Establishing a Sprachbund in the Western Lingnan region : conceptual and methodological issues

Szeto, Pui Yiu

De Gruyter Mouton
2022-04-26

Szeto, P Y & Yurayong, C 2022, 'Establishing a Sprachbund in the Western Lingnan region : conceptual and methodological issues', Folia Linguistica, vol. 56, no. 1, pp. 25-55. https://doi.org/10.1515/flin-2021-2

http://hdl.handle.net/10138/351128
10.1515/flin-2021-2003

publishedVersion

Pui Yiu Szeto* and Chingduang Yurayong

Establishing a Sprachbund in the Western Lingnan region: conceptual and methodological issues

https://doi.org/10.1515/flin-2021-2003
Received January 31, 2021; accepted June 1, 2021; published online November 29, 2021

Abstract: It is well-known that Tai-Kadai languages have affected the typological profiles of Southern Sinitic varieties. For example, compared with their northern sisters, Southern Sinitic varieties display a stronger tendency towards head-initial structures, as in the N–N compounds for expressing the sex of animals and in post-verbal temporal adverbs. Given that the Tai-Kadai languages in China have been in contact with Sinitic for over two millennia, it is quite natural to find signs of Sinitic influence therein. Most remarkably, pre-verbal adjunct phrases and pre-nominal relative clauses, which are extremely atypical of VO languages but distinctive of Sinitic, are attested in some Tai-Kadai languages in Southern China. The prevalence of such typologically unusual traits among different linguistic groups in the Lingnan region of Southern China provides strong support for its status as a linguistic area. Devising and adopting a 'mutualist' approach, we analyse the typological data of over 280 language varieties, which we believe illustrates and strongly supports the idea that Western Lingnan qualifies as a linguistic area in its own right according to criteria widely recognized by areal linguists. The approach proposed in this study can be applied to other putative linguistic areas around the world to study the mechanisms and outcomes of contact-induced change under a specific set of ecological conditions.

Keywords: language contact; linguistic area; linguistic typology; Sinitic; Tai-Kadai

1 Introduction

The Lingnan region, whose geographical range roughly corresponds to that of the present-day Chinese provinces of Guangdong (including the special administrative

***Corresponding author: Pui Yiu Szeto**, Department of Linguistics, The University of Hong Kong, Pokfulam, Hong Kong; and Department of Asian and North African Studies, Ca' Foscari University of Venice, Venice, Italy, E-mail: szetopy@connect.hku.hk. https://orcid.org/0000-0001-8444-4060
Chingduang Yurayong, Research Institute for Languages and Cultures of Asia, Mahidol University, Nakhon Pathom, Thailand; and Department of Cultures, University of Helsinki, Helsinki, Finland, E-mail: chingduang.yur@mahidol.edu. https://orcid.org/0000-0002-0676-981X

Figure 1: The Lingnan region.

regions of Hong Kong and Macau), Guangxi, and Hainan (Figure 1), represents an area of considerable ethnolinguistic diversity. This area was home to a range of aboriginal tribes collectively known as Baiyue 百越 'hundred Yue' in Chinese historiography (Ramsey 1987). The Song of the Yue Boatman (*Yuèrén gē* 越人歌), which was arguably sung in an early form of Tai-Kadai (Zhengzhang 1991), could represent the best preserved record of the Baiyue language(s). This important piece of linguistic evidence, combined with findings from toponymy[1] (Shao and Gan 2018) and ethnology (Chen 1999; Wang 2004), strongly suggests that a significant proportion of Baiyue people were ancestors of the present-day Tai-Kadai-speaking population. The Chinese conquest of the Lingnan region during the Qin dynasty (221–206 BCE) paved the way for the first major wave of Han Chinese migration into the region, resulting in contact and admixture between the Han Chinese and indigenous peoples for over two millennia. Genetic evidence provides strong support for such a contact and admixture scenario, in which the genetic structure of the Southern Han Chinese (especially those from the Lingnan region), while noticeably different from the Northern Han Chinese (Chen et al. 2009), displays signs of convergence towards the Tai-Kadai-speaking

1 For example, place names beginning with the Chinese character 那 (Cantonese *naa5*) are commonly found in Guangdong, Guangxi, and Hainan; 那 is likely to be related to *na*ㄴ, which means 'field' in Zhuang. Another example is 六 (Cantonese *luk6*), which may be related to the Zhuang word *luək*ㅓ 'valley'.

population (Gan et al. 2008; He et al. 2020; Sun et al. 2013; Wen et al. 2004). Currently, Tai-Kadai languages in China are mainly spoken in the western part of the Lingnan region, where it continues to be a hot spot of ethnolinguistic contact.

This study is chiefly concerned with the language contact phenomena in the Lingnan region. In particular, we focus on the mutual convergence between the Sinitic and Tai-Kadai languages in the region,[2] thereby evaluating the status of the Lingnan region as a potential linguistic area. Through this case study, we also aim to contribute towards the discussion of the notion of 'linguistic area', an issue which lacks general consensus in the field of areal linguistics.

2 Mutual convergence between Sinitic and Tai-Kadai

While there is considerable territorial overlap between Sinitic and Tai-Kadai in the Lingnan region, in order to argue for the existence of a linguistic area, we have to demonstrate the occurrence of typological convergence therein (see Section 3 for further discussion). In other words, the Sinitic languages in the region (hereafter Lingnan Sinitic) are expected to be structurally more similar to Tai-Kadai than their sister languages spoken elsewhere, whereas the Tai-Kadai languages in the region (hereafter Lingnan Tai-Kadai) should be structurally more similar to Sinitic. Given also earlier studies on Tai-Kadai substrate vocabulary items in Lingnan Sinitic (see e.g. Bauer 1996; Li 1990; Bian et al. 2011), the former phenomenon is in fact quite well-established. The typological variation within the Sinitic branch can be explained from the perspective of language contact, where Northern Sinitic shows signs of convergence towards Altaic languages (i.e. Altaicization) and Southern Sinitic towards Mainland Southeast Asian (especially Tai-Kadai) languages (i.e. Taicization) (Hashimoto 1976, 1985; cf. Chappell et al. 2007; de Sousa 2015). Although the validity of Altaicization is sometimes questioned (Bennet 1979), the computer-aided analyses conducted by Szeto (2019) provide strong support for both Altaicization and Taicization. Table 1 summarizes the general tendencies of this north-south contrast.

Studies on the opposite direction of contact influence are mainly about lexical borrowings (e.g. Li 1976; Bauer 1996; Pittayaporn 2014). However, given the dominance of Sinitic in the Lingnan region, it is not surprising that its influence on Tai-Kadai extends to various grammatical domains as well (Enfield 2019; Li 2008). This section reviews several typological features that reflect mutual convergence between Sinitic and Tai-Kadai in the region.

2 Languages of other genetic affinities are also spoken in the Lingnan region. We will discuss why Sinitic and Tai-Kadai warrant special attention in later sections of the paper.

Table 1: The north-south contrast in Sinitic (based on Chappell 2015: 17).

Northern Sinitic (Altaicization)	Southern Sinitic (Taicization)
Stress-based and fewer tones	More tones
Larger proportion of polysyllabic words	Larger proportion of monosyllabic words
Simpler syllable structure	More complex syllable structure
Smaller inventory of classifiers	Larger inventory of classifiers
Predominantly modifier–modified	Modified–modifier order possible
Pre-verbal adverbs	Post-verbal or clause-final adverbs possible
IO–DO order in ditransitives	DO–IO order in ditransitives
Marker–Standard–Adjective comparatives	Adjective–Marker–Standard comparatives
Causative speech-act verbs as passive marker	'Give' verbs as passive marker

2.1 Taicization of Lingnan Sinitic

Both Sinitic and Tai-Kadai languages are predominantly VO. Sinitic languages intriguingly exhibit head-final characteristics in nominal structures (associated with OV order) and a mixture of head-initial (associated with VO order) and head-final characteristics in verbal structures (Chappell et al. 2007; Dryer 2003) (see Table 2).[3] Tai-Kadai languages, by contrast, are typically strictly head-initial (see Section 2.2).

Given areal influence from Tai-Kadai and other southern minority languages, it is only natural for Lingnan Sinitic to show a relatively strong tendency towards head-initial structures. For instance, although the modifier–modified word order is

Table 2: Word order features of Sinitic (modified from Chappell et al. 2007: 189).

Head-final structures	Head-initial structures
Adjective–Noun[a]	Verb–Object
Numeral–Classifier–Noun	Auxiliary–Verb
Demonstrative–Classifier–Noun	Verb–Modifying adverbial complements
Relative clause–Noun	Preposition–Noun phrase
Genitive–Noun	Complementizer–Sentence
Adverb–Verb	
Intensifier–Adjective	
Standard of comparison–Adjective	
Prepositional phrase–Verb	

[a]Adjective–noun order actually does not correlate with OV languages (Dryer 1992). Nonetheless, it still represents a noteworthy word-order feature of Sinitic since it is rare in the rest of the Sino-Tibetan family.

3 Only major patterns of Sinitic are listed in the table. Minor patterns like verb-auxiliary and verb-adverb are not listed therein.

dominant in virtually all Sinitic varieties, the modified–modifier order is present in some structures of the southern varieties, such as the N–N compounds for expressing the sex of animals,[4] which is consistent with the word order of their Tai-Kadai neighbours as shown in (1) (Matthews 2007):

(1) a. *ki˧˥-**koŋ**˧˥* [Liuzhou Mandarin]
b. *kai˧-**pau**˩* [Yongbei Zhuang]
chicken-**male**
'rooster'

As Peyraube (2015) demonstrates, post-verbal adverbs are common in Tai-Kadai and Hmong-Mien languages. Lingnan Sinitic also tends to follow the head-initial structure in verbal phrases involving the temporal adverb 'first'. As noted by Matthews (2007: 229), "Perhaps the most well-known peculiarity of Cantonese syntax is the adverb *sin1* 'first' which almost uniquely follows the verb", which presents a case of salient departure from standard Chinese usage but closely matches that of Thai:

(2) a. *ngo5 zau2 **sin1*** [Cantonese]
b. *pʰom˧ pai˧˥ **kɔːn**˨˩* [Thai]
1SG go **first**
'I am going/leaving first.'

Regarding the double object dative constructions, the predominant word order in Northern Sinitic is verb–indirect object–direct object, i.e. [V IO DO] (see [3]). The [V DO IO] order, on the other hand, is a southern feature which is relatively rare in the Sinitic branch but common in Tai-Kadai languages, see (4). However, this order is arguably not a 'foreign' feature – it is found in Old Chinese, existing alongside the [V IO DO] and [V DO P IO] orders as a minority pattern (Xu and Peyraube 1997). Instead of introducing a new feature to the Sinitic branch, the non-Sinitic languages may have triggered the development of the [V DO IO] order from a minor to a major use pattern in Lingnan Sinitic, a phenomenon commonly observed in situations of language contact (cf. Heine and Kuteva 2005).

(3) Standard Mandarin
wǒ gěi tā qián
1SG give 3SG money
'I give him money.'

4 Interestingly enough, even Tibeto-Burman languages in Mainland Southeast Asia, such as those in the Baic, Lolo-Burmese, and Kuki-Naga branches, also consistently use post-nominal animal sex marking despite their head-final syntax; so this isogloss seems to have a broader extension than just the Lingnan region.

(4) a. *ngo5 bei2 cin2 keoi5* [Cantonese]
b. *pʰomꜜ haiꜘ ŋɤnꜛ kʰawꜜ* [Thai]
1SG give money 3SG
'I give him money.'

The effect of Taicization is evident in other domains of grammar as well. The distribution of several phonemes in Sinitic shows a north-south contrast, which corresponds well to the Altaicization-Taicization hypothesis. The retroflex fricative initial [ʂ-], which is common in Northern Sinitic (though not necessarily so in some Northeast Mandarin dialects) but rare in Lingnan Sinitic (Cao 2008: P045), is a case in point. The prevalence of the retroflex fricative in Altaic languages (e.g. Manchu *ʂun* 'sun') may have favoured the retention of this phoneme in Northern Sinitic, which was also present in Middle Chinese (Pulleyblank 1991). The high front rounded vowel [y] is another phoneme which was present in Middle Chinese but lost in many Lingnan Sinitic varieties (Cao 2008: P117). Interestingly, this vowel is very rare in Tai-Kadai languages but common in Turkic languages (e.g. Uyghur *køŋyl* 'mood'). In fact, [y] has been reconstructed to Proto-Turkic, Proto-Mongolic, and Proto-Tungusic but may have shifted its place of articulation as a result of vowel rotation, that is from palatal-velar to tongue root harmony in some modern Mongolic and Tungusic languages (Barrere and Janhunen 2019).

Another areal feature of Lingnan Sinitic likely to be related to Tai-Kadai influence is the non-distinction between the morphemes for plain negative and existential negative. In Far Southwestern China (Western Guangdong, Hainan, Guangxi, and adjoining regions in Hunan, Guizhou, and Yunnan), most Sinitic varieties deploy one and the same morpheme for the plain and the existential negative (Cao 2008: G033) (see [5–6]), unlike what we see in Sinitic varieties spoken elsewhere, which are by and large consistent with the Standard Mandarin pattern,[5] shown in (7).

(5) Liuzhou Mandarin
a. ***mɐi***ꜜ *toŋꜛ* [Plain negative]
NEG know/understand
'to not know/understand'
b. ***mɐi***ꜜ*-tə*ꜘ *kiaꜛkiaꜜ* [Existential negative]
NEG-get family.education
'poor upbringing (*lit.* to have no family education)'

5 This is inherited from the Old Chinese negation system, where morphemes with p/f initial denoted plain negation while those with m/w initial denoted existential negation (Pulleyblank 1995).

(6) Haikou Min

a. ***vo***˨˩ *ka*˥ [Plain negative]
NEG suitable
'not suitable'

b. ***vo***˨˩ *huaŋ*˧ ***vo***˨˩ *iaŋ*˧˩ [Existential negative]
NEG wind NEG wave
'windless and waveless'

(7) Standard Mandarin

a. ***bù*** *xīwàng* [Plain negative]
NEG hope
'to hope not'

b. ***méi****-yǒu* *xīwàng* [Existential negative]
NEG-exist hope
'hopeless'

The distribution of this feature across Sinitic closely coincides with the geographical range of the Tai-Kadai languages spoken in China. This observation, coupled with the fact that Tai-Kadai languages tend not to distinguish between the plain and the existential negative morphemically, strongly suggests that this areal feature is linked with Tai-Kadai influence:

(8) Lao

a. *kʰɔj*˦˩ ***bɔː***˦ *mɛn*˦ *naːŋ˧faː*˩˧ [Plain negative]
1SG NEG be angel
'I am no angel.' (title of the Lao song ຂ້ອຍບໍ່ແມ່ນນາງຟ້າ)

b. *kʰɔj*˦˩ ***bɔː***˦ *miː*˩˧ *ŋɤn*˩˧ [Existential negative]
1SG NEG exist money
'I don't have money.'

2.2 Sinicization of Lingnan Tai-Kadai

Compared with Sinitic, Tai-Kadai languages tend to be much more predominantly head-initial (see [9–12]). Nonetheless, under contact pressure from Sinitic for over two millennia, many Tai-Kadai languages in China have developed head-final structures, often coexisting side by side with the native head-initial structures (13–14) (see also Enfield 2019).

(9) Tai Phake (Morey 2008: 222)
Noun–Adjective
kʰon² *suŋ⁶*
person tall
'a tall person'

(10) Lao
Noun–Relative clause
ma:˨ *to:˨* *tʰi:˦* *kʰɔ:j˩* *hen˨*
dog CLF REL 1SG see
'the dog that I saw'

(11) Nung (Saul and Freiberger Wilson 1980: 29)
Verb–Adverb
mạhn *má* ***hạu*** *lái*
PL dog **bark** much
'The dogs are barking very much.'

(12) Thai
Verb phrase–Prepositional phrase
kin˦ ***kʰa:w˩˧*** *tʰi:˩˧* *ba:n˩˧*
eat **rice** at home
'to eat at home'

(13) Mulao (Bo 2002: 59–60)
a. **Noun**–Adjective
naŋ˨ *ŋe˩˧*
dog yellow
'yellow dog'
b. Adjective–**Noun**
ŋe˩˧ ***naŋ˨***
yellow **dog**
'yellow dog'

(14) Sui (Zhang 1980: 44)
a. **Adjective**–Intensifier
di˩ *ɕo˦*
far very
'very far'

b. Intensifier–**Adjective**
naŋ˥ ***tik˥***
very **full**
'very full'

Even more striking is the presence of two extremely unusual word-order features in these Tai-Kadai languages, which are unmistakably transferred from Sinitic. Among the head-final structures in Sinitic (see Table 2), the 'VO + pre-nominal relative clause' (15) (Dryer 2013) and the 'oblique (adjunct phrase) + VO' (16) (Dryer [with Gensler] 2013) combinations are almost unique cross-linguistically (see also Hawkins 1994). The presence of these features in Lingnan Tai-Kadai reflects the power and reality of Sinicization (17–18) (see also Wu 2008b, 2009).

(15) a. ***[kǎoshì bù jígé de]*** *xuéshēng* [Standard Mandarin]
b. ***[haau2si3 m4 hap6gaak3 ge3]*** *hok6saang1* [Cantonese]
exam NEG **pass** PRT student
'students who failed the exam'

(16) a. *wǒ* ***[zài jiā]*** *chīfàn* [Standard Mandarin]
b. *ngo5* ***[hai2 uk1kei2]*** *sik6faan6* [Cantonese]
1SG LOC **home** eat.rice
'I eat at home.'

(17) Yongbei Zhuang (Zhang et al. 1999: 404)
[kjoŋ˧ dam˧ na˨] *ɕuŋ˥* *tuk˦* *wun˨* *ba:n˥* *ɣau˨*
CLF **sow** **field** all be person village 1PL
'The people sowing the field are all from our village.'

(18) Cun (Ouyang 1998: 236)
ku˦niaŋ˧˩-mən˦ *tsiŋ˧* ***[tθai˧ tθən˦ bian˦]*** *tshaŋ˨* *kɔ˦*
young.lady-PL now LOC **village side** sing song
'The young ladies are singing on the edge of the village.'

Another notable example of word-order shift is observed in the comparative construction. The Marker–Standard–Adjective order in (Northern) Sinitic, as in (19), is highly uncommon cross-linguistically (Ansaldo 1999, 2010); but has been introduced to many Tai-Kadai languages in China, most likely via Southwest Mandarin (20–21).

(19) *wǒ* ***bǐ*** *tā* *gāo* [Standard Mandarin]

(20) *jie˩* ***pi˥*** *mən˩* *wu:ŋ˥* [Chadong] (Li et al. 2012: 156)
1SG **COMP** 3SG tall
'I am taller than him/her.'

(21) Langjia Buyang (Li J 1999: 75)
ali˩˧ *mu˥* *pi:ŋ˩* *ni˩˧* ***pi˧˥*** *pi:ŋ˩* *qon* *li˩˧* *ŋa˩˧* *manɛn˥*
CLF pig year this **COMP** year previous CLF that fat
'The pig this year is fatter than the one last year.'

3 Establishing a linguistic area

After reviewing a range of features indicative of mutual convergence between Sinitic and Tai-Kadai, we now move on to visualize their areal patterns, thereby identifying a zone of significant convergence. We will also take additional data into account when discussing whether the convergence zone[6] should be considered a linguistic area. Finally, we will discuss the 'mutualist' approach adopted in this study for identifying a linguistic area.

3.1 Visualizing the areal patterns

Mutual convergence between Sinitic and Tai-Kadai mainly occurs in Southern China, especially the Lingnan region. Although this general statement is very likely to be true, it is helpful to take a closer look at the empirical data. In order to achieve a comprehensive coverage of these two language groups, we consider all the Sinitic and Tai-Kadai varieties recorded in the database of Szeto and Yurayong (2021), with a handful of additional datapoints (Figure 2) (see Supplementary Material 1 for the information of the language varieties). At the same time, we take into account six Taicization and seven Sinicization features (Table 3), which can be readily located within the *Linguistic Atlas of Chinese Dialects* (Cao 2008) and the respective reference grammars of various language varieties. The Taicization score

6 We use the term 'convergence zone' in a general sense to refer to a region which sees structural convergence between languages, while 'linguistic area' and 'Sprachbund' are technical terms with more specific meanings. See Sections 3.4–3.5 for further discussion.

Figure 2: Geographical location of the Sinitic and Tai-Kadai varieties.

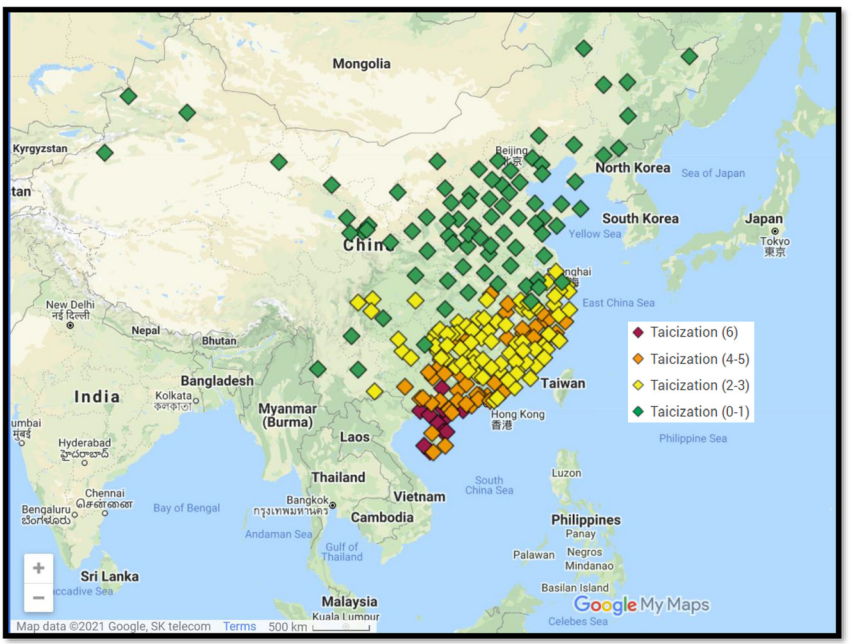

Figure 3: Distribution of Taicization scores.

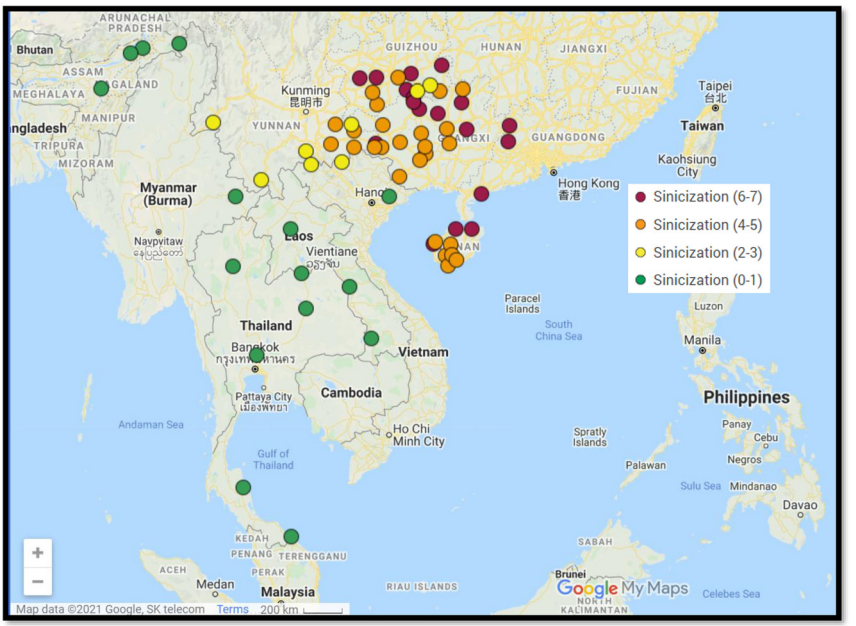

Figure 4: Distribution of Sinicization scores.

Table 3: The Taicization and Sinicization features.

Taicization	Sinicization
Absence of retroflex fricative initial	Numeral–Classifier–Noun
Absence of high front rounded vowel	Pre-nominal demonstrative
Modified–modifier order in animal sex marking	Genitive–Noun
Post-verbal temporal adverb in verb phrase	Intensifier–Adjective
[V DO IO] order in double object dative constructions	Pre-nominal relative clause
Identical morphemes for plain negative and existential negative	Pre-verbal adjunct phrase
	Standard–Adjective order in comparatives

of a Sinitic variety equals the number of Taicization features present in it, and vice versa. The distribution of Taicization scores is shown in Figure 3 and that of Sinicization in Figure 4.[7]

7 A similar evaluation method of areal features by scores was earlier applied to the investigation of Balkanization of languages in the Balkan Sprachbund by Lindstedt (2000) with satisfactory results.

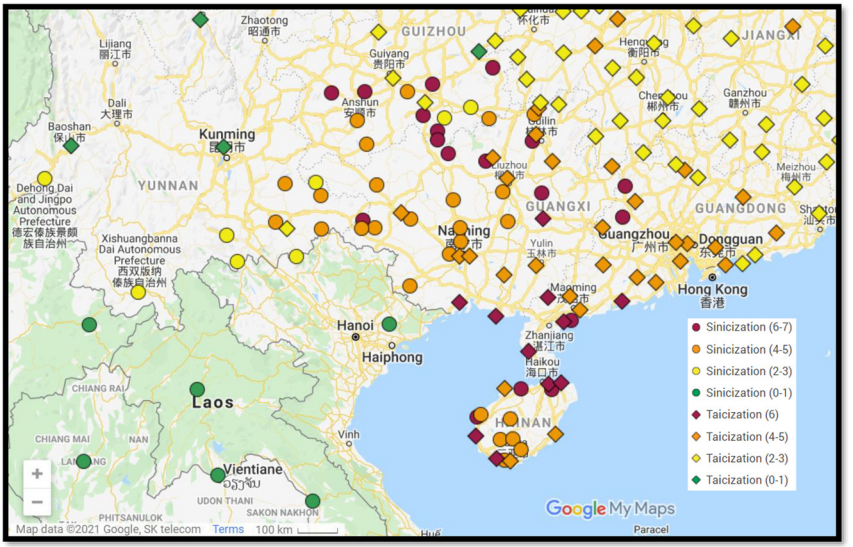

Figure 5: Distribution of Taicization and Sinicization scores.

A noticeable areal pattern can be observed in Figures 3 and 4, i.e. the effects of both Taicization and Sinicization are confined to Southern China.[8] Remarkably, though not unexpectedly, all the Tai-Kadai languages spoken outside China show hardly any sign of Sinicization, highlighting a clear typological divide between them and their sister languages spoken inside China. Zooming in, we can see that the core zone of convergence largely coincides with the geographical range of the Lingnan area (except that the former does not cover Central Eastern Guangdong) (see Figure 5).[9] Unlike their more Taicized sisters, Sinitic varieties in Central Eastern Guangdong have generally retained the high front rounded vowel and distinct morphemes for the plain and the existential negative.

3.2 Bringing two supplementary areal features into the picture

So far the focus has been on typological features which are manifestly attributable to either Sinitic or Tai-Kadai influence. While this is a reasonable approach given the prominence of these two groups of languages in the region, to better

8 See Supplementary Material 2 for the Taicization score of Sinitic languages, and Supplementary Material 3 for the Sinicization score of Tai-Kadai languages.

9 Available online at https://www.google.com/maps/d/u/0/edit?mid=1uE-UZUqQew7FQ76019Rz-REO8426nAMr&usp=sharing.

circumscribe a potential linguistic area, it may be helpful to consider some supplementary features as well. We managed to identify two features that seem particularly common in the region, but not among the Sinitic and Tai-Kadai languages spoken elsewhere.

In Sinitic, the voiceless lateral fricative initial [ɬ-] is largely confined to Guangxi and Southwestern Guangdong (Cao 2008: P046) (e.g. Nanning Pinghua *ɬam*˧˥ 'three', Yulin Yue *ɬam*˧˧ 'heart'). Interestingly, for Tai-Kadai, this phoneme is also more commonly found in the Lingnan region (e.g. Yongnan Zhuang *ɬi:n*˧ 'garden', Jizhao *ɬa*˨˩ 'to touch'). It is noteworthy that the distinction between voiceless lateral fricative [ɬ] and voiceless lateral approximant [l̥] sometimes merely reflects different linguists' terminology and transcription practices (Maddieson and Emmorey 1984). Despite the phonetic differences between this pair of laterals, we have to resort to treating [ɬ] and [l̥] as a single set of phonemes known as voiceless laterals since this study involves a large number of languages documented by linguists of different backgrounds. Such a non-distinction is arguably unproblematic for the purpose of this study especially given that no language in the world is known to make a phonemic contrast between these two phonemes (Maddieson 1984). Taking such an approach, the presence of voiceless laterals in some Sinitic varieties in the region might be considered a manifestation of Taicization because [l̥-] is believed to be present in Proto-Tai (Li 1977; Pittayaporn 2009). However, the situation is complicated by the fact that this phoneme can also be reconstructed to Proto-Hmong-Mien (as [ɬ-] according to Chen (2013), and as [l̥-] according to Wang and Mao (1995); the aspirated lateral [hl-] reconstructed by Ratliff (2010) represents another way of transcribing the phoneme) (see Section 3.3 for further discussion about the role of Hmong-Mien in this potential linguistic area), and the rarity of this phoneme among Tai-Kadai languages spoken outside the region. Therefore, we tentatively consider voiceless laterals an 'area-specific' feature.

Further, in many Lingnan Sinitic varieties, the 'bare classifier' construction [CLF N] can occur in subject position to mark definiteness (as shown in [22]) (Matthews 2007; Cao 2008: G014; cf. Wang 2015). This construction is also common among the Tai-Kadai languages in the region (23). Unlike the typological features we have discussed in Section 2, the origin of this feature is less clear as it is not common in either the Sinitic or the Tai-Kadai languages spoken elsewhere, in which a demonstrative is usually required to signal definiteness (24–25).

(22) Cantonese

[bou6	***sau2gei1]***	*m4*	*gin3-zo2*
CLF	**mobile**	NEG	see-PFV

'The mobile is lost.'

(23) Central Bouyei (Yu 1980: 19)
[*tu*˩˧ *ma*˧˥] *taɯ*˩˧ *zaːn*˩˧
CLF **dog** watch.over home
'The dog watches over the home.'

(24) Standard Mandarin
zhè *dào* *cài* *hěn* *hǎo* *chī*
this CLF food very good eat
'This is a delicious dish.'

(25) Thai
saːw˧ *suəj*˧ *kʰon*˦ ***nan***˥ *maj*˨˩ *cʰaj*˨˩ *pʰuː*˨˩ *jiŋ*˧
girl beautiful CLF **that** NEG COP person female
'That beautiful girl is not a female (i.e. she is a kathoey).'

Given that this feature is fairly common among Hmongic languages (26–27), its presence in Lingnan Sinitic and Tai-Kadai could be related to Hmongic influence. This, however, remains a matter for conjecture.

(26) Thailand Green Hmong (Kunyot 1984: 34)
[*tu̱*˩˩ ***tu*˥]** *mo̱ŋg̱*˩˩ *mo̱ːŋg̱*˥ *yua*˧˥ *tau*˦ *ndau*˥
CLF **son** go go buy get cloth
'The son went and went a long way and bought cloth.'

(27) Qiandong Miao (Wang 1985: 56)
[*lɛ*˦ *ʨi*˦] *ɳaŋ*˦ *haŋ*˧˥*noŋ*˧˥
CLF **bowl** LOC here
'The bowl is here.'

The distribution of these two area-specific features is shown in Figure 6[10] (see also Supplementary Material 4). Naturally, two features *per se* may not suffice to reveal a clear areal pattern. Combining Figures 5 and 6, however, we can tentatively locate the 'areal hotbed' (in Güldemann's (2008) terminology) in Guangxi and its adjoining region in Southwestern Guangdong, which partly agrees with the grammaticalization area in Central Southern Guangxi identified by Huang and Wu (2018).

10 Available online at https://www.google.com/maps/d/u/0/edit?mid=1RN6VbaCToekesKnbVPHqFvYnc4kKlRTt&usp=sharing.

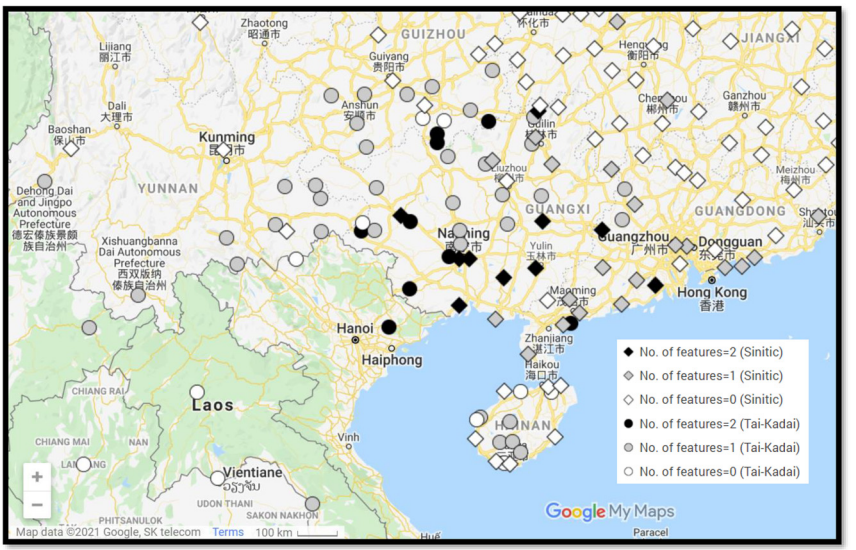

Figure 6: Distribution of the area-specific features.

3.3 Other languages in the area

In addition to Sinitic and Tai-Kadai, languages belonging to the Hmong-Mien, Austroasiatic, and Austronesian families are also found in the Lingnan region. This section discusses why these three groups of languages, especially the latter two, are only of secondary significance in the present case.

From a geographical perspective, Hmong-Mien appears to be an ideal candidate for the study of areal convergence as its territory spans Southern China and several Southeast Asian countries, much like that of Tai-Kadai. It may come as a surprise that the typological divide between the Tai-Kadai languages spoken inside and outside China, primarily caused by Sinicization, is much less obvious in Hmong-Mien languages. For example, pre-nominal relative clauses as in (28), preverbal adjunct phrases (29), and Standard–Adjective comparatives (30), the three highly marked word order features for VO languages characteristic of Sinitic (cf. Section 2.2), are all attested in Thailand Mien. This phenomenon is probably due to the relatively recent migration events of the Hmong-Mien populations, whose emigration from China only began in the 17th century (Chen 2013), around a millennium later than that of the Tai populations (Pittayaporn 2014).

(28) Thailand Mien (Saeliao 2012: 256)
[*mei*˩ *dzo*˧ *ȵɛi*˦ *lui*˦*xou*˧] *ga:i*˦ *ȵaʔ*˩
2SG **wash** **PRT** **clothes** dry PFV
'The clothes you washed have dried.'

(29) Thailand Mien (Saeliao 2012: 161)
mei˩ **[*kan*˦ *xai*˧*dau*˦]** *ta:i*˩ ***mei***˩
2SG **from** **where** come Q
'Where do you come from?'

(30) Thailand Mien (Saeliao 2012: 302)
tɕɛ˦ ***pei***˥ *a:p*˥ *kau*˩ *tɕai*˧
chicken **COMP** duck more expensive
'Chicken is more expensive than duck.'

This by no means implies that Hmong-Mien languages are irrelevant to the present case study. In fact, as we discussed above, the supplementary features we identified in Section 3.2 could be related to Hmong-Mien. Furthermore, we observe that the Classifier–Noun–Demonstrative order is commonly found in Tai-Kadai languages in China (31–32), which deviates from their canonical head-initial structure (Noun–Classifier–Demonstrative) but coincides with that of many Hmong-Mien languages (33–34).

(31) Yongbei Zhuang (Zhang et al. 1999: 410)
ʔan˧ *ɣa:n*˩ *nai*˩
CLF family this
'this family'

(32) Langjia Buyang (Li J 1999:60)
amiə˨˩ *qatui*˩ *ȵa*˨˩
CLF tree that
'that tree'

(33) Younuo (Mao and Li 2007: 65)
pəŋ˧ *no*˦ *no*˦
CLF book this
'this book'

(34) Thailand Mien (Saeliao 2012: 302)
lu˥ *ce*˧ *nua*˧
CLF house this
'this house'

Table 4: Typological features of the Austroasiatic and Austronesian languages.[a]

	Bolyu	Vietnamese	Tsat
Sinicization features			
Numeral–Classifier–Noun	+	+	+
Pre-nominal demonstrative	–	–	–
Genitive–Noun	+	–	+
Intensifier–Adjective	+	+	+
Pre-nominal relative clause	+	–	+
Pre-verbal adjunct phrase	+	+	+
Standard–Adjective comparatives	+	–	+
Area-specific features			
Voiceless lateral initial	+	+	–
[CLF N] construction in subject position with definite reference	+	+	–

[a]Information source: Bolyu (Li X 1999), Vietnamese (Ouyang et al. 1984), Tsat (Zheng 1997).

Signs of Hmong-Mien influence, notwithstanding, given the lack of a clear typological distinction between the Hmong-Mien languages inside and outside the area, they can only play a subsidiary role in identifying the potential linguistic area under the approach adopted in this study (see Sections 3.4–3.5 for further discussion).

Austroasiatic (Bolyu and Vietnamese in Guangxi) and Austronesian (Tsat in Hainan) languages are small minorities in the Lingnan region. They demonstrate obvious signs of convergence towards Sinitic, and those in the 'areal hotbed' Guangxi even feature the voiceless lateral initial and 'bare classifier' construction, two areal traits which seem peculiar to the region (see Table 4). Although these languages are not known to have structurally influenced their linguistic neighbours,[11] their typological profiles serve to provide further evidence for areal convergence in the Lingnan region.

3.4 Evaluating the status of the convergence zone

Having analysed the areal traits in the Lingnan region, we can confidently assert that the region has seen significant structural convergence between languages of different genealogical affiliations. Next, we have to deal with two related matters: does the Lingnan region constitute a linguistic area, and, more importantly, what specific kinds of contact-induced change have occurred therein?

Although its precise definition differs from linguist to linguist, the concept of linguistic area (*jazykovoj sojuz* 'language union' in its original Russian form

11 See Alves (2020) for discussion on Sinitic-Vietic contact and the role of Sinicization in the diversification of Sinicized Northern Vietic from their more conservative Southern Vietic sisters.

introduced by Trubetzkoy (1923), or Sprachbund in German) can be generally understood as a geographically delimited area containing languages sharing some structural similarities as a result of contact instead of accident or inheritance (Hickey 2017; Thomason 2001). It is noteworthy that recent studies tend to share the view that it is more worthwhile to focus our attention on the mechanisms and outcomes of contact-induced change within a given geographical area, instead of getting entangled in making a binary decision as to whether the area qualifies as a Sprachbund (Campbell 2017; Hickey 2017). Even so, it may still be revealing to assess the status of the Lingnan region. Campbell (2017) outlines four major criteria routinely utilized by linguists to determine a linguistic area, i.e. number of shared traits, trait weight, number of language families, and trait bundling. We will discuss each of these criteria in turn, with special reference to the Lingnan region.

3.4.1 Number of shared traits

As Campbell (2017: 25) puts it, "it is not a matter of some minimum number of shared traits, but rather the more, the merrier: some areas are more securely established because they contain many shared diffused traits, and other areas with fewer shared traits are weaker". Obviously, while it is possible to establish a linguistic area based on a single feature (Hickey 2017), it is always better to have a broader range of features in order to establish the linguistic area more securely. With 15 features covering the domains of phonology, word order, and grammatical morphemes, the number of shared traits we have identified in this study should be more than adequate.

3.4.2 Trait weight

Trait weight, the second criterion, should arguably be considered in connection with the first. Although the number of shared traits is often believed to be proportional to the strength of the linguistic area concerned, different features may actually bear different levels of significance. Typologically rare features are more areally significant because the clustering of such features within an area can effectively rule out the possibility of chance resemblance and highlight the typological distinctiveness thereof (cf. Hickey 2017). On the other hand, if interconnected features are counted as independent features, then the strength of the linguistic area in question may be overestimated. For example, as basic word order correlates significantly with other features related to constituent order (Greenberg 1963), in contact scenarios involving shifts in basic word order, it makes more sense to view the basic word order and its interrelated constructions as a single set

of traits (albeit a complex and significant one). See Campbell et al. (1988) for a case study of the Ethiopian linguistic area.

In our case, the pre-nominal relative clause, pre-verbal adjunct phrase, and adjective-final comparative are all typologically unusual features among VO languages (see Section 2.2). The prevalence of these rare traits in the Lingnan region firmly secures its status as a linguistic area. At the same time, although the Sinicization features identified are all related to head-final constructions, given that no basic word order change is observed in the Lingnan region, these features serve to assess the differing levels of Sinicization (Sinicization scores) of different languages in the region rather than that they exaggerate the degree of structural convergence, thereby circumscribing the core region of the linguistic area.

3.4.3 Number of language families

To ensure that the shared traits in a linguistic area are not inherited from a common ancestor, some scholars believe that it is necessary to have two or more language families involved in a linguistic area (e.g. Emeneau 1980; Aikhenvald and Dixon 2001). Nonetheless, given that the Balkan Sprachbund, one of the best recognized linguistic areas, primarily comprises languages from various branches of the Indo-European family (with minor contributions from Turkic), the multiple language-family requirement does not appear to be strictly followed or widely adopted in the field (see e.g. Lindstedt (2000)). Anyway, with no fewer than five distinct language families[12] involved in our area of interest, our case can surely stand up to the strictest scrutiny as far as this requirement is concerned.

3.4.4 Trait bundling

An idealized linguistic area may have an easily recognizable boundary, inside which languages consistently share certain traits absent in their sister languages outside the area. While this might be possible if centuries of intense contact had indeed occurred in a geographically isolated region, in reality, the isoglosses of different areal traits seldom overlap neatly (Emeneau 1980). This is hardly surprising given that a linguistic area is a dynamic and diverse phenomenon (Thomason 2001) often disrupted by "latecomers, earlier drop-outs, and temporary passers-by" (Stolz 2002: 265). This is not to deny the importance of looking for areas in which many traits bundle together, but we have to acknowledge the fact

12 Suggestions have been made for macrofamilies like Austric and Sino-Tai, arguing for some distant genetic relationship between these language families. These claims remain controversial and are beyond the scope of this study.

that the boundary line of a linguistic area, however we define it, is almost always fuzzy in nature. This is the reason why Campbell (2017) advocates against the use of overly geographically oriented criteria for establishing linguistic areas (see also Dahl 2001 for a similar view).

We mainly make use of Sinicization and Taicization scores to visualize the areal patterns in the present case study (see also Lindstedt (2000) for a similar use of scores to locate and distinguish the centre of the Balkan Sprachbund from the periphery). Given the inevitable presence of uneven convergence, this visualization method is easier to interpret than a bunch of discontinuous (and sometimes arbitrary), non-coinciding isoglosses on the one hand; on the other, it also rightly reflects the lack of a clear boundary of a linguistic area. In our case, while we can identify an areal hotbed in Guangxi and Southwestern Guangdong (see Figures 5 and 6), the area of convergence can be extended to Hainan and Central Western Guangdong, or even Southern Guizhou and Southeastern China at large. Naturally, the more we expand the linguistic area, the more exceptions and irregularities there will be. As the essence of our approach lies at mutual convergence (see Section 3.5), we will occupy the 'middle ground' and locate the linguistic area in Guangxi, Hainan, and Central Western Guangdong (hereafter Western Lingnan). As we will discuss below, the Western Lingnan Sprachbund emerged as a result of mutual convergence between the Sinitic and Tai-Kadai languages spoken therein, which was in turn caused by bi/multilingualism (which may or may not be reciprocal).

3.5 On the mutualist approach

According to Campbell (1985), there are two main approaches to establishing a linguistic area. In the 'circumstantialist' approach, a linguist simply compiles a list of structural similarities found in the languages belonging to a geographical area, assuming that they emerged as a result of contact (instead of chance or inheritance). In the 'historicist' approach, a linguist has to provide concrete evidence to demonstrate that the shared traits are indeed products of language contact, as well as indicate the direction of diffusion, during which historical information about the languages and populations is usually required. Although the latter approach is far preferable, in many cases linguists have no access to such historical information and have to resort to the former approach (see also a similar problem for Vietic in Alves (2020)).

With a reasonably rich body of literature about the history of the languages and peoples in the Lingnan region, we fortunately are able to adopt the historicist approach in this study. Moreover, to establish the Western Lingnan linguistic area,

we have developed the mutualist approach, which can be considered a subtype of the historicist approach. As we discuss below, this approach can arguably help establish linguistic areas securely.

In the mutualist approach, the first (and most crucial) step is to identify at least two genealogical units of languages (belonging to different language families or distinct branches within a family) demonstrating signs of mutual structural convergence within a particular region. These are the core members for establishing the linguistic area. Next, a contact scenario is assumed, involving two linguistic groups, namely A and B. If a given trait is common in A inside the area and rare in A outside the area, and if it is common in B in general, we can (tentatively) list it as a trait appearing in A inside the area due to contact influence from B. Likewise, if a given trait is rare in A inside the area and common in A outside the area, and if rare in B in general, we can (tentatively) list it as a trait which faded out in A inside the area due to contact influence from B (cf. the concept of 'active adaptation' and 'passive adaptation' put forward by Janhunen (2007)). In rare cases where the earlier stages of a particular linguistic group are relatively well-studied (in our case, Sinitic), this historical information should also be taken into account whenever appropriate. In our case, Sinitic and Tai-Kadai represent the core members which have demonstrably undergone mutual convergence (see Section 2). To aid our analysis, we may also look, after considering these primary traits, for supplementary traits (possibly of uncertain origin) that seem peculiar to the area (see Section 3.2).

There are often members in a linguistic area that cannot serve to illustrate mutual convergence due to various constraints (e.g. migration history for Hmong-Mien, small number of representatives for Austroasiatic and Austronesian). However, as in our case, they can still provide corroborating evidence for the existence of a linguistic area by showing signs of convergence towards the core members (Section 3.3).[13]

As languages may differ substantially in power and size, contact influence is not always mutual. In the mutualist approach, however, we emphasize that influence must be mutual in order to establish a secure linguistic area that must have arisen by structural convergence. In a contact setting with a clearly dominant language, speakers of other languages may acquire the dominant language through bi/multilingualism or even language shift, introducing many non-native features to the 'feature pool' of the dominant language (Mufwene 2001). Illustrative examples have been reported from Southern Burma and the Reef Islands, where populations of dominant languages are outnumbered by non-native speakers, causing structural changes in the dominant languages (Næss and Jenny 2011).

13 This is analogous to the status observed for Ladino, Romani, and Turkish as non-core but still significant members of the Balkan Sprachbund in terms of overall language sociology and ecology (see e.g. Lindstedt 2000).

Therefore, despite the overwhelming dominance of the language, as long as there is direct contact between its speakers and those of other languages, non-native features will inevitably infiltrate into the dominant language (unless there is a huge imbalance in population size, which may render the effect of the non-dominant languages negligible). Therefore, in Mufwene's (2001) terminology, the dominant language can only win a 'Pyrrhic victory'.[14] In cases where the dominant language remains unaffected, there is probably very limited contact between its speakers and those of other languages, suggesting a loose communication network. All kinds of linguistic innovations and changes observable at the communal level must logically stem from the idiolectal level (Mufwene 2001; Szeto et al. 2019); a loose communication network between speakers of different languages implies a low level of contact intensity, which can hardly result in a proper linguistic area.

We view our approach as a relatively conservative one which requires solid evidence to establish a linguistic area securely. As Muysken (2008) shows, the level of evidence required is inversely proportional to the size of the linguistic area postulated. Obviously, the difficulty of adopting a historicist approach increases with the size of the area because we are bound to encounter an increasing number of information gaps. In addition, under the mutualist approach, we often have to compare the typological profile of a given linguistic group spoken inside and outside the postulated linguistic area. This imposes a limit on the size of the linguistic area because many linguistic groups (especially branches within a family) in the world do not have extensive geographical coverage, such as is the case with Koreanic, a small and relatively homogenous language family, forming a linguistic area with Japonic (Yurayong and Szeto 2020). Consequently, the geographical extent of the Western Lingnan linguistic area identified in this study may be smaller than that of some reasonably well-established linguistic areas. A case in point is the Mainland Southeast Asian (MSEA) linguistic area (Enfield 2005, 2019), which may cover the entirety of the Western Lingnan linguistic area. With its large size, the MSEA linguistic area virtually encompasses all the Tai-Kadai, Hmong-Mien, and Austroasiatic languages, making it extremely difficult (if not impossible) to evaluate the convergence phenomena of these linguistic groups under the mutualist approach.

Despite the presence of a number of common features, given that some areal traits (especially those related to Sinicization) are restricted to the Western Lingnan linguistic area, as well as the ample evidence for mutual convergence in the region, we take the view that Western Lingnan should be regarded as a linguistic

14 This may not be the case if we talk about standardized languages. The focus of areal linguistics, however, is typically on regional vernaculars.

area in its own right instead of a micro-area under the MSEA linguistic area. Admittedly, this is essentially a terminological issue. Adopting Campbell's (2017: 28) terminology, Western Lingnan can be referred to as a "linguistic area *sensu stricto*" and MSEA a "trait-sprawl area". Possibly, we may even further delimit the geographical range of the Western Lingnan linguistic area by identifying more areal traits. The grammaticalization patterns recognized by Huang and Wu (2018) are promising candidates,[15] but at this stage we do not have enough empirical data to include these features in our analysis.

3.6 Summary and discussion

By now it should be obvious that Western Lingnan meets all the traditional criteria for a Sprachbund, in which we are able to identify an array of structural traits (whose origins are largely traceable) shared among languages of different genealogical affiliations, effectively ruling out the possibility of chance resemblance. Given the difficulty (or even futility) of establishing a fixed set of criteria for defining a Sprachbund to be followed universally, it is crucial to stress that the mutualist approach proposed in this study is not meant to be a set of prescriptive methods for defining Sprachbünde in general; instead, our aim is to identify a specific kind of linguistic area (which can be referred to as mutualist Sprachbund) whose convergence phenomena arose as a result of long-term intense contact, bi/multilingualism, and possibly genetic admixture as well. As we mentioned in Section 3.4, the specific kinds of contact-induced change occurring within a linguistic area deserve more attention than a precise definition of a linguistic area. In the present case, apart from the prevalence of bidirectional influence, we can see that even highly unusual syntactic features can diffuse through a mutualist Sprachbund. While one may argue that this is facilitated by the relatively close typological distance between Sinitic and Tai-Kadai languages in general, it is noteworthy that similar convergence phenomena are also observed in Tsat, making it clearly typologically distinct from its Chamic (and Austronesian at large) sisters in the MSEA region (Thurgood 2010). This suggests that the contact settings of a mutualist Sprachbund may likely result in

15 Wu (2008a) suggests that the A-not-A question (e.g. Cantonese *nei5 gong2-m4-gong2* [2SG say-NEG-say] 'Are you going to say it?') is a characteristic construction of Sinitic that has diffused across minority languages in Southern China (including Tai-Kadai) but is not found in their sister languages outside China. This is not true. Although we cannot rule out the possibility that the A-not-A question originated from Sinitic, this construction is in fact present in Tai-Kadai languages outside China as well (e.g. Thai *kɛː˦ tɕaʔ˩ pʰuːt˥˩-mai˥˩-pʰuːt˥˩* [2SG IRR say-NEG-say] 'Are you going to say it?'). In Tai-Kadai languages outside China, the association of this construction with the informal register may have resulted in its omission from some reference grammars.

radical restructuring if languages of distinct typological profiles are involved. In addition to the Western Lingnan Sprachbund, the Amdo Sprachbund[16] provides another case in point, in which the Sinitic languages have undergone radical restructuring due to influence from their neighbouring Altaic and Tibetic languages (Dede 2007; Sandman 2016; Szeto et al. 2018; Szeto forthcoming; Xu 2021; Zhou 2019).

4 Conclusions

Through devising and adopting the mutualist approach, this study demonstrates that Western Lingnan qualifies as a strongly supported linguistic area according to criteria widely recognized by areal linguists, suggesting that the grammatical phenomena of languages within this region are best studied from the perspective of language contact. As its name implies, concrete evidence for mutual convergence lies at the heart of this approach. We look forward to applying the mutualist approach to the study of other putative linguistic areas in the world, thereby fine-tuning its methodology and at the same time contributing to the discussion of the linguistic area concept. While we agree with the view that linguistic area is not a uniform phenomenon which can be defined or identified in a single way, we wish to put forward the mutualist approach in order to identify a specific kind of linguistic area known as mutualist Sprachbund. Should we manage to identify a good number of mutualist Sprachbünde around the globe, we will have a chance to look into the common patterns of contact-induced change under certain ecological settings, which can contribute to the study of contact linguistics at large.

Abbreviations

1/2/3	First/second/third person
CLF	classifier
COMP	comparative
COP	copula
IRR	irrealis
LOC	locative
NEG	negation
PFV	perfective
PL	plural

16 While it is beyond the scope of this study to evaluate whether the Amdo Sprachbund constitutes a mutualist Sprachbund, Janhunen (2007) provides a detailed discussion of the typological interaction between different language groups within the Sprachbund, which is largely compatible with the mutualist approach proposed in this study.

PRT	particle
REL	relative
Q	question particle/marker
SG	singular

Acknowledgements: Earlier versions of this paper were presented at the Workshop on Linguistic Minorities in Asia held at the Hong Kong Jockey Club University of Chicago Academic Complex in March 2019, the 29th Meeting of the Southeast Asian Linguistics Society (SEALS 29) held at the Tokyo University of Foreign Studies in May 2019, and the 24th International Conference on Yue Dialects held at the University of Macau in November 2020. We would like to thank the audience for their insightful feedback. We are also grateful for the very helpful comments and suggestions from the editors and anonymous reviewers of this journal.

References

Aikhenvald, Alexandra Y. & Robert M. W. Dixon. 2001. Introduction. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 1–26. Oxford: Oxford University Press.

Alves, Mark. 2020. Initial steps in reconstructing Proto-Vietic syntax. In Mathias Jenny, Sidwell Paul & Mark Alves (eds.), *Austroasiatic syntax in areal and diachronic perspective*, 46–81. Leiden: Brill.

Ansaldo, Umberto. 1999. *Surpass comparatives in Sinitic: Typology and grammaticalization*. Stockholm: Stockholm University Press.

Ansaldo, Umberto. 2010. Surpass comparatives in Sinitic and beyond. *Linguistics* 48. 919–950.

Barrere, Ian G. & Janhunen Juha. 2019. Mongolian vowel harmony in a Eurasian context. *International Journal of Eurasian Linguistics* 1(1). 46–77.

Bauer, Robert S. 1996. Identifying the Tai Substratum in Cantonese. In *Proceedings of the fourth international symposium on languages and linguistics*, 1806–1844. Salaya: Institute of Language and Culture for Rural Development, Mahidol University.

Bennett, Paul. 1979. A critique of the Altaicization hypothesis. *Cahiers de Linguistique Asie Orientale* 6. 91–104.

Bian, Chenglin, Meixin Huang & Miaoyan Lu (卞成林、黄美新、陆淼焱). 2011. Guixinan Yueyu he Zhuangyu de jiechuguanxi luelun (桂西南粤语和壮语的接触关系略论) [On contact relationship between Cantonese and Zhuang language in Southwestern Guangxi]. *Baisexueyuan Xuebao* (百色学院学报) [*Journal of Baise University*] 24(1). 66–69.

Bo, Wenze (薄文泽). 2002. *Mulaoyu yanjiu* (木佬语研究) [A study of Mulao]. Beijing: The Ethnic Publishing House.

Campbell, Lyle. 1985. Areal linguistics and its implications for historical linguistic theory. In Jacek Fisiak (ed.), *Proceedings of the sixth international conference of historical linguistics*, 25–56. Amsterdam: John Benjamins.

Campbell, Lyle. 2017. Why is it so hard to define a linguistic area? In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 19–39. Cambridge: Cambridge University Press.

Campbell, Lyle, Vit Bubenik & Leslie Saxon. 1988. Word order universals: Refinements and clarifications. *Canadian Journal of Linguistics* 33. 209–230.
Cao, Zhiyun (曹志耘) (ed.). 2008. *Hanyufangyan dituji* (汉语方言地图集) [Linguistic atlas of Chinese dialects]. Beijing: The Commercial Press.
Chappell, Hilary M. 2015. Linguistic areas in China for differential object marking, passive, and comparative constructions. In Hilary M. Chappell (ed.), *Diversity in Sinitic languages*, 13–52. Oxford: Oxford University Press.
Chappell, Hilary M., Li Ming & Alain Peyraube. 2007. Chinese linguistics and typology: The state of the art. *Linguistic Typology* 11. 187–211.
Chen, Jieming, Houfeng Zheng, Jin-Xin Bei, Liangdan Sun, Wei-hua Jia, Tao Li, Furen Zhang, Mark Seielstad, Yi-Xin Zeng, Xuejun Zhang & Jianjun Liu. 2009. Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *The American Journal of Human Genetics* 85. 775–785.
Chen, Liankai (陈连开) (ed.). 1999. *Zhongguominzushi gangyao* (中国民族史纲要) [A brief ethnic history of China]. Beijing: China Financial & Economic Publishing House.
Chen, Qiguang (陈其光). 2013. *Miaoyao yuwen* (苗瑶语文) [Miao and Yao language]. Beijing: China Minzu University Press.
Dahl, Östen. 2001. Principles of areal typology. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals: An international handbook*, vol. 2, 1456–1470. Berlin: De Gruyter Mouton.
Dede, Keith. 2007. The deep end of the feature pool: Syntactic hybridization in Chinese dialects. *Journal of Chinese Linguistics* 35(1). 58–80.
Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68(1). 81–138.
Dryer, Matthew S. 2003. Word order in Sino-Tibetan languages from a typological and geographical perspective. In Thurgood Graham & Randy J. LaPolla (eds.), *The Sino-Tibetan languages*, 43–55. London: Routledge.
Dryer, Matthew S. 2013. Relationship between the order of object and verb and the order of relative clause and noun. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
Dryer, Matthew S. (with Orin D. Gensler). 2013. Order of object, oblique, and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
Emeneau, Murray B. 1980. *Language and linguistic area: Essays by Murray B. Emeneau*. Stanford, CA: Stanford University Press.
Enfield, Nick J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34. 181–206.
Enfield, Nick J. 2019. *Mainland Southeast Asian languages: A concise typological introduction*. Cambridge: Cambridge University Press.
Gan, Rui-Jing, Shang-Ling Pan, Laura F. Mustavich, Zhen-Dong Qin, Xiao-Yun Cai, Ji Qian, Cheng-Wu Liu, Jun-Hua Peng, Shi-Lin Li, Jie-Shun Xu, Li Jin & Hui Li & The Genographic Consortium. 2008. Pinghua population as an exception of Han Chinese's coherent genetic structure. *Journal of Human Genetics* 53. 303–313.
Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of language*, 73–113. Cambridge, MA: MIT Press.

Güldemann, Tom. 2008. The Macro-Sudan belt: Towards identifying a linguistic area in northern Sub-Saharan Africa. In Bernd Heine & Derek Nurse (eds.), *A linguistic geography of Africa*, 151–185. Cambridge: Cambridge University Press.

Hashimoto, Mantaro. 1976. Language diffusion on the Asian continent: Problems of typological diversity in Sino-Tibetan. *Computational Analysis of Asian and African Languages* 3. 49–63.

Hashimoto, Mantaro (桥本万太郎). 1985. *Yuyandili leixingxue* (语言地理类型学) [Areal typology of language]. Beijing: Peking University Press.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

He, Guanglin, Zheng Wang, Jianxin Guo, Mengge Wang, Xing Zou, Renkuan Tang, Jing Liu, Han Zhang, Yingxiang Li, Rong Hu, Lan-Hai Wei, Gang Chen, Chuan-Chao Wang & Yiping Hou. 2020. Inferring the population history of Tai-Kadai-speaking people and southernmost Han Chinese on Hainan Island by genome-wide array genotyping. *European Journal of Human Genetics* 28. 1111–1123.

Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Hickey, Raymond. 2017. Areas, areal features and areality. In Raymond Hickey (ed.), *The Cambridge handbook of areal linguistics*, 1–16. Cambridge: Cambridge University Press.

Huang, Yang & Fuxiang Wu. 2018. Central Southern Guangxi as a grammaticalization area. In Sylvie Hancil, Tine Breban & José Vicente Lozano (eds.), *New trends in grammaticalization and language change*, 105–134. Amsterdam: John Benjamins.

Janhunen, Juha. 2007. Typological interaction in the Qinghai linguistic complex. *Studia Orientalia* 101. 85–102.

Kunyot, Taweesak. 1984. *General characteristics of Hmong Njua grammar*. Bangkok: Mahidol University MA thesis.

Li, Fang-Kuei. 1976. Sino-Tai. *Computational Analysis of Asian and African Languages* 3. 39–48.

Li, Fang-Kuei. 1977. *A handbook of comparative Tai*. Manoa: University Press of Hawaii.

Li, Jinfang (李錦芳). 1990. Yueyu zhong de Zhuangdongyuzu yuyandiceng chuxi (粵语中的壮侗语族语言底层初析) [A preliminary study on the Tai-Kadai substrate in Yue]. *Zhongyangminzudaxue Xuebao* (*Zhexueshehuikexueban*) (中央民族大学学报[哲学社会科学版]) [*Journal of the Minzu University of China* (*Philosophy and Social Sciences Edition*)] 6. 71–76.

Li, Jinfang (李锦芳). 1999. *Buyangyu yanjiu* (布央语研究) [A study of Buyang]. Beijing: China Minzu University Press.

Li, Rulong, Xiaoying Hou, Tiansong Lin & Kai Qin (李如龙、侯小英、林天送、秦恺). 2012. *Chadongyu yanjiu* (茶洞语研究) [A study of Chadong]. Beijing: The Ethnic Publishing House.

Li, Xulian (李旭练). 1999. *Laiyu yanjiu* (侏语研究) [A study of Bolyu]. Beijing: China Minzu University Press.

Li, Yunbing (李云兵). 2008. Yuyanjiechu dui yixie nanfang minzuyuyan yuxu de yingxiang (语言接触对一些南方民族语言语序的影响) [The influence of language contact on word order of some minority languages in Southern China]. *Minzu Yuwen* (民族语文) [*Minority Languages of China*] 5. 17–34.

Lindstedt, Jouko. 2000. Linguistic Balkanization: Contact-induced change by mutual reinforcement. *Studies in Slavic and General Linguistics* 28. 231–246.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian & Karen Emmorey. 1984. Is there a valid distinction between voiceless lateral approximants and fricatives? *Phonetica* 41. 181–190.

Mao, Zongwu & Yunbing Li (毛宗武、李云兵). 2007. *Younuoyu yanjiu* (优诺语研究) [A study of Younuo]. Beijing: The Ethnic Publishing House.
Matthews, Stephen. 2007. Cantonese grammar in areal perspective. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Grammars in contact: A cross-linguistic typology*, 220–236. Oxford: Oxford University Press.
Morey, Stephen. 2008. The Tai languages of Assam. In Anthony V. N. Diller, Jerold A. Edmondson & Yongxian Luo (eds.), *The Tai-Kadai languages*, 207–253. Abingdon: Routledge.
Mufwene, Salikoko S. 2001. *The ecology of language evolution*. Cambridge: Cambridge University Press.
Muysken, Pieter. 2008. Introduction: Conceptual and methodological issues in areal linguistics. In Pieter Muysken (ed.), *From linguistic areas to areal linguistics*, 1–24. Amsterdam: John Benjamins.
Næss, Åshild & Mathias Jenny. 2011. Who changes language? Bilingualism and structural change in Burma and the Reef Islands. *Journal of Language Contact* 4(2). 217–249.
Ouyang, Jueya (欧阳觉亚). 1998. *Cunyu yanjiu* (村语研究) [A study of Cun]. Shanghai: Shanghai Far East Publishers.
Ouyang, Jueya, Fang Cheng & Cuirong Yu (欧阳觉亚、程方、喻翠荣). 1984. *Jingyu jianzhi* (京語簡志) [A sketch grammar of Vietnamese]. Beijing: The Ethnic Publishing House.
Peyraube, Alain. 2015. Grammatical change in Sinitic languages and its relation to typology. In Hilary M. Chappell (ed.), *Diversity in Sinitic languages*, 53–78. Oxford: Oxford University Press.
Pittayaporn, Pittayawat. 2009. *The phonology of Proto-Tai*. Ithaca, NY: Cornell University PhD dissertation.
Pittayaporn, Pittayawat. 2014. Layer of Chinese loanwords in Proto-Southwestern Tai as evidence for the dating of the spread of Southwestern Tai. *Manusya: Journal of Humanities Special Issue* 20. 47–68.
Pulleyblank, Edwin G. 1991. *Lexicon of reconstructed pronunciation in early Middle Chinese, late Middle Chinese and early Mandarin*. Vancouver: UBC Press.
Pulleyblank, Edwin G. 1995. *Outline of classical Chinese grammar*. Vancouver: UBC Press.
Ratliff, Martha. 2010. *Hmong-Mien language history*. Canberra: Pacific Linguistics, Australian National University.
Robert, Ramsey S. 1987. *The languages of China*. Princeton: Princeton University Press.
Saeliao, Thanyalak (刘玉兰). 2012. *Taiguomianyu cankaoyufa* (泰国勉语参考语法) [A reference grammar of Thailand Mien]. Beijing: Minzu University of China PhD thesis.
Sandman, Erika. 2016. *A grammar of Wutun*. Helsinki: University of Helsinki PhD thesis.
Saul, Janice E. & Nancy F. Wilson. 1980. *Nung grammar*. Dallas, Texas: Summer Institute of Linguistics.
Shao, Huijun & Yu'en Gan (邵慧君、甘于恩). 2018. *Yueyucihui jiangyi* (粵語詞彙講義) [Lectures on Yue lexicon]. Hong Kong: The Commercial Press.
de Sousa, Hilário. 2015. The far Southern Sinitic languages as part of Mainland Southeast Asia. In Nick J. Enfield & Comrie Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art*, 356–440. Berlin: De Gruyter Mouton.
Stolz, Thomas. 2002. No sprachbund beyond this line! On the age-old discussion of how to define a linguistic area. In Paolo Ramat & Thomas Stolz (eds.), *Mediterranean languages: Papers from the MEDTYP workshop, Tirrenia, June 2000*, 259–281. Bochum: Brockmeyer.

Sun, Hao, Chi Zhou, Xiaoqin Huang, Shuyuan Liu, Keqin Lin, Yu Liang, Kai Huang, Jiayou Chu & Zhaoqing Yang. 2013. Correlation between the linguistic affinity and genetic diversity of Chinese ethnic groups. *Journal of Human Genetics* 58. 686–693.

Szeto, Pui Yiu. 2019. *Typological variation across Sinitic languages: Contact and convergence*. Hong Kong: The University of Hong Kong PhD thesis.

Szeto, Pui Yiu. forthcoming. Revisiting the Amdo Sprachbund: Genes, languages, and beyond. *Himalayan Linguistics*.

Szeto, Pui Yiu, Umberto Ansaldo & Stephen Matthews. 2018. Typological variation across Mandarin dialects: An areal perspective with a quantitative approach. *Linguistic Typology* 22(2). 233–275.

Szeto, Pui Yiu, Stephen Matthews & Virginia Yip. 2019. Bilingual children as "laboratories" for studying contact outcomes: Development of perfective aspect. *Linguistics* 57(3). 693–723.

Szeto, Pui Yiu & Chingduang Yurayong. 2021. Sinitic as a typological sandwich: Revisiting the notions of Altaicization and Taicization. *Linguistic Typology* 25(3). 551–599.

Thomason, Sarah G. 2001. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.

Thurgood, Graham. 2010. Hainan Cham, Anong, and Eastern Cham: Three languages, three social contexts, three patterns of change. *Journal of Language Contact* 3(2). 39–65.

Trubetzkoy, Nikolai S. 1923. Vavilonskaja bašnja i smešenie jazykov (Вавилонская башня и смешение языков) [The tower of Babel and the confusion of languages]. *Evrazijskij vremennik* (*Евразийский временник*) [*Annals of Eurasia*] 3. 107–124.

Wang, Fushi (王辅世). 1985. *Miaoyu jianzhi* (苗语简志) [A sketch grammar of Miao]. Beijing: The Ethnic Publishing House.

Wang, Fushi & Zongwu Mao (王辅世、毛宗武). 1995. *Miaoyaoyu guyingouni* (苗瑶语古音构拟) [A phonological reconstruction of Proto-Miao-Yao]. Beijing: China Social Sciences Press.

Wang, Jian. 2015. Bare classifier phrases in Sinitic languages: A typological perspective. In Hilary M. Chappell (ed.), *Diversity in Sinitic languages*, 110–133. Oxford: Oxford University Press.

Wang, Zhonghan (王钟翰) (ed.). 2004. *Zhongguominzushi gaiyao* (中国民族史概要) [A brief ethnic history of China]. Taiyuan: Shanxi Education Press.

Wen, Bo, Hui Li, Daru Lu, Xiufeng Song, Feng Zhang, Yungang He, Feng Li, Yang Gao, Xianyun Mao, Liang Zhang, Ji Qian, Jingze Tan, Jianzhong Jin, Wei Huang, Ranjan Deka, Bing Su, Ranajit Chakraborty & Li Jin. 2004. Genetic evidence supports demic diffusion of Han culture. *Nature* 431. 302–305.

Wu, Fuxiang (吴福祥). 2008a. Nanfangyuyan zhengfanwenju de laiyuan (南方语言正反问句的来源) [The origin of A-not-A questions in the minority languages in Southern China]. *Minzu Yuwen* (民族语文) [*Minority Languages of China*] 1. 3–18.

Wu, Fuxiang (吴福祥). 2008b. Nanfangminzuyuyan chusuojieciduanyu weizhi de yanbian he bianyi (南方民族语言处所介词短语位置的演变和变异) [Variation and change of the locative prepositional phrase in the minority languages in Southern China]. *Minzu Yuwen* (民族语文) [*Minority Languages of China*] 6. 3–18.

Wu, Fuxiang (吴福祥). 2009. Nanfangminzuyuyan guanxixiaojujiegoushiyuxu de yanbian he bianyi: Jiyu jiechuyuyanxue he yuyanleixingxue de fenxi (南方民族语言关系小句结构式语序的演变和变异——基于接触语言学和语言类型学的分析) [The changes and variations in the word order of relative clause and head noun in the minority languages in Southern China]. *Yuyan yanjiu* (语言研究) [*Studies in Language and Linguistics*] 29(3). 72–85.

Xu, Dan (徐丹). 2021. Zhuizong Ganqingyidai yuyanquyu nei Hanyu ji qi bianti gebiaoji xingcheng de guiji (追踪甘青一带语言区域内汉语及其变体格标记形成的轨迹) [Tracing the case forming in Sinitic languages and varieties in the Gansu-Qinghai linguistic area]. *Yuyan kexue* (语言科学) [*Linguistic Sciences*] 20(3). 272–286.
Xu, Liejiong & Alain Peyraube. 1997. On the double object construction and the oblique construction in Cantonese. *Studies in Language* 21(1). 105–127.
Yu, Cuirong (喻翠容). 1980. *Buyiyu jianzhi* (布依语简志) [A sketch grammar of Bouyei]. Beijing: The Ethnic Publishing House.
Yurayong, Chingduang & Pui Yiu Szeto. 2020. Altaicization and de-Altaicization of Japonic and Koreanic. *International Journal of Eurasian Linguistics* 2(1). 108–148.
Zhang, Junru (张均如). 1980. *Suiyu jianzhi* (水语简志) [A sketch grammar of Sui]. Beijing: The Ethnic Publishing House.
Zhang, Junru, Min Liang, Jueya Ouyang, Yiqing Zheng, Xulian Li & Jianyou Xie (张均如、梁敏、欧阳觉亚、郑贻青、李旭练、谢建猷). 1999. *Zhuangyufangyan yanjiu* (壮语方言研究) [A study of Zhuang dialects]. Chengdu: Sichuan Nationality Publishing House.
Zheng, Yiqing (郑贻青). 1997. *Huihuihua yanjiu* (回辉话研究) [A study of Tsat]. Shanghai: Shanghai Far East Publishers.
Zhengzhang, Shangfang. 1991. Decipherment of Yue-Ren-Ge (Song of the Yue boatman). *Cahiers de Linguistique Asie Orientale* 20(2). 159–168.
Zhou, Chenlei. 2019. A special case marking system in the Sinitic languages of Northwest China. *Journal of Chinese Linguistics* 47(2). 425–452.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/folia-2021-2003).