s video, on the other hand, refers to the language in (apparently) Kelantan Malay as *Hokkien Kelate*. This and Standard Malay *Hokkien Kelantan* seem to be the prevailing ways to refer to the language online, although *Hokkien Kelantan* could also refer to the straight Hokkien spoken in Kota Bharu. As noted in [§2.12](#), Kelantan Malay *Kelate* counterintuitively seems to translate to KELANTAN in Local Hokkien; *Hokkien Kelate* would translate to HOKKIEN KELANTAN.

The most intuitive ISO 639-3 code for Local Hokkien would be [hkl]—H for *Hokkien*, and K-L for *Kelantan*. This code is available. Since Kelantanese forms the greater part of the *Cina Kampung* identity, it is fitting that two of the three letters correspond to Kelantan. If it should happen [hkl] has already been reserved for Haklau, we alternatively suggest [hsk]—HOKKIEN SIAM KELANTAN—for Local Hokkien.

It is well advised that, while our suggestions follow from the preceding analysis, we do not have extensive personal experience of Kelantan Local Hokkien-speaking society.

²⁹⁷ The Hokkien cognate of this word means *Chinese people*. The Local Hokkien word only refers to the *Cina Kampung*. We use Teo’s spelling approximately.

²⁹⁸ Teo (1993) does not give the native forms for these.