Language classification in the Pearl River Region

Grambank – The Next Generation; MPI EVA Leipzig

Hilário de Sousa

Centre de recherches linguistiques sur l'Asie orientale
membre associé

Dienstag 24 November 2026
🔥🐴 丙午十月十六 🌕

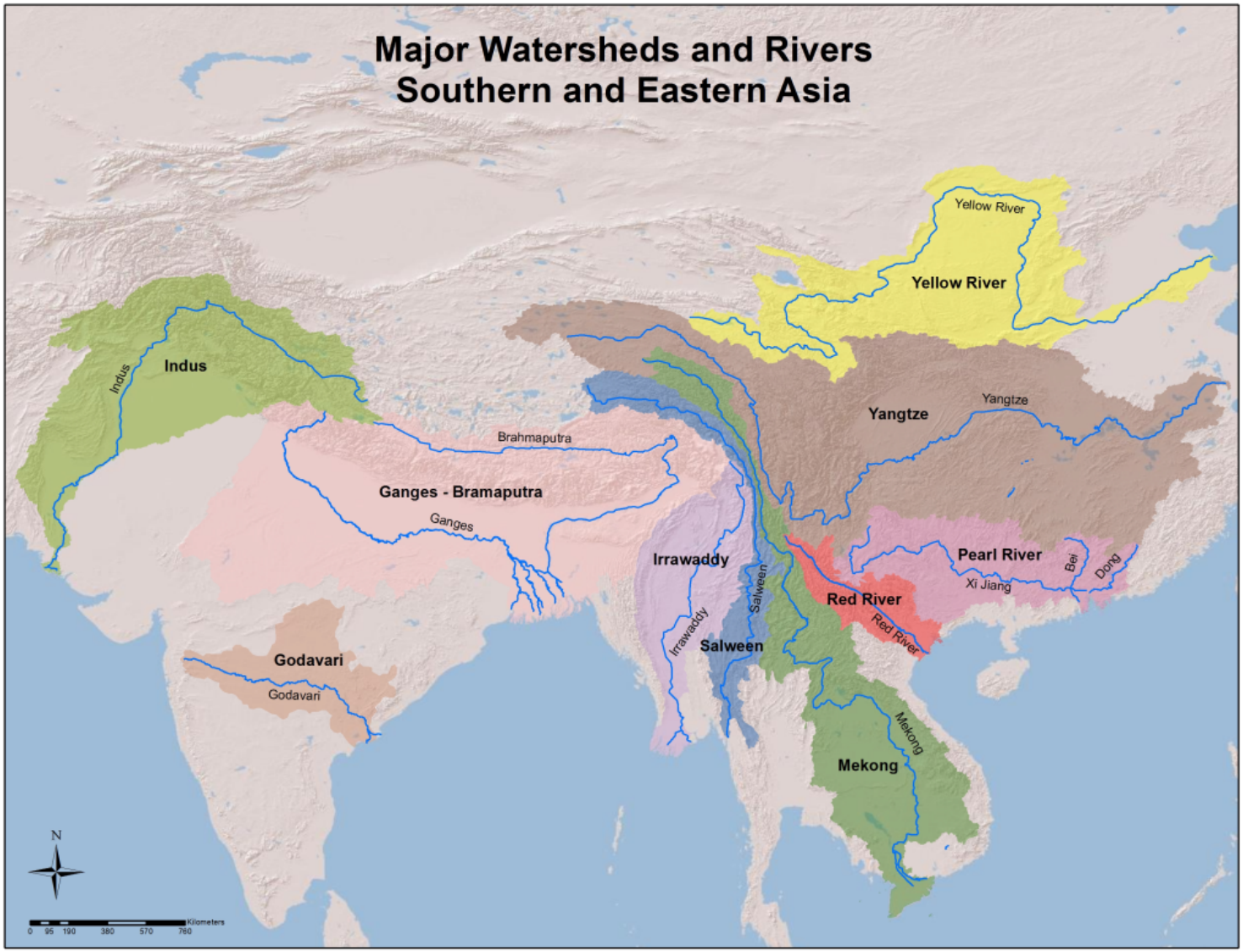
hilario@bambooradical.com

https://hilario.bambooradical.com

Myself

- **Hilário de Sousa** Portuguese /iˈlaɾiu dɨ ˈsouzɐ/ (Macau ㄧ+ㄱ*, SEP H+L*), NZ English /hɪˈlaːɹɪɐʉ/~/hɪˈlɛːɹɪɐʉ/, Dutch /ɦiˈlaːɾijoː/, French /ilaʁjo/, etc.
- any animate pronoun, Dr.
- `https://hilario.bambooradical.com`
- *membre associé* of CRLAO 🇫🇷; 🏠: 🇳🇱
- Types of past / forthcoming publications involvements:
  - Grambank (*hurra!* 🥳)
  - Switch-reference & co., Melanesia
    Mid-North New Guinea, Southern Vanuatu
  - Psycholinguistics, Semantic typology
  - Writing systems
  - Chinese diaspora in Oceania
  - Southern China as part of Mainland Southeast Asia
    (areal linguistics / typology)
  - Phonology, morphosyntax, history in Far Southern China:
    Yuè, Pínghuà, Zhōngshān Mǐn, Sino-Vietnamese

Pearl River Basin/Region



https://openrivers.lib.umn.edu/article/hydrology-and-world-history/watershed-map3/

Pearl River Basin/Region



Figure 1: Pearl River Basin

Based on `https://en.wikipedia.org/wiki/Pearl_River#/media/File:Zhujiangrivermap.png`

Pearl River Basin/Region



Figure 2: Some watersheds in East Asia

Based on
https://decolonialatlas.wordpress.com/wp-content/uploads/2019/11/world-watersheds-blank.png

Language families

Ethnolinguistic situation before the Hàn expansion was something like this: *(i.e. there are various theories)*



Fig. 1. Ethnic divisions in South China before the Han's southward expansion. Dashed lines are the ethnic divisions. Dotted lines are the province borders. Dashed dotted lines are the country borders. Solid lines are rivers and sea lines. Important cordilleras between ethnic areas are marked in the map.

Li, Hui et al. 2007. Mitochondrial DNA diversity and population differentiation in southern East Asia. *American Journal of Physical Anthropology* 134(4). 481–488. https://doi.org/10.1002/ajpa.20690
https://loca.fudan.edu.cn/Doc/A35.pdf

Language families

Kra-Dai (Tai-Kadai) 壯侗～侗台

- Tai 台～壯傣～傝 (NT: unaspirated)
- Be (Ong-Be) 臨高
- Kam-Sui 侗水 (侗 dòng dung6)
- Hlai 黎
- Kra 仡央 (仡佬–布央)
- Biao–Lakkja 標–拉珈

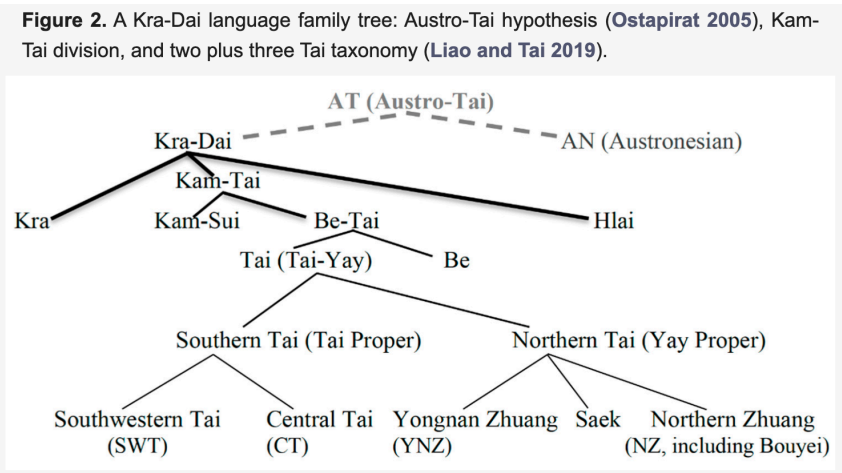
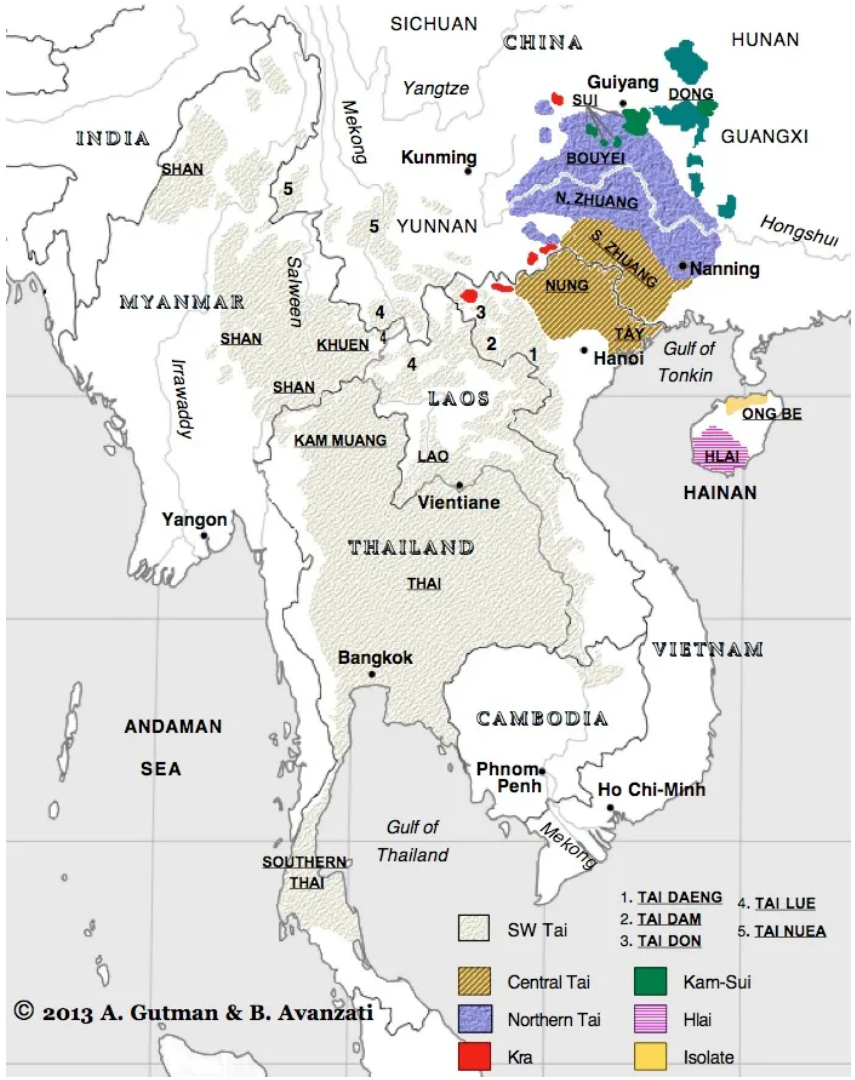


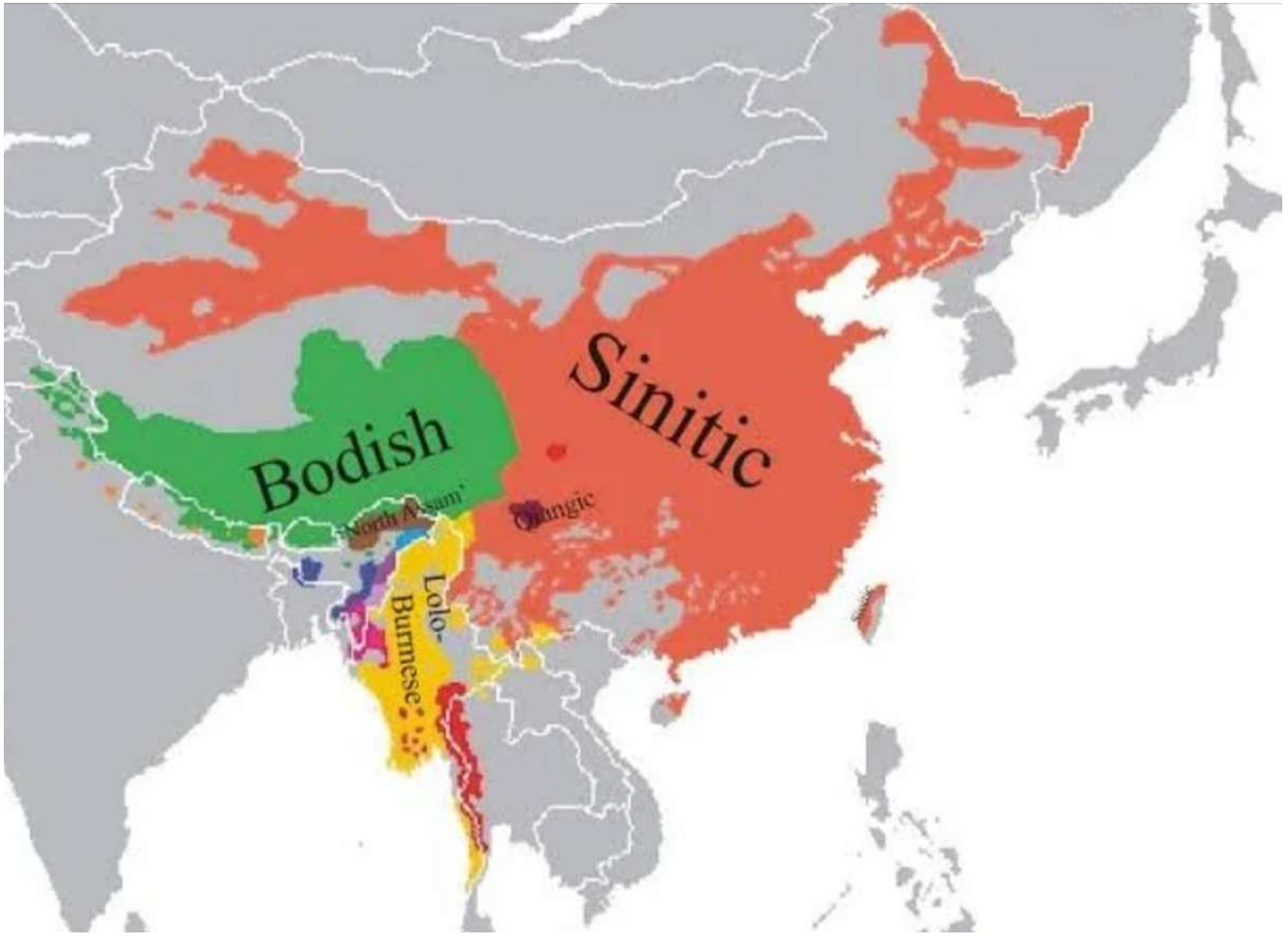
Figure 2. A Kra-Dai language family tree: Austro-Tai hypothesis (**Ostapirat 2005**), Kam-Tai division, and two plus three Tai taxonomy (**Liao and Tai 2019**).

Liao, Hanbo. 2023. Tonal Behavior as of Areal and Typological Concerns: Centering on the Sinitic and Kam-Tai Languages in Lingnan. *Languages* 8(2). 148. https://doi.org/10.3390/languages8020148



Language families

Sino-Tibetan 漢藏 / Trans-Himalayan 跨喜瑪拉雅



"Tibeto-Burman" 藏緬 = non-Sinitic Sino-Tibetan languages (one node? multiple nodes?)

Is Sinitic a first order branch?

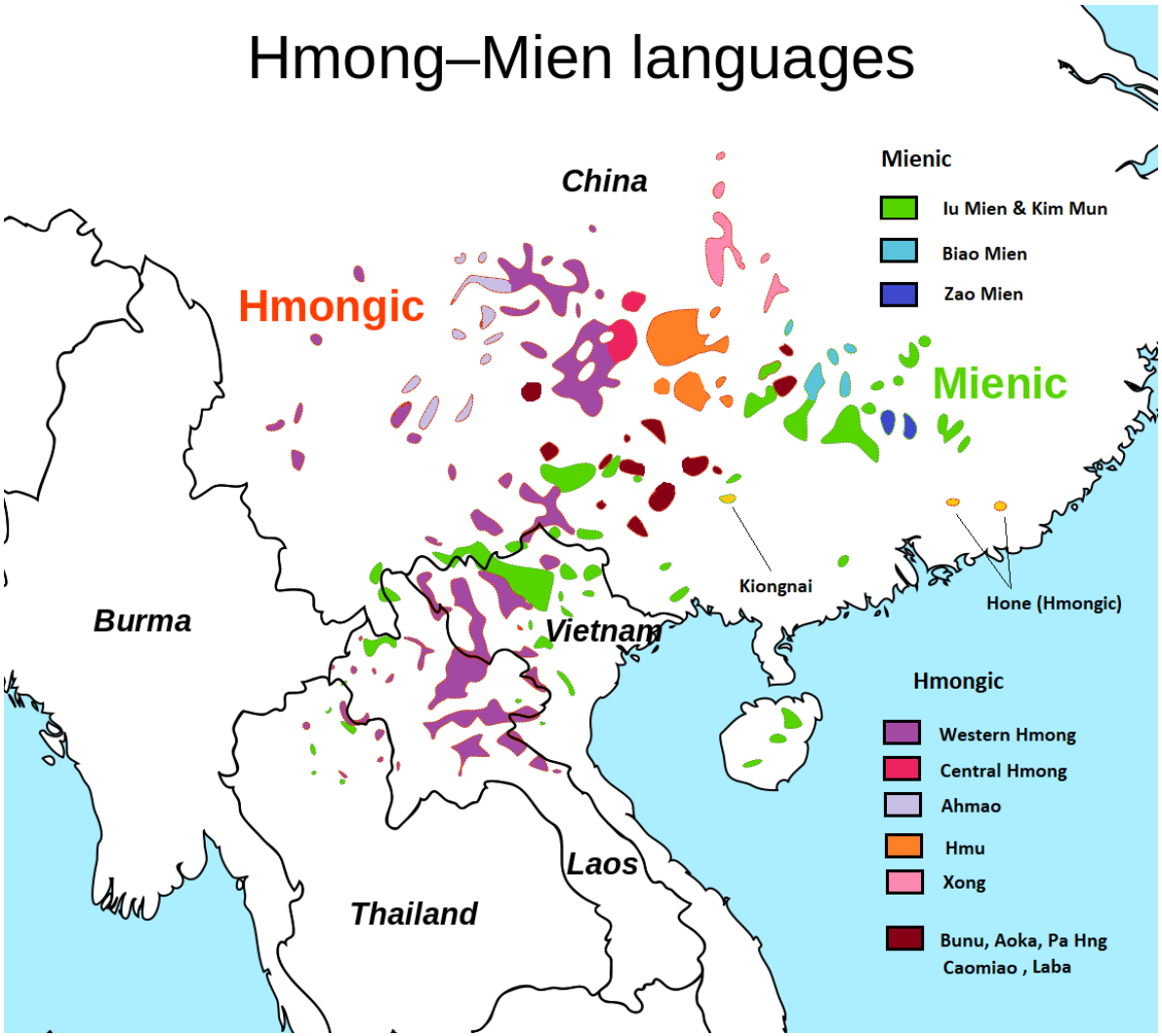
Pearl River Basin: Sinitic and Loloish ("Luóluó 倮倮", i.e. Yí 彝 and relatives)

Language families

Hmong-Mien 苗瑤

Slight influences in the Pearl River region

e.g. Yue-Hashimoto (1991) claims that Cantonese mass classifier 啲 di1 is from Hmgon-Mien

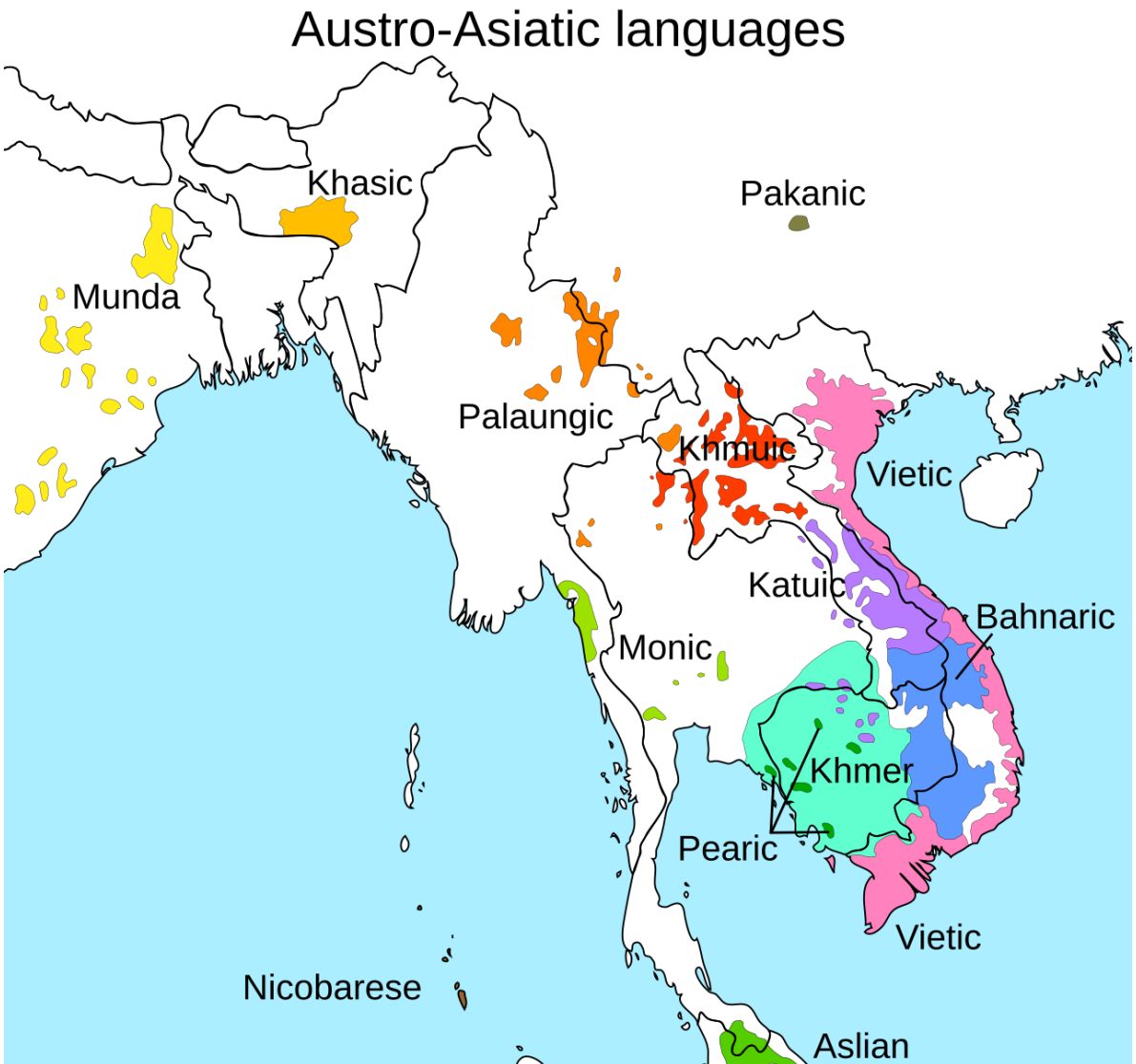


Yue-Hashimoto, Anne Oi-kan. 1991. The Yue dialect. In Wang, William S. Y. (ed.), *Languages and Dialects of China* (Journal of Chinese Linguistics Monograph Series Number 3), 292–322.

Language families

Austroasiatic 南亞

Pearl River Basin:
Bolyu 侾 (Pakanic)
Vietnamese
a little bit



Language classification



Figure 3: Some approaches to language/dialect classification

Sinitic languages

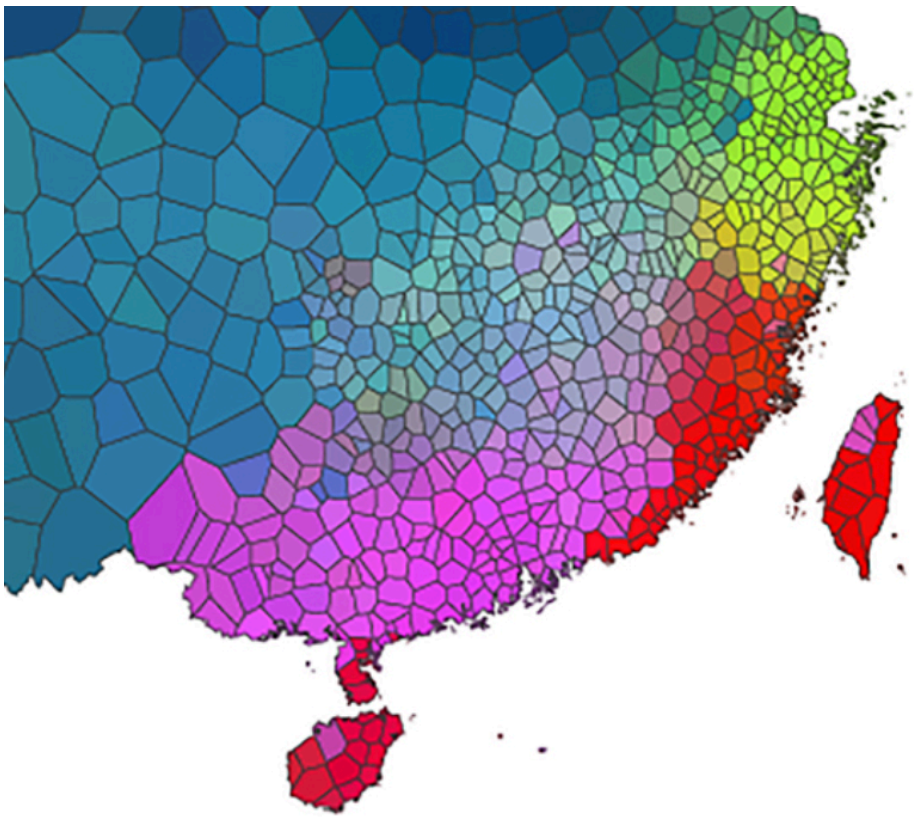


Figure 4: RGB–Multidimensional Scaling map of Southeastern Sinitic

Huang, He & Grieve, Jack & Jiao, Lei & Cai, Zhuo. 2024. Geographic structure of Chinese dialects: a computational dialectometric approach. *Linguistics* 62(4). 937–976. https://doi.org/10.1515/ling-2021-0138
Based on data from the *Linguistic Atlas of Chinese Dialects* 汉语方言地图集

Sinitic languages

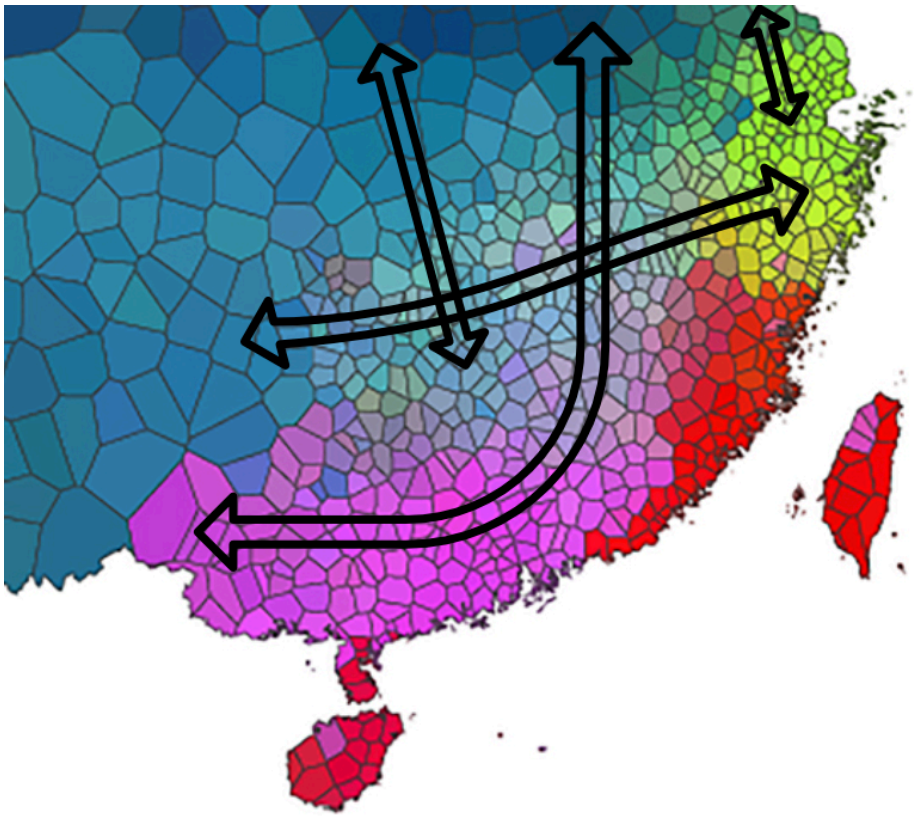


Figure 5: Some Sinitic dialect continua

Base map from Huang et al. (2024), based on 510 features in the *LACD*

Sinitic languages



Figure 6: Dial. groups in Language Atlas of China 1st ed. *phonology based*

Huang, He & Grieve, Jack & Jiao, Lei & Cai, Zhuo. 2024. Geographic structure of Chinese dialects: a computational dialectometric approach. *Linguistics* 62(4). 937–976. https://doi.org/10.1515/ling-2021-0138
Wurm, S.A. & Li, Rong (eds.). 1987. *Language Atlas of China.* 1st edn. Hong Kong: Longman Group (Far East) Ltd.

Sinitic languages

- 官話 Guānhuà / Gun1 Waa2
- 晉 Jìn / Zeon3
- 湘 Xiāng / Soeng1
- 贛 Gàn / Gam3
- 徽 Huī / Fai1
- 吳 Wú / Ng4
- 閩 Mǐn / Man5
- 客家 Kèjiā / Haak3 Gaa1
- 粵 Yuè / Jyut6
- 平話 Pínghuà / Ping4 Waa2
- (+ smaller unclassified groups)



Figure 7: Southern Sinitic dialect groups

https://en.wikipedia.org/wiki/Sinitic_languages#/media/File:Map_of_sinitic_languages_full-en.svg
Based on 1st ed. of the *Language Atlas of China* 中国语言地图集

Pinghua–Yue

Many Yuè dialects are not intelligible to the people of Canton
without prior exposure.

e.g. Cantonese speakers listening to Táishān 台山



Can Cantonese Speakers Understand Taishanese? (ft....

39K weergaven • 4 jaar geleden

https://www.youtube.com/watch?v=nuXg6KhpwXc



广州人爆笑挑战台山话！未入门就放弃！

2018–05–04

https://www.bilibili.com/video/BV1dW41137NS/

Pinghua–Yue LAC 2012



Figure 8: Subgroups within Pinghua & Yue in LAC-1 *phonology based*

https://en.wikipedia.org/wiki/Yue_Chinese#/media/File:Ping_and_Yue_dialect_map.svg

Wurm, S.A. & Li, Rong (eds.). 1987. *Language Atlas of China.* 1st edn. Hong Kong: Longman Group (Far East) Ltd.

Pinghua–Yue LAC 2012

	‘heart’ 心	‘door’ 門	‘sugar’ 糖	‘ten’ 十	‘seven’ 七	‘six’ 六
Mid. Chin.	si**m**^{A}	muə**n**^{A}	da**ŋ**^{A}	dz̥i**p**	ts^{h}i**t**	lju^{w}**k**
Guìlín Mand.	ɕin˧	mən˩	t^{h}aŋ˩	sz̩˩	tɕ^{h}i˩	nu˩
Línguì N Píng.	san˧	man˩	t^{h}ɔ̃˩	ɕe˩	ts^{h}e˥	li˩
Nánníng S Píng.	ɬɐm˧	mun˩	taŋ˩	ɬɐp˨	tʃ^{h}ɐt˨	lʊk˨
Nánníng Canto.	ɬɐm˥	mun˩	t^{h}ɔŋ˩	ʃɐp˨	tʃ^{h}ɐt˥	lʊk˨
Sino-Viet.	*tâm*	*môn*	*đường*	*thập*	*thất*	*lục*

Northern Pínghuà: Under the influence of SW Mandarin (+ Xiāng ∼ S Húnán), Middle Chinese codas of -m -p -t -k have not been preserved; fewer tones on average

Southern Pínghuà: usually conservative of of the Middle Chinese codas of -m -n -ŋ -p -t -k, similar to Yuè and Sino-Vietnamese; more tones on average

Línguì N. Píng. = Liǎngjiāng 兩江; Nánníng S. Píng. = Wèizǐlù 位子淥

Yue–Pinghua Sung 2026

Dialectometry, e.g. Sung 2026 (: 38)

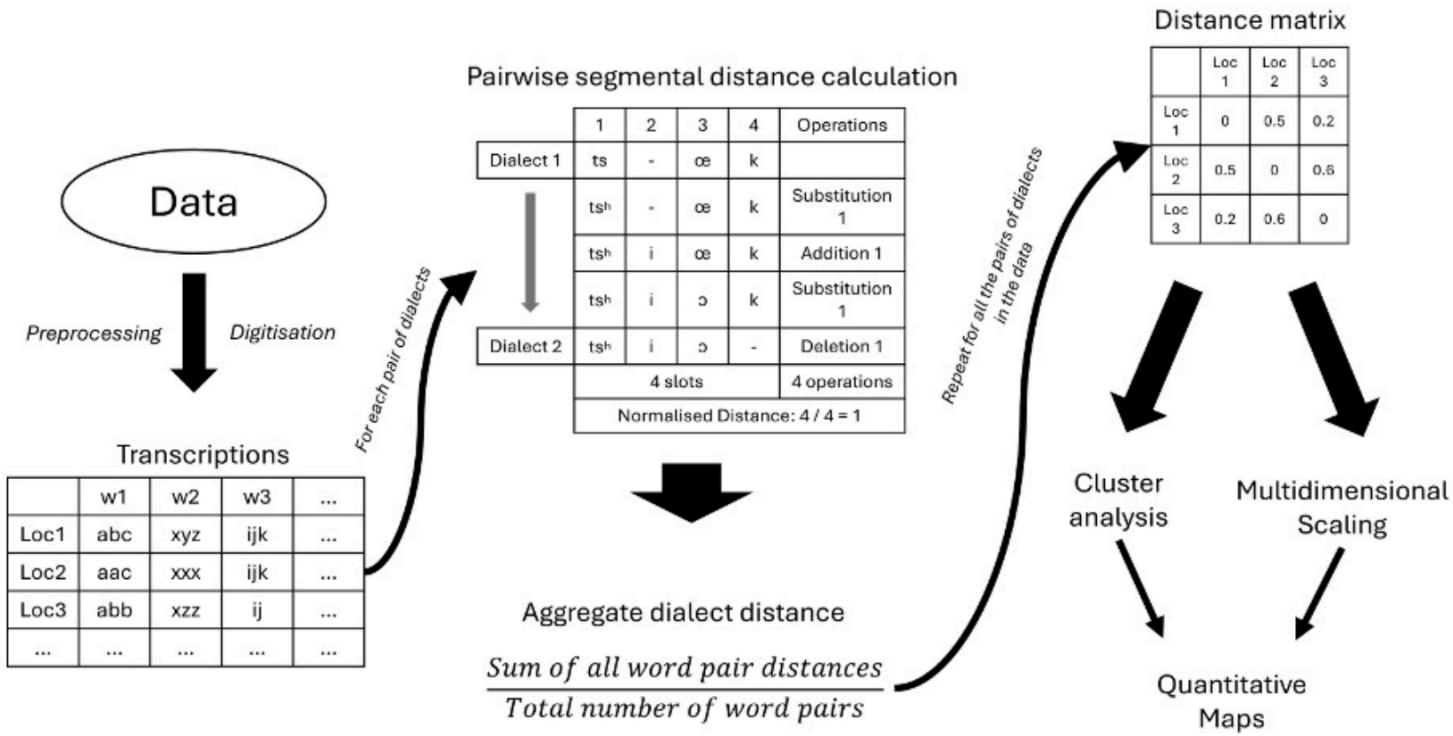


Figure 3.2: Workflow of Dialectometry using Levenshtein Distance

Sung, Ho Wang Matthew. 2026. *Advancing Explanatory and Tonal Dialectometry* (LOT 709). Amsterdam: Landelijke Onderzoekschool Taalwetenschap. https://doi.org/10.48273/LOT0709

Yue–Pinghua Sung 2026

Pairwise segmental distance calculation, e.g. 雀 'bird':

Dialects	Onset	Medial	Nucleus	Coda	Operations	Costs
Guangzhou	ts	-	œ	k		
	tsʰ	-	œ	k	Substitution from ts to tsʰ	1
	tsʰ	**i**	œ	k	Insertion of i	1
	tsʰ	i	**ɔ**	k	Substitution from œ to ɔ	1
	tsʰ	i	ɔ	-	Deletion of k	1
Bao'an	tsʰ	i	ɔ	-		4

Figure 3.3: Illustration of Levenshtein Distance

Sung (2026: 39)

Yue–Pinghua Sung 2026

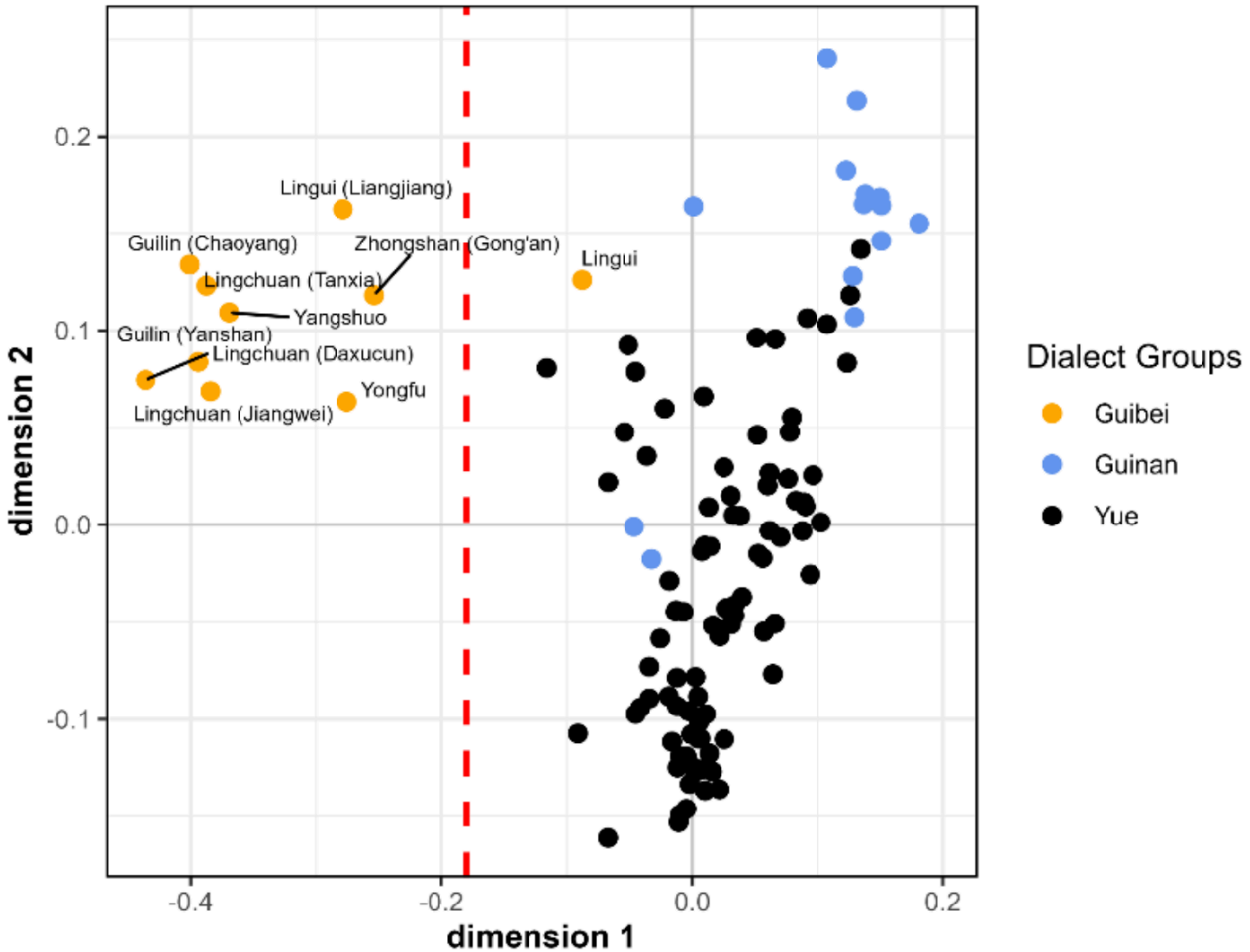


Figure 5.1: MDS plots of the segmental distances of 113 Yue and Pinghua dialects (Levenshtein distance), $r^2 = 0.59$

"Lingui" = Línguì Wǔtōng 臨桂五通

Sung (2026:66)

Yue–Pinghua Sung 2026

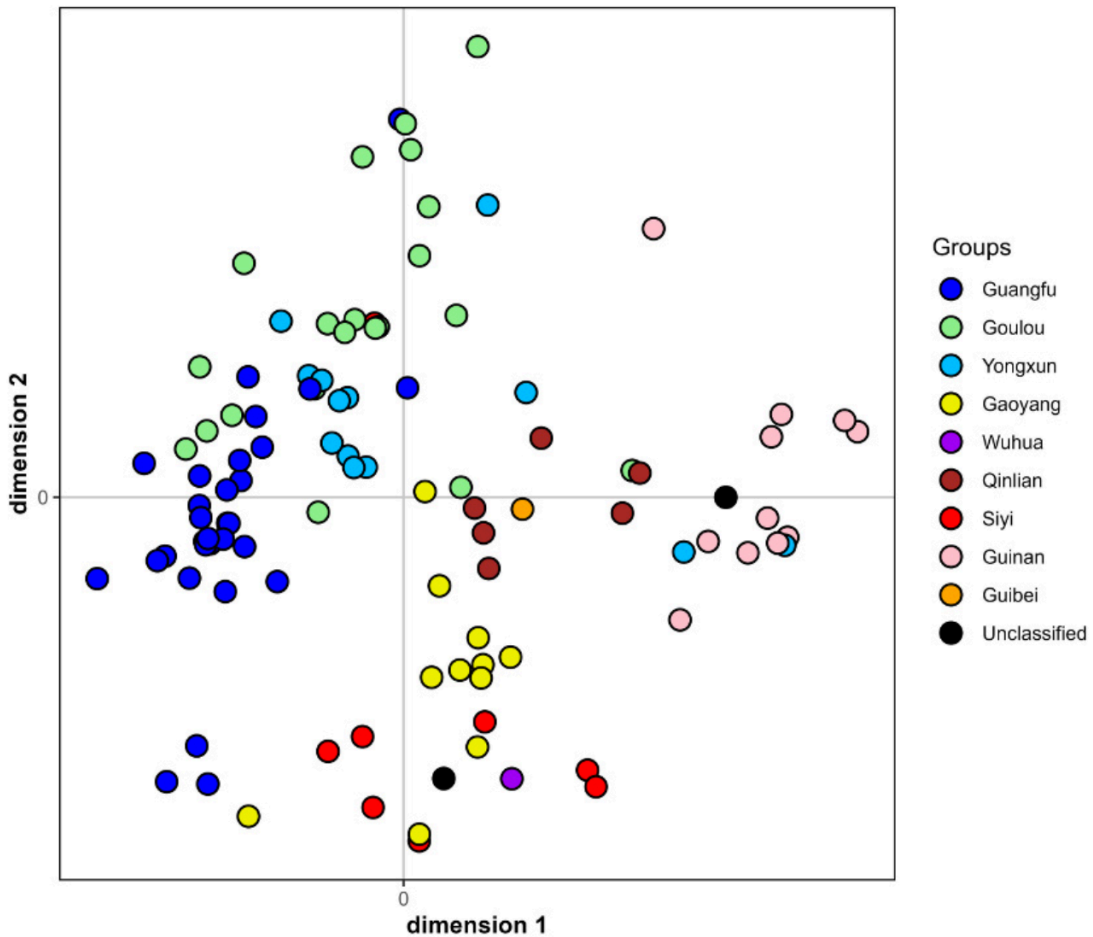


Figure 5.5: MDS plot of 104 Yue dialects (Figure 5.3) with LAC dialect group annotation (Dimension 1 & 2)

Sung (2026:70)

Yue–Pinghua Sung 2026

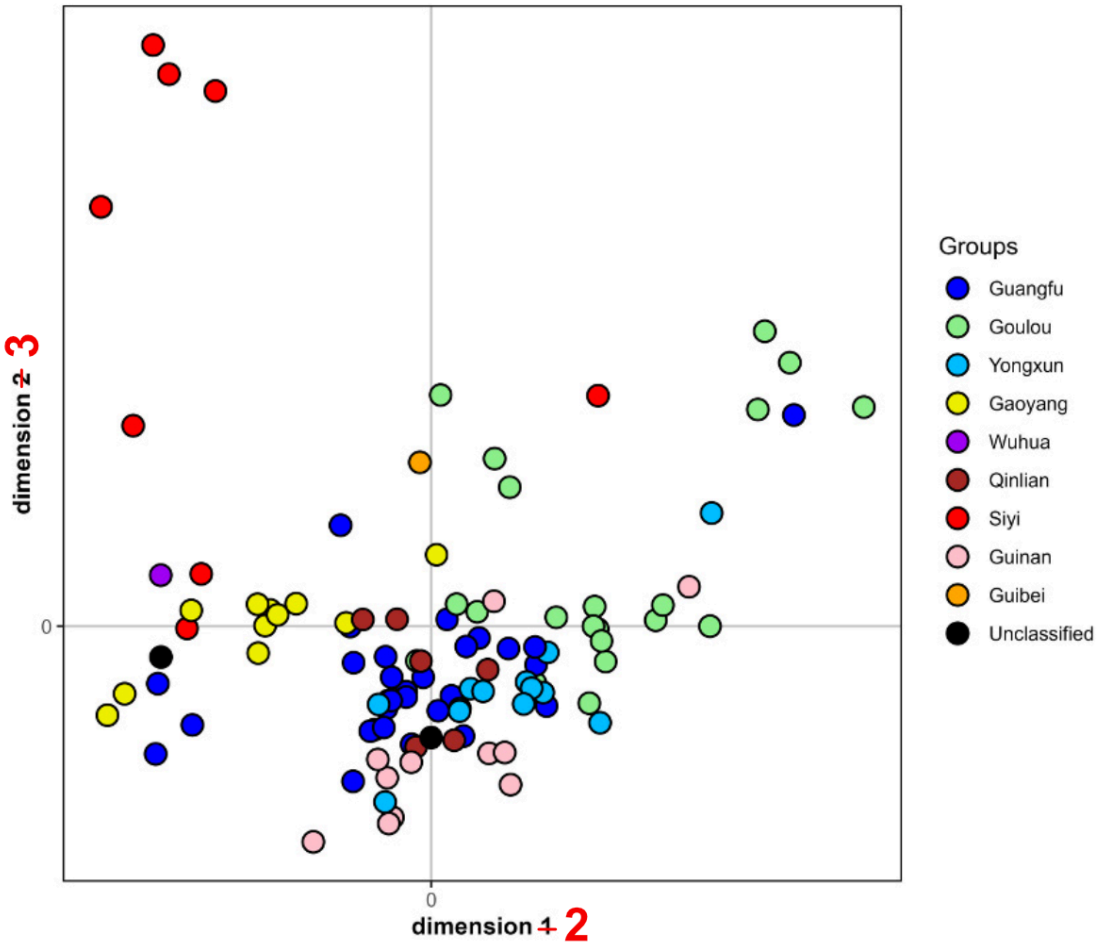


Figure 5.6: MDS plot of 104 Yue dialects (Figure 5.3) with LAC dialect group annotation (Dimension 2 & 3)

Sung (2026:71)

Yue–Pinghua Sung 2026

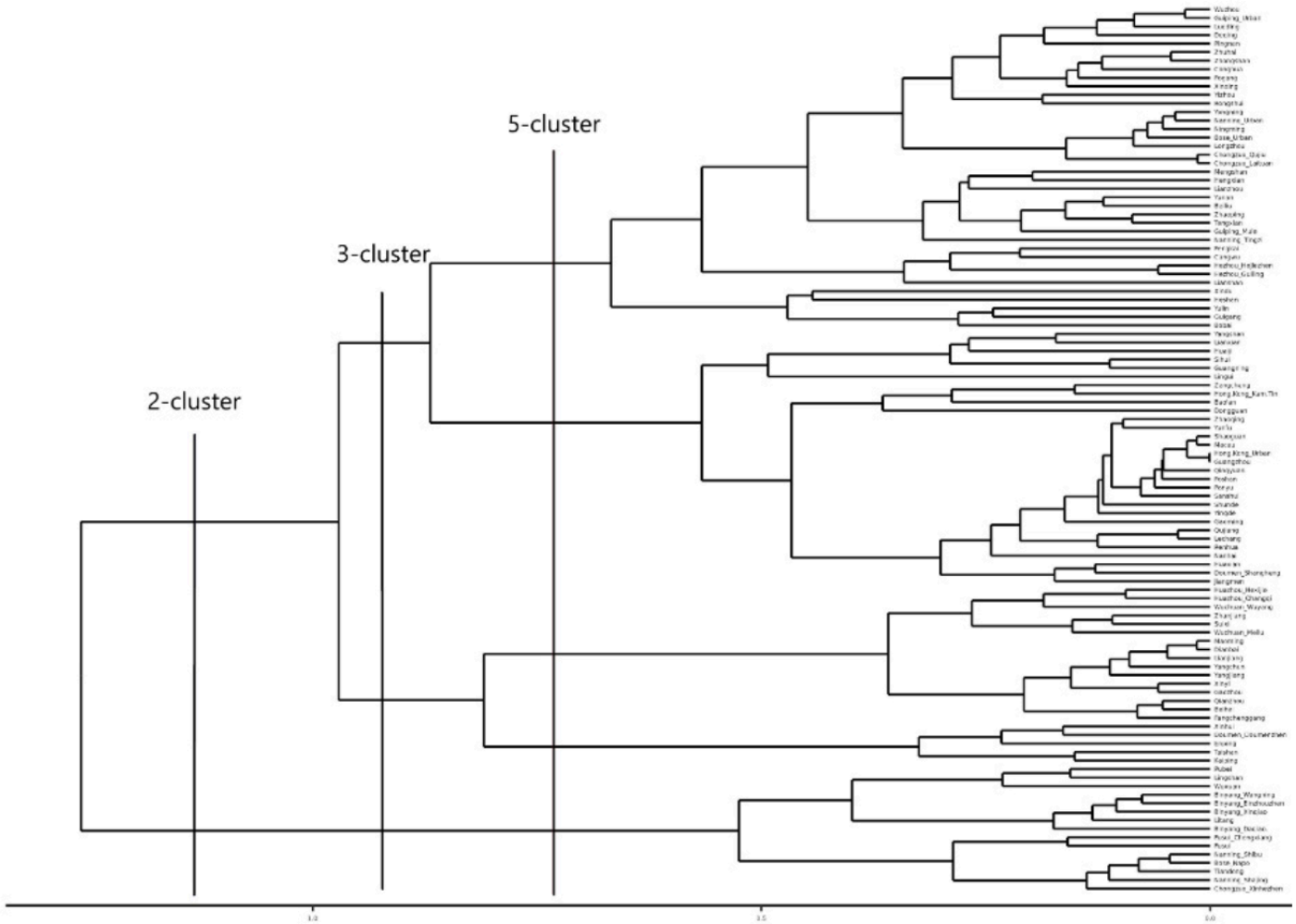


Figure 5.12: Dendrogram of 104 Yue segmental distances using Ward's method

Sung (2026:79)

Yue–Pinghua Sung 2026

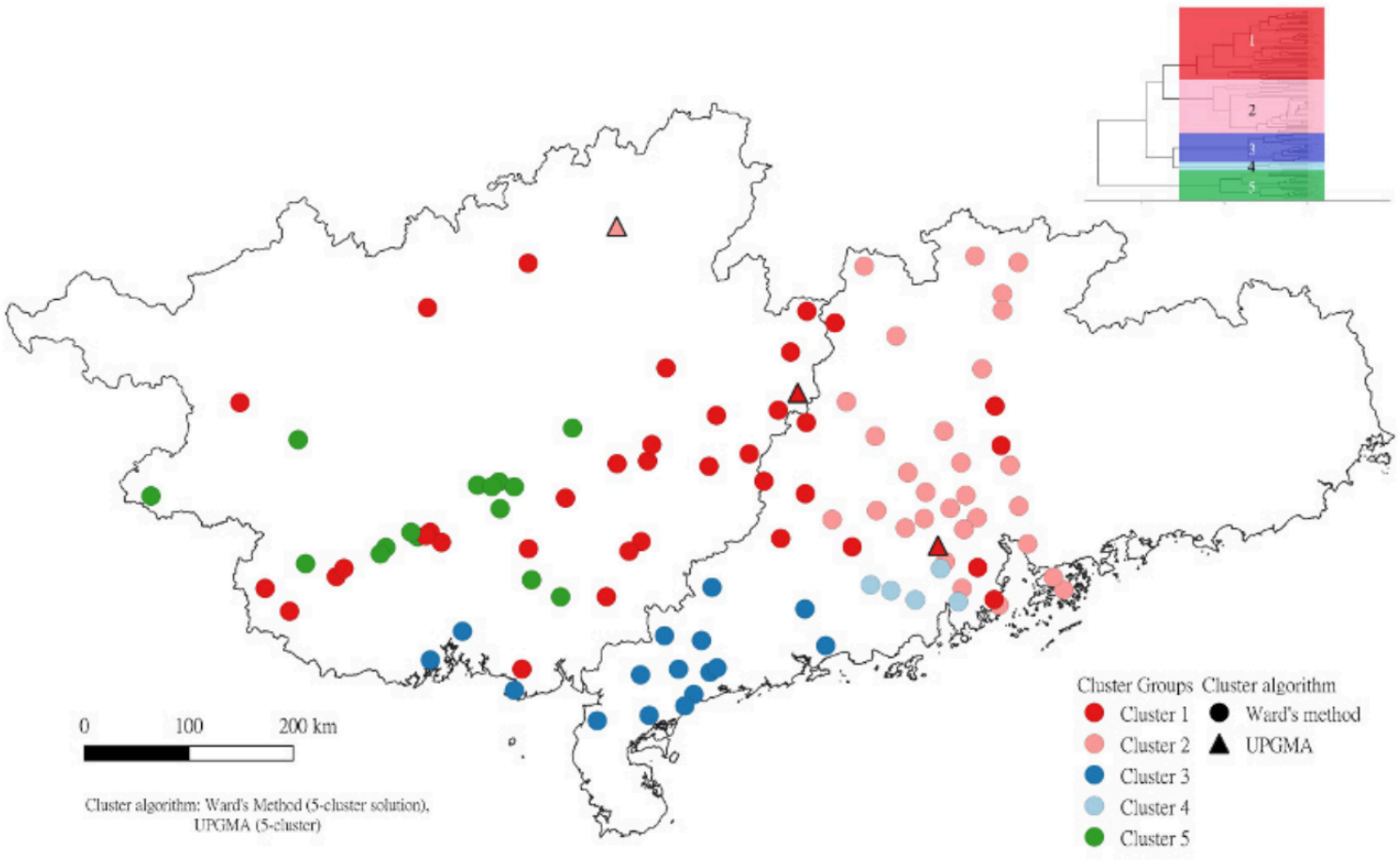


Figure 5.15: Cluster map based on the Ward's method with dialect islands

Sung (2026:83)

Yue–Pinghua Sung 2026



Figure 8.3: MDS plot of aggregate tone distances between Yue dialects, $r^2 = 0.49$

Tones:

Cantonese or similar

Sìyì 四邑

Sung (2026:163)

Yue–Pinghua Sung 2026

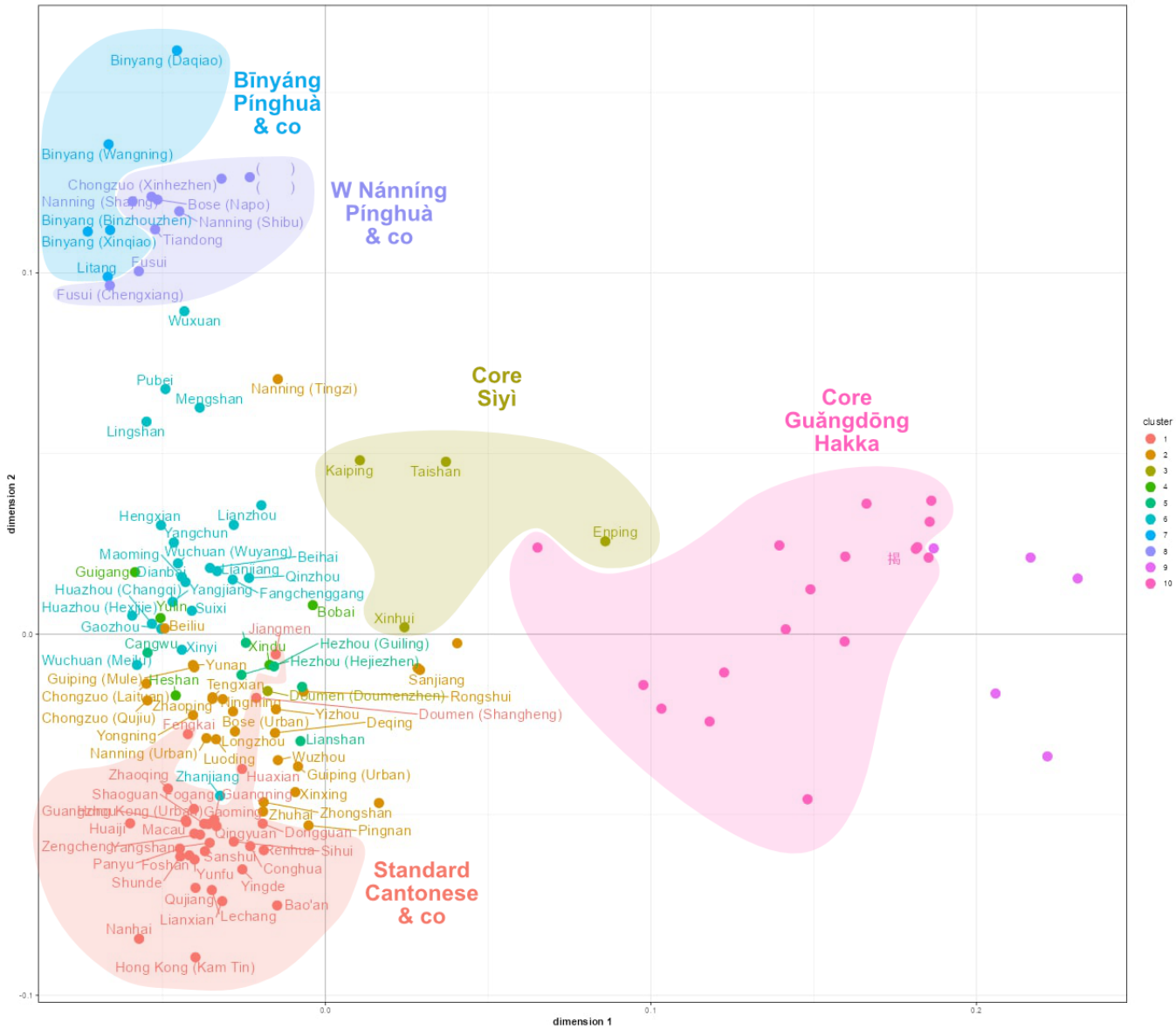
Amongst the many merits of Sung (2026):

- criticisms of the subgrouping of Yue and Pinghua dialects in the LAC
- dealing with a very large number of features (no cherry-picking of which features should have more weighting)
- [and many other merits…]

One criticism: Sung (2026) has used the LAC's definition of Yue and Pinghua. My questions are, e.g.

- What is Yue? Why should the Sìyì dialects be considered Yue (linguistically speaking)?
- Can some of the Hakka / Hakka-like dialects be considered Yue? Is the Yue–Hakka boundary as clear-cut as assumed?

Yue–Pinghua Sung & de Sousa forthcoming



(work in progress)

MDS plot (10-cluster solution, Ward's Method)

Yue–Pinghua Sung & de Sousa forthcoming

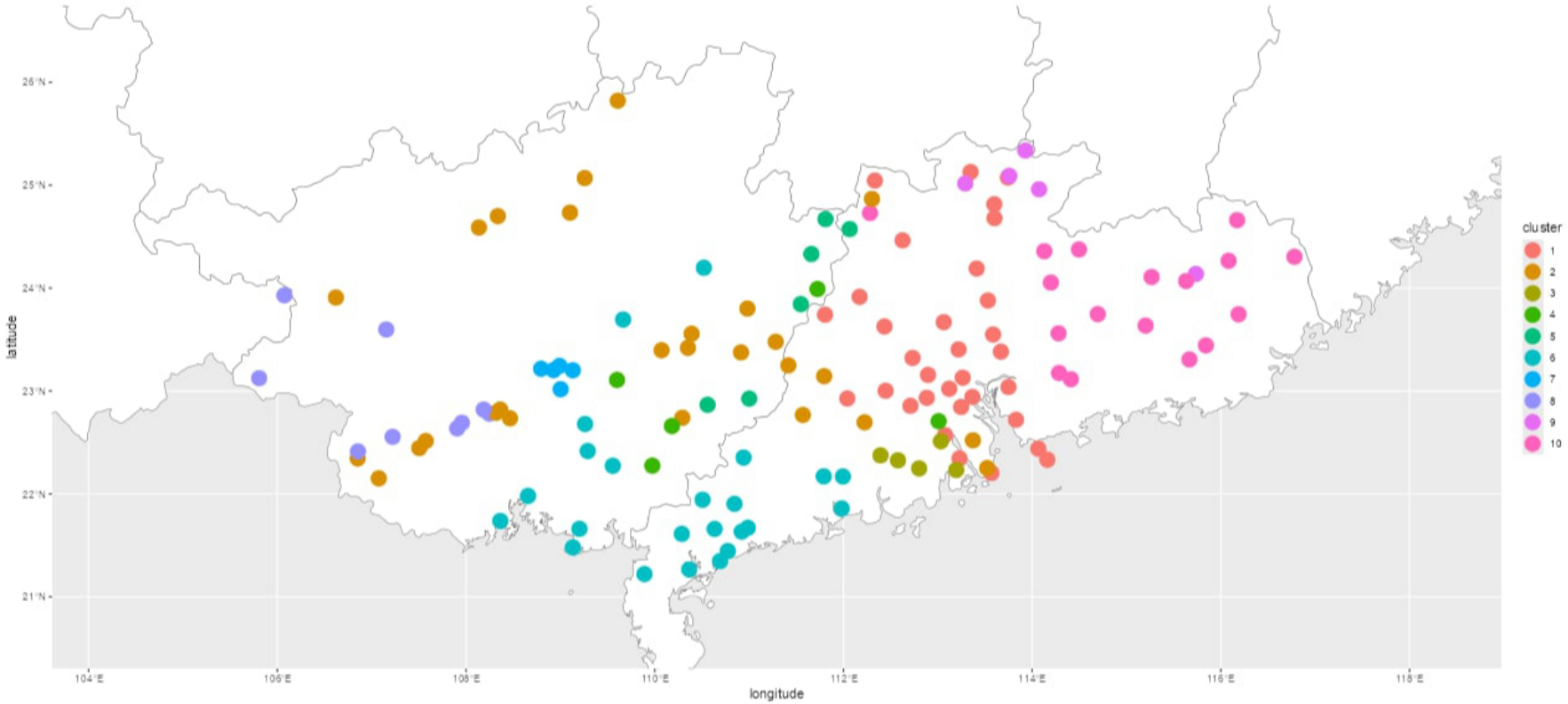
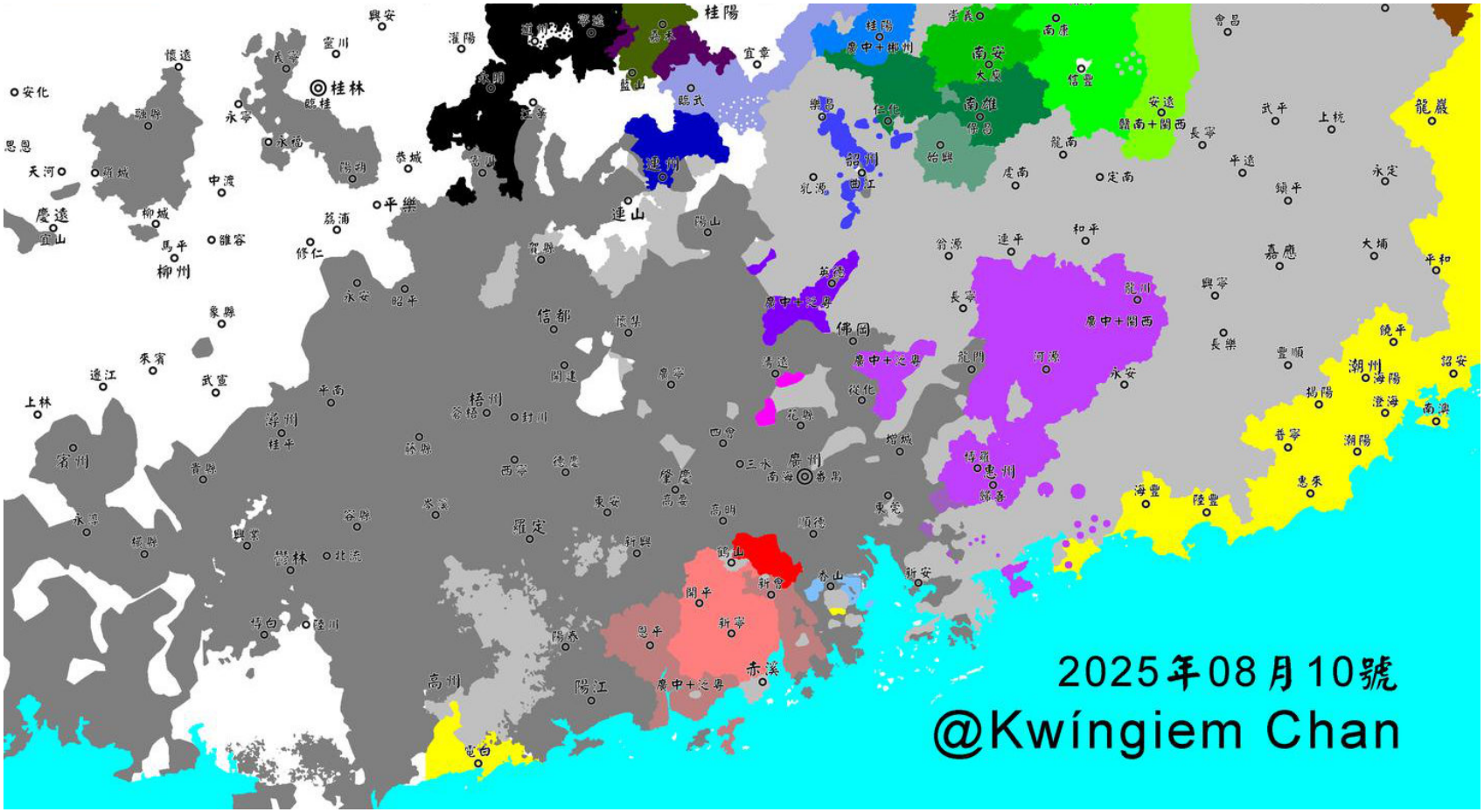


Figure 9: Cluster Map (Ward's Method)

Yue–Pinghua Chan 2025

Historical linguistics, e.g. Chan 2025



Chan, Kwíngiem 陳泂龔. 2025. 泛粵語的語言特徵、分區分片與原始語構擬. Paper presented at Forum on Languages in South China (FoLiSC)-5. https://zhuanlan.zhihu.com/p/2037678751337722052, https://www.bilibili.com/video/BV1yce5zmEcK/?vd_source=921d60258b6f28b75a778221169178bd

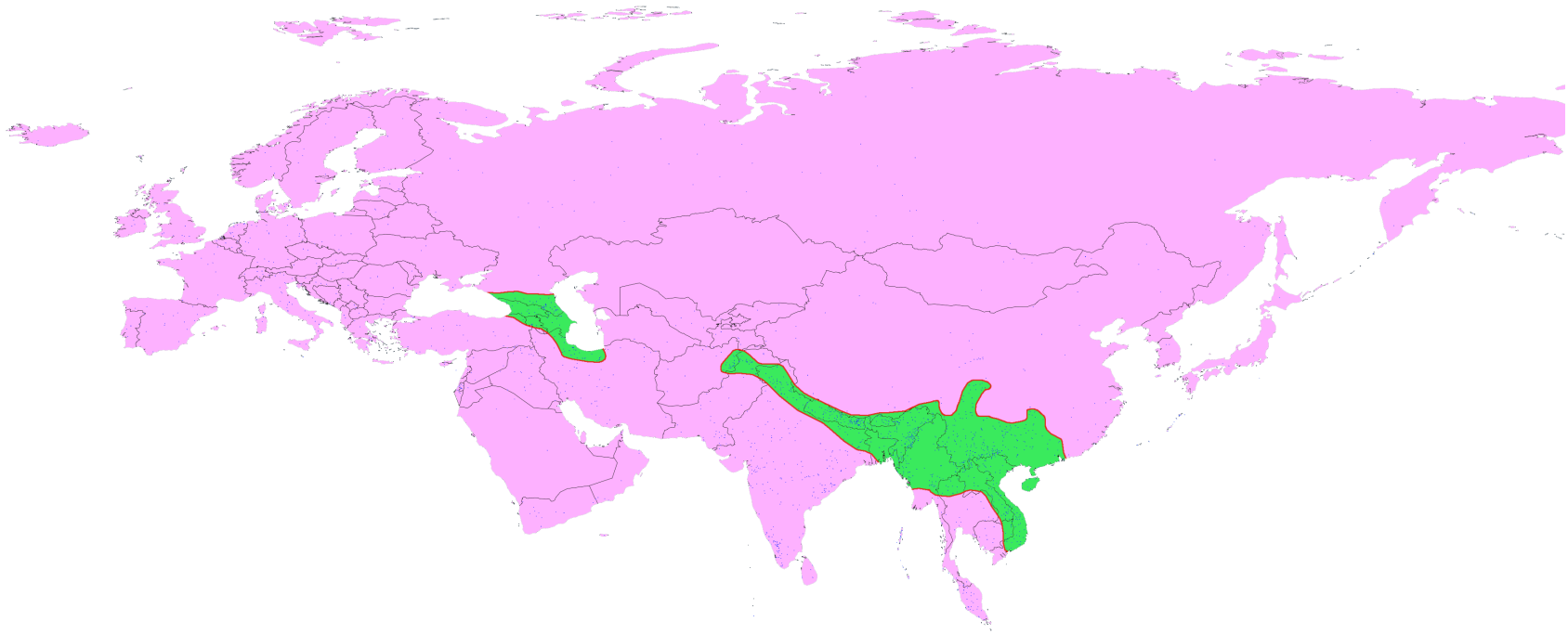
This map comes from something else: https://zhuanlan.zhihu.com/p/30911789139

Yue–Pinghua Chan 2025

Pan-Yuè innovations: (here talking about just one)
(rather) consistent vowel breaking of *i* in *-iC*:
-iC > *-jəC* (> *-əC*)
nucleus something like [ɐ ə ɯ]

	‘nine’ 九	*surname* 陳	‘ten’ 十	‘gold’ 金
MC	kjuw^{B}	ɖin^{A}	dʑip^{D}	kim^{A}
HK Hakka	kiw˩	ts^{h}in˨	sip˥	kim˧
Huizhou City	kiw˧	tɕ^{h}in˨	ɕip˩	kim˦
Taishan	kiw˥	ts^{h}in˦	sip˧	kim˦
Danzhou ^{Lt}	kjew˥	sɔn˨	sip˥	kim˦
Sino-Viet	*cửu*	*trần*	*thập*	*kim*
Danzhou ^{Col}	kɔw˦	sɔn˥	tɔp˥	kɔm˧
Cantonese	kɐw˧	ts^{h}ɐn˨	sɐp˦	kɐm˥
Yangjiang Yue	kjɐw˩	ts^{h}ɐn˧	sɐp˥	kjɐm˦
Nanning Ping	kəw˦	tʃɐn˩	ɬɐp˦	kɐm˧

Language?



There are more languages native to the green area than to the pink area

Roughly 900 languages each

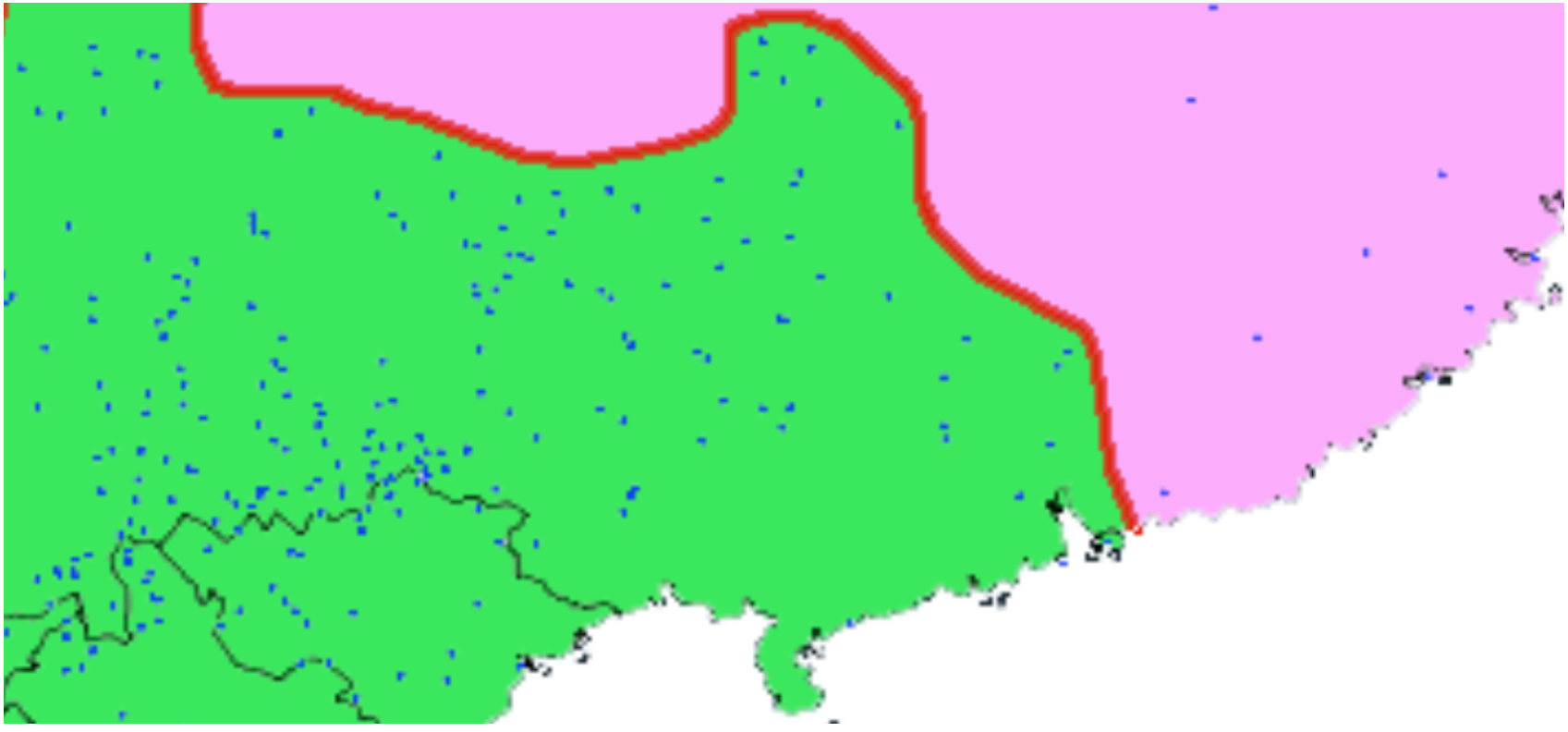
Each blue dot represents a language

Based on glottolog language location data

by reddit u/LIST- (?)

https://www.reddit.com/r/LinguisticMaps/comments/ge1j0d/there_are_more_languages_native_to_the_green_area/

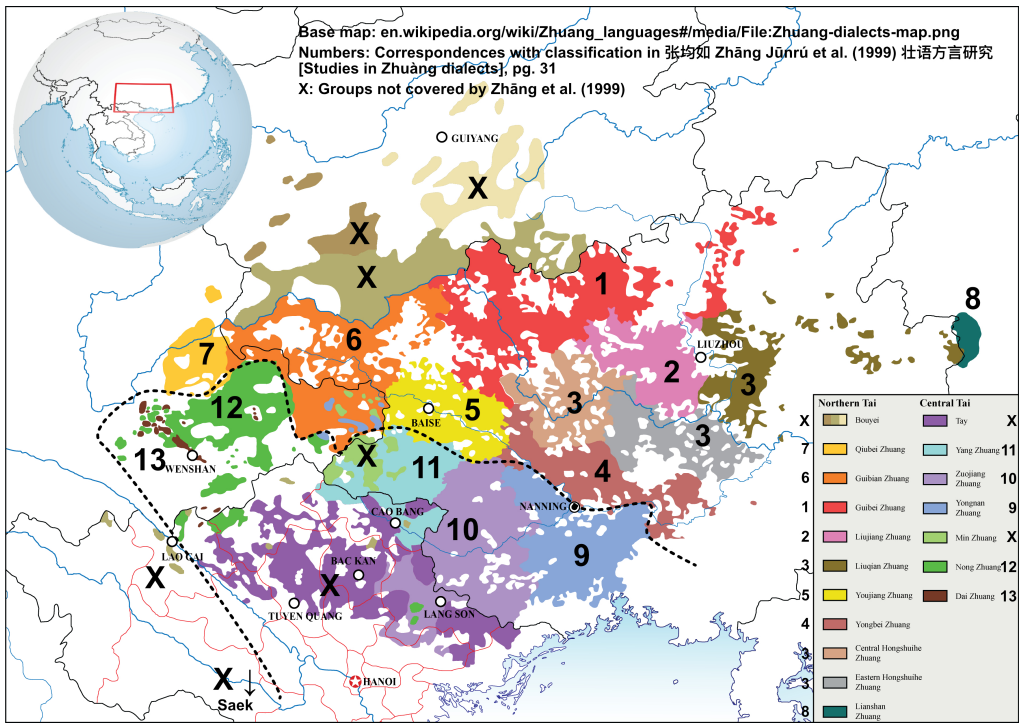
Language?



Language?

Glottolog 5.4 counts **10** languages amongst what China considers Northern Zhuàng.

- Northern Daic-Sek (15)
  - Northern Daic (14)
    - Northern Tai (11)
      - **Bouyei**
        - **Guibei Zhuang** N Zhng
      - Hongshui He (3)
        - **Liuqian Zhuang** N Zhng
        - Western Hongshui He (2)
          - **Central Hongshuihe Zhuang** N Zhng
          - **Eastern Hongshuihe Zhuang** N Zhng
      - Lianshan-Liujiang (2)
        - **Lianshan Zhuang** N Zhng
        - **Liujiang Zhuang** N Zhng
      - Unclassified Northern Tai (1)
        - **E**
      - Yei Zhuang (2)
        - **Guibian Zhuang** N Zhng
        - **Qiubei Zhuang** N Zhng
      - **Youjiang Zhuang** N Zhng
    - Yongnan-Yongbei (2)
      - **Yongbei Zhuang** N Zhng
      - **Yongnan Zhuang** *(depends on classf.)*
    - **Zuojiang Zhuang** *(def. not N Daic)*
  - **Saek**

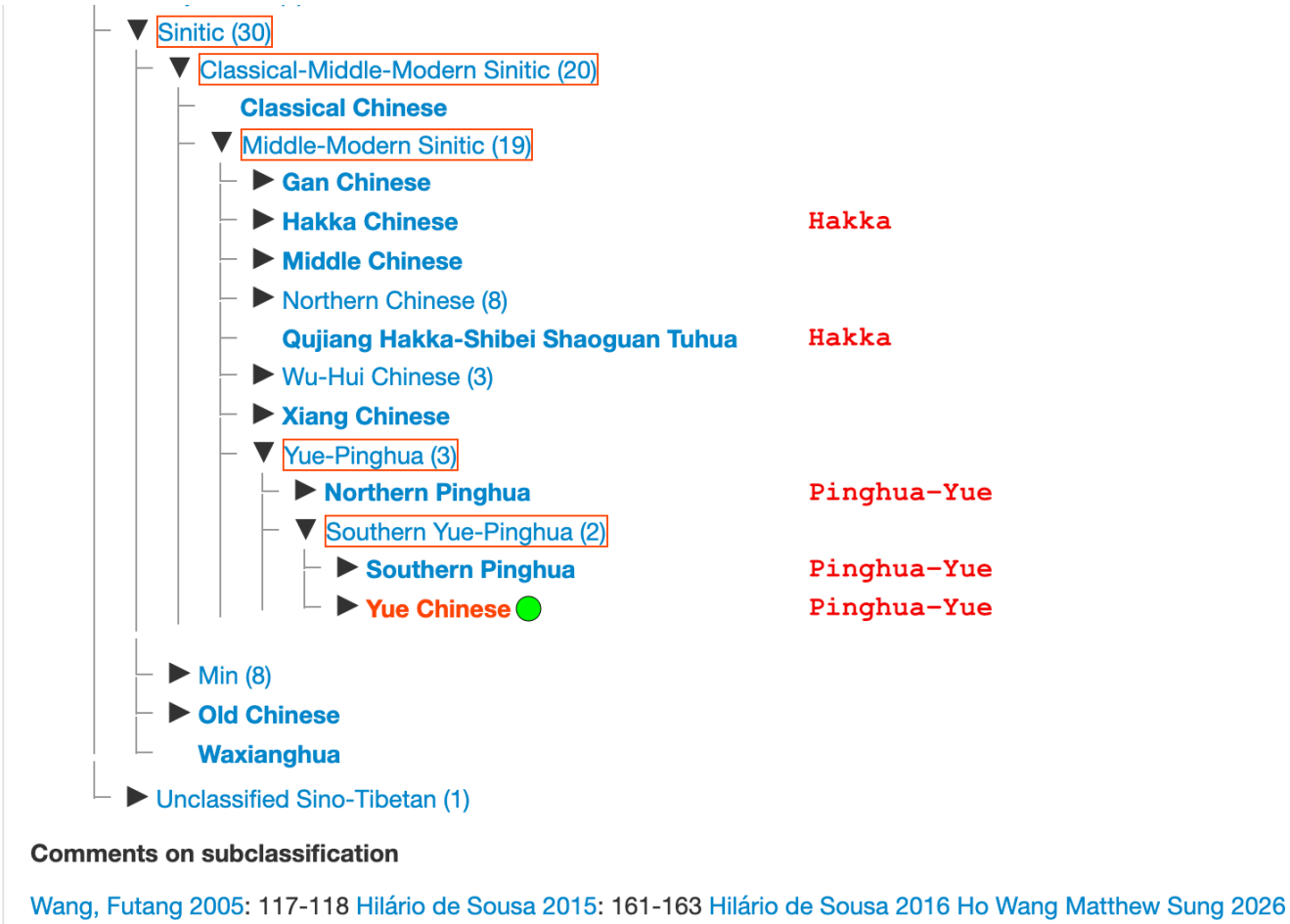


(North of the dotted line is what Glottolog calls 'Northern Daic-Sek')

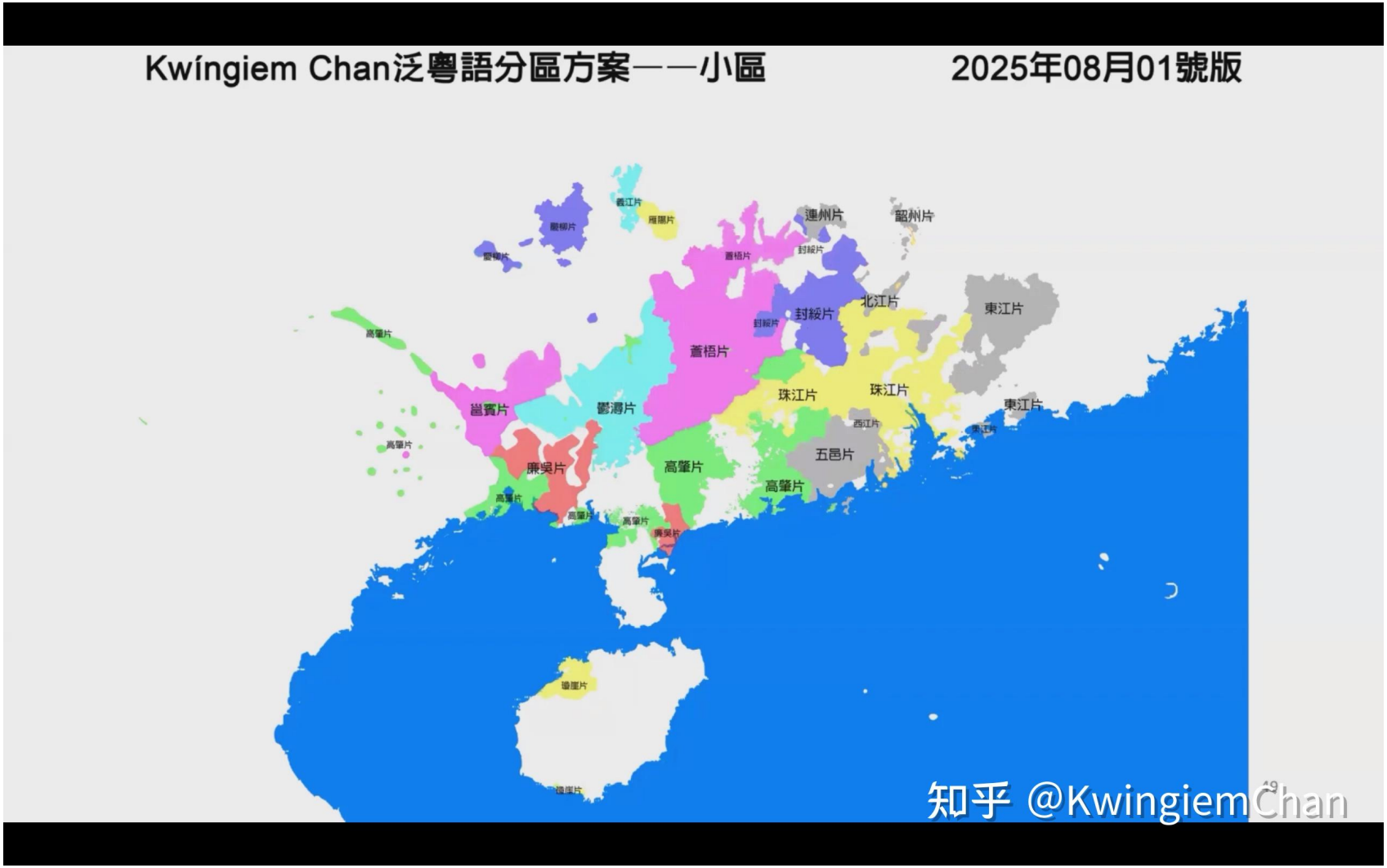
https://hilario.bambooradical.com/study-notes.html#zhuang-dialects/

Language?

Glottolog 5.4 counts (just!) **3** languages for Pínghuà–Yuè, and **2** languages for (what LAC considers) Hakka 😬 😢



Yue–Pinghua Chan 2025



Something like this? Yuè–Pínghuà as 8? 10? 12? languages?

https://zhuanlan.zhihu.com/p/685087866