

Hilário de Sousa

The far southern Sinitic languages as part of Mainland Southeast Asia

1 Introduction

Within the Mainland Southeast Asian (MSEA) linguistic area (e.g. Matisoff 2003; Bisang 2006; Enfield 2005, 2011, Comrie 2007), some languages are said to be in the core of the language area, while others are said to be in the periphery. In the core are the Mon-Khmer and Kra-Dai languages. The core languages generally have:

- Analytic morphological profile with many sesquisyllabic or monosyllabic words
- Strong syntactic left-headedness, including prepositions and SVO word order
- Phonemic tonal contrasts and/or phonational contrasts

The Chamic languages (Austronesian) and the Hmong-Mien languages are also in the region, and are typologically relatively similar to the Mon-Khmer and Kra-Dai languages. On the other hand, there are the Sino-Tibetan languages in the northern and western periphery; their linguistic properties are somewhat less MSEA-like. For instance, in contrast to the strong syntactic left-headedness that is typical of MSEA languages, Burmese is OV and right-headed in general.¹ On the other hand, Mandarin has the left-headed traits of VO word order and preposition. However, Mandarin is otherwise strongly right-headed (e.g. right-headed noun phrases, adjunct-verb order). These two languages also have fewer lexical tones than most tonal languages in MSEA.

The aim of this paper is to discuss some of the phonological and word order typological traits amongst the Sinitic languages, and to compare them with the typological profiles of some MSEA languages. While none of the Sinitic languages could be considered to be in the core of the MSEA language

□□

¹ Nonetheless, Burmese still has some left-headed traits like post-nominal adjectives ('stative verbs') and numerals. In fact, it is more common for OV languages to have NAdj order (e.g. Dryer 2013). The North Asian type of consistent right-headedness for the OV and AdjN word orders is actually cross-linguistically slightly rarer.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

area, the Far Southern Sinitic languages, namely Yuè, Píngguà, and the Sinitic dialects in Hǎinán Island and Léizhōu Peninsula (largely corresponding to Chappell's (2012, 2013) 'Southern Zone') are typologically closer to the non-Sinitic MSEA languages to the south and west than the other Sinitic languages. Studies on the MSEA linguistic area would benefit from taking a closer look in a wider range of Sinitic languages, and include at least the Far Southern Sinitic languages as part of the MSEA linguistic area.

The rest of the paper is structured as follows. In Section 2, I present a brief overview of the Sinitic languages; I outline the history of the Sinitic languages, and also the genealogical relationships within and beyond the Sinitic language family. In Section 3, I discuss the typological features that are canonical of MSEA, and Comrie's (2007, 2008a) discussions on this based on the data from WALS. In Section 4, I discuss some of the MSEA-like phonological traits in the Sinitic languages. In Section 5, I discuss the variation in word order amongst the Sinitic languages. A conclusion is presented in Section 6.

2 The Sinitic languages

The Sinitic languages are the descendants of the historically attested Chinese language. The periodization of the Chinese language differs amongst linguists, with historical syntacticians usually favouring terms like 'Archaic Chinese' and 'Medieval Chinese', and historical phonologists usually favouring terms like 'Old Chinese' and 'Middle Chinese'.² The earliest attested stage of the Chinese language is 'Pre-Archaic Chinese', as exemplified by the fourteenth to eleventh century BCE oracle bone scripts (Shāng Dynasty). The earliest phonologically reconstructible form of Chinese is 'Old Chinese', which is reconstructed with the help of the Book of Odes/ *Shījīng*, the earliest collection of rhyming texts, composed between tenth to seventh century BCE (Western Zhōu and early Eastern Zhōu Dynasties). The diversity and time depth of the modern Sinitic language is comparable to that of the Romance languages (e.g. Norman 2003: 82). Around the same time that Vulgar Latin

□□

2 Historical syntacticians and phonologists of Archaic/Old Chinese deal with morphology in different ways. Historical phonologists of Old Chinese often reconstruct single-consonant affixes that are not necessarily indicated in the writing system, e.g. 王 **ɣʷaŋ* 'king', 王 **ɣʷaŋ*-s 'be king' (Baxter and Sagart 2014). On the other hand, historical syntacticians usually only look at the syntax and morphology of the strings of characters in texts.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

was spread by Roman conquests, spoken Chinese was spread by the expansions of the Qín (221 BCE – 206 BCE) and Hàn Empires (206 BCE – 220 CE) from the Yellow River Region. Based on lexical and phonological innovations, Sagart (2011) dates the most recent common ancestor of the modern Sinitic languages to about the third or second century BCE,³ with Xiānghuà 鄉話 (also known as Wǎxiāng(huà) 瓦鄉(話)) being the earliest branch. The Sinitic languages are often called ‘Chinese dialects’. The term ‘dialect’ is a (mis)translation of the Chinese term 方言 (Mandarin *fāngyán*), which literally means ‘regional speech’. The Chinese term *fāngyán* is semantically wider than the notion of ‘dialect’ in English, and readily includes what would be considered separate languages of the same language family in Western linguistics.

The Language Atlas of China (Zhāng et al. in press; Wurm & Li et al. 1987) classifies the Sinitic languages into ten major dialect groups, plus other unclassified smaller varieties, based primarily on phonological criteria. Each major dialect group includes a number of dialects that are not mutually intelligible. The ten major dialects groups are (Xióng and Zhāng 2008):

- Jìn 晉
- Mandarin 官話
- Wú 吳
- Hū 徽
- Gàn 贛
- Xiāng 湘
- Mǐn 閩
- Hakka (or Kèjiā) 客家
- Yuè 粵
- Píng huà 平話

The smaller Sinitic languages which fall outside of the ten-group classification are:

□□

3 More specifically, a time that is later than 330 BCE, the year that Alexander III of Macedon invaded Central Asia, and during or before the earlier stages of the Hàn Dynasty (202 BCE – 220 CE). See Sagart (2011) for details.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

- The ‘patois’ (*tǔhuà* 土話 in Chinese) of Southern Húnán (*Xiāngnán Tǔhuà* 湘南土話), Northern Guǎngdōng (*Yuèběi Tǔhuà* 粵北土話) and Eastern Guǎngxī (*Guidōng Tǔhuà* 桂東土話)⁴
- The Dānzhōu 儋州 language in Northeastern Hǎinán (somewhat Yuè-like, with influence from other Sinitic and non-Sinitic languages in lowland Hǎinán)
- The Xiānghuà 鄉話 (a.k.a. Wǎxiāng(huà) 瓦鄉(話)) language in western Húnán (e.g. Wǔ and Shén 2010, Chappell forthcoming)
- The Sinitic first language of Blue Dress Miáo people in Southwestern Húnán and neighbouring Northern Guǎngxī (Qīngyī Miáo Rénhuà 青衣苗人話; Lǐ 2004)
- The Sinitic first language of the Shē 畲 people (somewhat Hakka-like) (Yóu 2002)

Externally, the Sinitic language family is a member of the larger Sino-Tibetan language family. There are two groups of languages that are thought to be very close to the Sinitic languages in some ways. Firstly, there are the Bái 白 languages in Yúnnán. Some argue that Proto-Bái is a sister of Old Chinese (e.g. Starostin 1995; Zhèngzhāng 1999; Wāng 2006, 2012), while others argue that Bái is a family of Tibeto-Burman languages that has been heavily influenced by Chinese (e.g. Matisoff 2001b, Lee and Sagart 2008). Also in Southwestern China is the recently discovered Càijiā 蔡家 language (Bó 2004) on the Yúnnán–Guìzhōu border. Zhèngzhāng (2010) argues that Càijiā is a sister of Bái (and hence also genealogically related to Sinitic, according to his theory). Sagart (2011) considers Càijiā a sister of Xiānghuà (or at least the Sinitic layer in Càijiā is related to Xiānghuà if Càijiā turns out not to be a Sinitic language). Wǔ and Shěn (2010: 30–42) point out the lexical similarities between Xiānghuà, Old Chinese, Càijiā and Bái.

A number of factors contributed towards the distribution and diversity of the Sinitic languages. Firstly, there are the usual political and geographical factors which influence the distribution of languages in general. With the Sinitic family, the boundaries amongst the Sinitic languages follow the boun-

□□

4 In the first edition of the Language Atlas of China (Wǔrm & Lǐ et al. 1987), the Northern Guǎngdōng Patois are called Sháozhōu Patois. Nowadays, this term only refers to the patois in Mid-Northern Guǎngdōng near Sháoguān 韶關. The term ‘Eastern Guǎngxī Patois’ is not actually used in the Language Atlas of China; this term is increasingly popular in referring to the Patois in Eastern Guǎngxī in the Hèzhōu 賀州 area (e.g. Chén and Liú 2009). These patois are considered a type of Northern Píngguà in the Atlas. However, they are better viewed as a geographical continuation of the neighbouring Patois of Southern Húnán.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

daries of the historical prefectures or counties to some degree.⁵ For instance, although nowadays the language area of Huī 徽 Chinese is split amongst the three modern provinces of Ānhuī, Zhèjiāng, and Jiāngxī, it largely corresponds to the historical prefectures of Huīzhōu 徽州 (plus a small portion of neighbouring areas). Waterways facilitate the migration of people and linguistic features along them, and mountains between drainage basins impede the migration of people and diffusion of features across them. For instance, Xiāng 湘 Chinese is largely confined within the drainage basin of the Xiāng 湘 and Zī 資 rivers (both tributaries of the Yangtze). Unusual amongst the world's languages is the fact that the language diversity in Northern China, where the Chinese language originated, is low, whereas the language diversity in Southern China, where Chinese people migrated to, is high. This has to do with terrain: Southern China is mountainous, whereas in Northern China, there is the North China Plain, where one language, Mandarin, is spoken. In Northern China, there is also the Jin dialect area which is linguistically diverse; correlating with this fact is the unevenness of the terrain of this area, which is not part of the North China Plain. On top of the historical political boundaries and physical geography, there is also the complicated migration history of the Chinese people. For instance, in the case of Mandarin, Mandarin expanded outward from the North China Plain area rapidly within the last few centuries towards the northeast, northwest, and southwest. Towards Manchuria (northeast), the ban on Hàn Chinese people settling in Manchuria began to relax in 1860. Towards Dzungaria (northwest), Northern Xīnjiāng Mandarin formed in about 1780 (Liú 1993:4). Towards the Yúnnán-Guizhōu Plateau (southwest), Mandarin speakers arrived during the Míng Dynasty (1368–1644). Due to the relatively late outward expansion, Mandarin dialects cover a huge area, and the mutual intelligibility amongst the Mandarin dialects, even for the far-flung ones, is relatively high (in comparison with other Sinitic groups).

The Sinitic languages are also notable in that most of the speakers have been under unified single regimes for most of their history. Chinese people in general recognize the hegemony of the Common Chinese language, of which the latest stage is Standard Mandarin. Even when China was not unified, people from the various Chinese states used varieties of the same (written and spoken) Common Chinese language as a lingua franca. The concept of

□□

5 County is one level below prefecture, and prefecture is one level below province. Unlike India, China has an informal policy of not allowing provincial boundaries and linguistic boundaries to coincide.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

there being a Common Chinese language began as early as the Western Zhou dynasty (11th century BCE – 771 BCE). Common Chinese is based on the language of the contemporary or preceding political centre of China, which is usually in the North China Plain, neighbouring Wèi River Valley, or Lower Yangtze Region. One factor which contributed to the diversity amongst the modern Sinitic languages is the influence of prestigious varieties, with Common Chinese being overwhelmingly influential. The various Sinitic languages preserved linguistic material from different historical stages of Common Chinese. For instance, out of the major branches of Sinitic, only Mǐn retained a phonological layer from Old Chinese. Early Middle Chinese, the stage of Common Chinese represented by the language of the rime dictionary *Qièyùn* 切韻 (published in 601 CE during Suí Dynasty), has basically wiped out all phonological diversity amongst the Sinitic languages other than Mǐn. (However, the lexical and grammatical diversity amongst the Sinitic languages predates Early Middle Chinese.) The tree model is ill-fitted to the Sinitic family, as the Sinitic languages have preserved multiple layers of phonological material from Common Chinese (see Wáng 2009). Not only with phonology, the Sinitic languages have accumulated various layers of lexicon and grammar from various historical stages of Common Chinese ('stratification' in Chappell 2012). To complicate the matter even further, the non-standard Sinitic languages often create hybrid constructions from native material and material from Common Chinese ('hybridisation' in Chappell 2012). Other than influence from Common Chinese, there is also diffusion amongst the various non-standard Sinitic languages (e.g. the influence of Cantonese on Hakka and Mǐn in Guǎngdōng Province), making the classification of the Sinitic languages a notoriously difficult task.

The last major factor that contributes to the diversity of the Sinitic languages is the variation in areal influence from neighbouring non-Sinitic languages. This is where MSEA linguistics comes into Sinitic linguistics, the primary concern in this paper. Hashimoto (1978) and (1986) are the first major works that discuss Altaic influence on Northern Chinese, and Tai and Hmong-Mien influence on Southern Chinese. The historical interactions between Chinese people and their northern versus southern neighbours were drastically different. Northern China was dominated by various North Asian peoples, and sometimes Tibeto-Burman peoples, intermittently for more than one thousand years during the last two thousand years. The most influential dynasties were Mongolic (e.g. the Khitan Liáo Dynasty, 907–1125 CE) or Tungusic (e.g. the Jurchen Jīn Dynasty, 1115–1234 CE). There have been two dynasties where North Asians governed China as a whole rather than just Northern China: the Mongol Yuán Dynasty (1279–1368 CE) and the Manchu Qīng Dynasty (1644–20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

1912 CE). There were also dynasties headed by Turkic people (e.g. the various Shato Turk Dynasties during the Five Dynasty period, 907–979 CE), Qiangic people (e.g. the Tangut Xīxia Dynasty, 1038–1227 CE), and people of other Northern or Western ethnicities.⁶ Northern Chinese was influenced greatly by the North Asian languages, Mongolian and Manchu in particular, due to the North Asian languages being politically powerful, and also to the fact that many of the North Asian people shifted into speaking Chinese.⁷ For instance, under Altaic influence, in Mandarin and Jīn there are fewer tones, fewer classifiers, and many syntactic environments where clauses are verb final (Sinitic languages are normally verb-medial). In northwestern China, under the influence of neighbouring Turkic, Mongolic and Tibetan languages, there are even varieties of Mandarin with postpositional case markers and usually verb-final constituent order, for instance the Far-Western Central Plains Mandarin dialects in Línxià 臨夏 (a.k.a. Hézhōu 河州) and Xīníng 西寧 areas (e.g. Dede 2007), and the Tángwàng 唐汪 language (Djamouri forthcoming). The following is an example: the verb ‘eat’ is clause-final, and the object ‘meat’ is marked by an object case postposition 哈 *xa*.⁸

(1) Huàngshuǐ Mandarin (Xīníng area)

狗 肉 哈 吃 了
dog meat [OBJ] eat PRF

□□

6 During the Sixteen Kingdoms period (304–439 CE), there were various polities headed by the Dī 氐 people, whose descendents might be the modern Baima Tibetans (but see counter-arguments in Chirkova (2008)), who speak a Bodic language (e.g. Sun 2003). There were also the Jié (<*kiat) 羯 people, the leaders of the Later Zhào state (319–351 CE), who were probably Yeneseian (Pulleyblank 1963: 264; Vovin 2000). There were also kings of other ethnicities. King Gāo Yún 高雲 / Ko Un 高云 (reign 407–409) of Later Yān (384–409) or Northern Yān (407–436) was a descendent of the Goguryeo royal family (≈ Korean) adopted into the Yān royal family. The Táng Dynasty General Ān Lùshān 安祿山, who founded the short-lived Yān 燕 Kingdom (756–763 CE), had a father who was perhaps of Sogdian origin, and a mother who was a Turkic Zoroastrian priestess.

7 This is particularly the case with the Manchus: there are currently more than 10 million ethnic Manchus, but only a handful of native Manchu speakers left. The rest have shifted into speaking Mandarin or other Sinitic languages. Even when Xibe, an offshoot of Manchu, is included, there are fewer than 30,000 speakers.

8 Nearby there is also the mixed language Wūtún 五屯 (e.g. Janhunen, Peltomaa, Sandman and Dongzhou 2008) of which the vocabulary is over 50% Mandarin, and the grammar is mostly Tibetan. The phonology and lexicon in Wūtún is not as obviously Sinitic-like as Tángwàng. See Zhōng (2007) on the language contact situation in this area.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

‘The dog ate the meat.’
(Dede 2007: 867)⁹

The situation with Southern China was the opposite: Chinese people cause disturbance to the Southern non-Sinitic people more often than the opposite. Before the arrival of Han Chinese people, in Southern China there were Kra-Dai,¹⁰ Hmong-Mien, Austronesian and Austroasiatic-speaking people.¹¹ China first set up administrative bases in the Pearl River region and in the lower Red River regions during the Qín Dynasty (221–207 BCE). From then onwards, the primary migration routes for Chinese people have been from Northern China to Southern China. The migration of Chinese people to Southern China intensified whenever Northern China was ravaged by natural disaster or war (Chinese people had many wars with North Asians). The southward migration of Chinese people caused the southward migration of some of the Southern indigenous people deeper into Southeast Asia. Some of the indigenous populations of Southern China were assimilated by the migrant Chinese population. Genetically, it is known that the patrilineage of many Southern Chinese males is of Northern Chinese origin, while the matrilineage of most Southern Chinese people is of Southeast Asian origin (Wen et al. 2004). There is also a study which concluded that Northern Pínghuà speakers are genetically primarily Southeast Asian in both their patrilineage and matrilineage (Gan et al. 2008).¹² Linguistically, many Southern Sinitic languages

□□

9 Linguistic publications in the Chinese world often have examples with only Chinese characters and no phonological transcription of the characters. In this paper I try to include examples with phonological transcription as much as possible. With no phonological transcriptions, it is not always easy to determine whether a particular Chinese character is used for a morpheme because the morpheme: a) is a reflex of the same character in older stages of Chinese; b) is homophonous with that character but etymologically different; or c) is synonymous with the character, but etymologically and phonetically not related.

10 ‘Kra-Dai’ is a name propagated by, e.g. Ostapirat (2000, 2005); Pittayaporn (2009), for the language family which is also known as Tai-Kadai.

11 Ostapirat (2005) argues for the close relationship between Kra-Dai and Austronesian, and Sagart (2004) argues that Kra-Dai people were Austronesian who migrated from Taiwan back to the Mainland. That some conservative Kra languages have segments in their sesquisyllabic words matching the segments in the disyllabic words in Austronesian languages is a strong support for the link between Kra-Dai and Austronesian families. Many Kra groups have legends of their ancestors coming from the east and having crossed the sea in big boats (Lí 1999: 2). If Sagart’s viewpoint is correct, this ‘sea’ could well be the Taiwan Strait. If not, perhaps this ‘sea’ refers to a larger water crossing like the Mouth of the Pearl River.

12 Gan et al. (2008) make their claim for Pínghuà people in general. However, all but one of their sampling groups are Northern Pínghuà-speaking.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

are claimed to have Southeast Asian substrata. For instance, Cantonese has an obvious Tai substratum (e.g. Ōuyáng 1989, Bauer 1996). Nearly all Southern Sinitic languages have been argued to have at least some Kra-Dai vocabulary (see Lǐ 2002: 94–149). Hakka is often said to be a Gàn-like Sinitic language that was influenced by the Hmong-Mien language originally spoken by the Shē 畲 people (e.g. Sagart 2002).¹³ Mǐn is argued by Norman and Mei (1976) to have an Austroasiatic substratum (but this theory is criticized by Sagart (2008)). Historically, corresponding roughly to the modern day Wú-speaking area (around and south of the mouth of the Yangtze) was the Yuè 越 kingdom (? – 222 BCE), of which the commoners were probably Kra-Dai-speaking. There are bilingual Chinese-Yuè 越 texts like the sixth century BCE Song of the Yuè (*Yuèréngé* 越人歌; Wěi 1981, Zhengzhang 1991),¹⁴ and the Record of Yuè (*Yuèjuéshū* 越絕書; Zhengzhang 1998), which was compiled in the first century CE. Currently there are still islands of non-Sinitic languages in Southern China that have not (yet) been totally engulfed by the surrounding Sinitic languages. There are two such languages in Guǎngdōng: the Kam-Sui language of Biāo (Liáng 2002) which is surrounded by Yuè, and the Hmongic language of Ho Ne (Ratliff 1998), which is surrounded by Hakka. Given that many non-Sinitic MSEA people were absorbed into the Chinese community, it is not surprising that the Southern Sinitic languages bear similarities with languages in the core of MSEA.

In the rest of this paper, I will outline the typological features of the Sinitic languages in reference to the surrounding typological zones, and concentrate on the linguistic features in the Southern Sinitic languages that are typical of MSEA but atypical of Sinitic languages as a whole.

□□

13 Shē people these days speak Sinitic dialects closely related to Hakka, with layers of Hmong-Mien and Kra-Dai vocabularies, and influences from their current Mǐn- and/or Wú-speaking neighbours (Yóu 2002). The Ho Ne people in Southern Guǎngdōng, who speak a Hmongic language (Ratliff 1998), are considered by the government to be the last remaining people who still speak the original Hmong-Mien language of the Shē people (Máo and Méng 1986). However, there are doubts that the Ho Ne people are actually Shē, based on the many cultural differences between Ho Ne and Shē Proper. Culturally, Ho Ne most closely resembles Yáo (≈ Mien) in Northern Guǎngdōng, and Ho Ne people do in fact consider themselves Yáo (according to Yóu 2002: 8–10).

14 There are competing theories in Vietnam that the language in *Yuèréngé* (*Việt Nhân Ca* in Vietnamese) is Vietic.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

3 The typology of the MSEA linguistic area and the Sinitic languages

The MSEA linguistic area is commonly understood to include the following groups of languages (e.g. Matisoff 2003; Bisang 2006; Enfield 2005, 2011; Comrie 2007):

- “Mon-Khmer” languages (i.e. Austroasiatic family minus the Munda branch)
- Kra-Dai languages
- Hmong-Mien languages
- Chamic languages
- Some of the surrounding Sino-Tibetan languages, e.g. Karen, Lolo-Burmese, some nearby Sinitic languages

Towards the north, the Sino-Tibetan languages and the strongly Chinese-influenced varieties of Kra-Dai and Hmong-Mien languages can be said to be on the periphery of the MSEA linguistic area.

We will start by discussing Comrie (2007, 2008a), which present a measurable framework in comparing the typological profiles of languages (albeit with pitfalls, as Comrie admits). Most studies on language areas begin by having preconceptions about what linguistic features are common in a linguistic area, and then the geographical extent of the said features are determined. Comrie (2007) takes a different approach. Instead of having a preconceived list of typological features, all the linguistic features in the World Atlas of Language Structures (WALS; Haspelmath, Dryer, Gil, and Comrie 2005) are examined to see whether there are typological features that distinguish MSEA from other areas. (See also Dahl (2008) on this approach.) The results of Comrie (2007) are largely congruent with the conclusions in other research on the MSEA linguistic area: there is a ‘core’ to the MSEA linguistic area with languages like Thai, Khmer, and Vietnamese which possess more canonical MSEA typological features, and a ‘periphery’, including languages like Indonesian, Burmese, and Mandarin which possess fewer canonical MSEA features. Comrie (2008a) follows similar methods, but concentrates on the Sinitic languages, comparing them with both MSEA and North Asia. There are twenty features that are said to be canonical of MSEA, and another set of twenty features that are said to be canonical of North Asia. Mandarin achieves

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

a score of 8 out of 20 for MSEA features (the lowest scored language out of the surveyed languages),¹⁵ and 11 out of 20 for North Asian features (the lowest scored language out of the surveyed languages, together with Nivkh). The conclusion is that Mandarin is typologically between MSEA and North Asia.¹⁶

The following are the twenty features that are said to be canonical in the MSEA linguistic area (Comrie 2008a):

- Having implosives
- Velar nasal used as onsets
- No front rounded vowels
- Complex tone systems
- Little affixation
- Having plural words
- No distributive numerals
- Obligatory use of numeral classifiers
- The perfect marker is synchronically a word meaning ‘finish’
- A number of left-headed traits
 - Verb – Object order
 - Preposition – NP order
 - Noun – Genitive order
 - Noun – Adjective order
 - Noun – Demonstrative order
 - Noun – Numeral order
 - Noun – Relative clause order
 - Adjective – Degree word order
- ‘Topic’ predicative possession construction (“possessor-TOPIC exist possessum)
- Verbal encoding for predicative adjectives

□□

15 Comrie (2007) has an extra MSEA feature that is not featured in Comrie (2008a): feature 45A ‘Politeness Distinctions in Pronouns’.

16 Instead of saying that Mandarin is ‘half-MSEA-like’ and ‘half-North-Asian-like’, one could also say that the MSEA and North Asian languages are typologically ‘half-Mandarin-like’. However, MSEA and North Asia serve as better typological standards of comparison due to their word order typological profiles being relatively normal: the MSEA languages are rather consistently left-headed, while the North Asian languages are very strongly right-headed. These contrast with the Sinitic languages, which have the very unusual typological profiles of being SVO, but otherwise strongly right-headed, as discussed in the rest of this paper.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

□ Different markings for nominal and locative predication

For this section, I have repeated the exercise using the twenty MSEA features in Comrie (2008a), with data from the 2011 online edition of WALS (Dryer and Haspelmath 2011), and added the following languages: Cantonese, Hakka, Eastern Kayah Li, Hmong, and Mien. Having more data from the Sinitic languages would be preferable (since this paper focuses on the Sinitic languages), but Cantonese and Hakka are the only non-Mandarin Sinitic languages with a reasonable amount of data in WALS. Eastern Kayah Li is chosen as a representative of the Karen languages. The Karen languages are interesting from a Sinitic point of view, as both the Sinitic and Karen families are SVO with mixed left-headed and right-headed typological profiles. Gaps in the WALS data are filled with the help of Matthews and Yip (2011) for Cantonese, Lo (1988) for Hakka, Solnit (1997) for Eastern Kayah Li, Wáng (1985) and Jarkey (1991) for Hmong,¹⁷ and Máo, Méng, and Zhèng (1982) for Mien. Based on the set of criteria used in Comrie (2007, 2008a), Cantonese, Hakka, and Mien (which score 9, 10, and 11 respectively) are comparable to Burmese (which scores 10) in terms of the distance between their typological profile and the MSEA typological canon. Eastern Kayah Li and Hmong score 14 and 13 respectively, which are closer to the score of 16 achieved by Khmer in the core of MSEA.¹⁸

□□

17 Data from various dialects of Hmong proper are used in this paper: Mong Njua (Green Hmong) data from WALS, Hmong Daw (White Hmong) data from Jarkey (1991), and Dànánshān Hmong data from Wáng (1985). Dànánshān Hmong is the standard variety of Western Hmongic (Chuānqíandiàn Miáo) chosen by the China government. These three dialects of Western Hmongic are very closely related to each other, and for the linguistic features discussed in this paper, the three dialects behave in the same way, unless specified. In the feature tables, 'Hmong' refers to Green Hmong and White Hmong, the varieties spoken by all Hmong speakers outside of China.

18 Amongst the Hmong-Mien languages, the Hmongic languages are generally less influenced by Chinese than the Mienic languages. The Hmongic languages are thus typologically more like the core MSEA languages than the Mienic languages (e.g. Ratliff 2010: 239–240).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

In the rest of this paper, I shall discuss further some of the phonological and word order issues discussed in Comrie (2007, 2008a), and some other related issues.

I shall also take this opportunity to introduce Chappell's (2012, 2013) classification of the Sinitic languages into four macro-areas (Chappell 2012: 5–6), with my own minor alterations, due to differences in linguistic criteria used.

- Northern zone:
Běijīng Mandarin, Northern (Jilǔ) Mandarin, Peninsular (Jiāoliáo) Mandarin, Northeastern Mandarin, Northwestern (Lányín) Mandarin, Central Plains (Zhōngyuán) Mandarin (portion), and Jin.
- Transitional zone:
Central Plains (Zhōngyuán) Mandarin (portion), Southeastern (Jiānghuái) Mandarin, Southwestern Mandarin, Xiāng, Xiānghuà (a.k.a. Wǎxiāng), Gàn, Mǐn-Gàn (i.e. Western Mǐn, which is strongly Gàn-influenced), and Hakka.
- Southeastern zone:
Wú, Huī, Mǐn.
- Far-Southern zone (≈ Chappell's "Southern Area"):
Yuè, Pínghuà, Sinitic languages in Léizhōu Peninsula and Hǎinán Island¹⁹.

□□
19 Some of the differences between the four typological zones in this paper and Chappell's (2012) four macro-areas are:

- the term 'Far-Southern zone' is used here instead of Chappell's 'Southern area'. The term 'Southern Sinitic' is ambiguous: it typically refers to the non-Mandarin Sinitic languages in Southern China, sometimes it also includes Southwestern Mandarin, and sometimes also Southeastern (Jiānghuái) Mandarin
 - Northern Wú and Huī are included here in the same Southeastern zone as Mǐn and Southern Wú. Northern Wú and Huī are more strongly influenced by Mandarin, and are hence sometimes treated differently from Southern Wú
 - the Mǐn exclaves in Léizhōu Peninsula and Hǎinán, which are spoken to the south of Yuè, are grouped together with Yuè in the Far-Southern zone. The Mǐn dialect of Hǎinán Island (a.k.a. Hainanese) is strongly influenced by the Kra-Dai language Ong-Be (i.e. the lowland indigenous language of Hǎinán), and the Mǐn dialect of Léizhōu
- 20140919 Draft of:* de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

In terms of word order, amongst the four zones, the languages with the most verb-medial traits are unsurprisingly in the Far-Southern zone; the Far-Southern Sinitic zone borders the Kra-Dai and Mien speaking areas, and many Kra-Dai and Mien speakers in China also speak Far-Southern Sinitic languages. As is to be expected, the languages in the Northern zone have a number of verb-final traits, being in contact with the North Asian languages. However, putting the aforementioned Far-Western Central Plains Mandarin dialects aside (which can be said to be actual SOV languages), the languages with the most verb-final traits are, surprisingly, in the Southeastern zone. This will be discussed in Section 5.

In the rest of this paper, unless specified, Sinitic data are provided by the seven members of the ERC Sinotype project,²⁰ based on their fieldnotes, their first-language knowledge, or their heritage-language knowledge. The following are the list of the team members and the data they contributed:

- Hilary Chappell: Gǔzhàng Xiānghuà (fieldnotes)
- Wěiróng Chén: Huì’ān Southern Mǐn (first language and field notes)
- Yùjjé Chén: Zhōukǒu Central Plains Mandarin (first lg and field notes)

- Xūpíng Lǐ: Yíchūn Gàn (fieldnotes)
Fùyáng Wú (first language)

- Peninsula is closely related to that of Hǎinán. Yuè and Píng huà have also been strongly influenced by Kra-Dai languages
- Western Mǐn is a Mǐn dialect that is strongly influenced by Gàn, and is here included in the same Transition zone as Gàn, rather than being in the Southeastern Zone together with other Mǐn dialects
- Chappell (2012) has the Hakka in Guǎngdōng in her ‘Southern area’, whereas other Hakka dialects to the north in the ‘Transitional area’

Chappell’s (2012) division of the Sinitic languages into four macro-areas is a refinement on Norman’s (1988: Section 8.1) division of the Sinitic languages into the typological zones of North (Mandarin), South (Yuè, Hakka, Mǐn), and Central (Xiāng, Gàn, and Wú). The four macro-areas in Chappell (2012) were based on the distribution of the various grammaticalization pathways of the passive and object marking constructions. However, it is noted (2012: 6) that the boundaries amongst the four macro-areas are approximate, and the boundaries would change slightly depending on the typological criteria used. The boundaries between the four typological zones proposed in this paper are also approximate, due to the paucity of data.

20 The Sinotype research project, funded by the European Research Council, was headed by Hilary Chappell, and hosted at École des hautes études en sciences sociales, from 2009 to 2013. See the acknowledgement section for more details.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

- Sing Sing Ngai: Shàowǔ Mǐn-Gàn, (a.k.a. Western Mǐn; fieldnotes)
Fúqīng Eastern Mǐn (heritage language)
Standard Cantonese (first language)
- Hilário de Sousa: Nánning Southern Pínghuà (fieldnotes)
Standard Cantonese (first language)
- Jiàn Wáng: Jǐxī Huī (fieldnotes)
Suīníng Central Mandarin (first language)

4 Phonology

In this section, I shall discuss the following phonological phenomena in the Sinitic languages and the MSEA languages to the south:

- Tones and onsets (Section 4.1)
- Codas (Section 4.2)
- Implosives (Section 4.3)
- Front rounded vowels (Section 4.4)
- ‘Apical’ vowels (Section 4.5)

We shall see that the Sinitic dialects in the Far-Southern zone and surrounding areas often have phonological traits that are typical of MSEA, but atypical for Sinitic languages. A summary of the phonological features is presented in Section 4.6.

Maps from the ‘Phonetics’ volume of the Linguistic Atlas of Chinese Dialects (LACD; Cáo et al. 2008) are shown. The maps are referred to by abbreviations like ‘Map P117’, where P stands for the Phonetics volume of LACD, and 117 for map 117 therein.

4.1 Tones and onsets

Most MSEA languages have phonemic use of pitch and/or phonational differences. Pitch and phonation are two closely related phenomena; both are primarily produced by configurations of the glottis. In this paper, ‘tone’ refers to systems where at least pitch contrasts have been phonemicized. Many of these phonemicized pitch systems also include phonational contrasts. (Languages where only phonational contrasts have been phonemicized are not

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

considered to be 'tonal' in this paper. See also Brunelle and Kirby (this volume).)

Many language families in this area with tones had an earlier stage where there were three tones for sonorant-ending syllables, and no tonal contrasts (or 'one' tone) for obstruent-ending syllables. This set of tonal contrasts is notated here as '3/1' tones. The proto languages with 3/1 tones include:

- Proto Kra-Dai
- Proto Hmong-Mien
- Middle Chinese
- Proto M̐n
- Proto Việt-Mường
- Proto Bái
- Proto Lolo-Burmese
- Proto Karen

The development of the three tones for sonorant-ending syllables is clear in some cases: one tone is related to an earlier **-h* (<**-s*), another to an earlier **-ʔ*, while the third is related to the lack of an obstruent at the end of a syllable. Haudricourt made this observation when comparing the tones in Vietnamese with the codas in cognates in other Mon-Khmer languages (Haudricourt 1954). The Sino-Tibetan languages have Written Tibetan as a reference. (Classical Tibetan was non-tonal; while many Tibetan varieties have developed tones, there are many Tibetan dialects in the periphery which remain non-tonal.) Written Burmese in fact still often marks the high tone with ̃, which is related to the Indic sign *visarga* : (-*h*), suggesting the high tone came from an earlier **-h*. There is also the case of Utsat, which, when compared with the other Chamic languages, developed tones in similar ways: normally a high tone developed out of **-h*, mid and low tones developed out of syllables with no obstruent ending, and rising and falling tones developed out of the plosive codas including **-ʔ* (Thurgood 1993).

Most languages in MSEA and East Asia have moved beyond this 3/1 tone system. The voicing of an onset influences the pitch value of a tone. Initially, the difference in pitch of a tone with different onsets might not be noticeable to speakers, but what typically happens is that the difference in pitch becomes more noticeably different. (The process of developing noticeable allotones are commonly referred to as 'tone-splitting'.) If the voicing contrast of the onsets is lost, the allotones become separate tonemes. Theoretically a language with 3/1 tones would thus end up with 6/2 tones. However, most languages do not have 6/2 tones, as the tones have gone through other

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

splits and mergers. For instance, while Northern Vietnamese has 6/2 tones, Southern Vietnamese has 5/2 tones as it has merged the *hỏi* and *ngã* tones. Standard Lao has 5/4 tones, and Central Thai has 5/3 tones; both have experienced different splits and mergers of the tones. Amongst the Sinitic languages, the Far-Southern languages, being closest to the core of MSEA, have the most tones on average. The Southeastern languages have slightly fewer tones,²¹ the Central languages have even fewer tones, and the Northern languages, being closest to the non- or less-tonal languages of North Asia, have the least tones on average. For example, prototypically (there are variations within each group):

- Yuè and Southern Píngguà dialects have 6/3 tones
- Mǐn and Wú dialects have 6/2 or 5/2 tones
- Hakka dialects have 4/2 tones
- Xiāng dialects have 5/1 tones
- Southeastern Mandarin dialects have 5/1 or 4/1 tones
- Jìn dialects have 4/1 tones
- Other Mandarin dialects have 4/0 or 3/0 tones
(Many Mandarin dialects have '0' tones as they have lost all plosive codas)

LACD Map P001 (Figure 1) shows the number of 'tone categories' amongst the Sinitic languages. ('Tone categories' in traditional Chinese linguistics refers to all the allotones in a language counted separately, including the tones for sonorant-ending syllables and tones for obstruent-ending syllables. For instance, Standard Cantonese has '9 tone categories' according to traditional Chinese linguistics; in my notation, Cantonese has '6/3' tones, i.e. 6 tonemes.) The Sinitic dialects with the highest number of tone categories are clearly concentrated in Far-Southern China, the area closest to the core of MSEA. Mandarin has the smallest number of tone categories, especially Northwestern Mandarin.

One prominent non-Sinitic historical tonal trait in Yuè and many Southern Píngguà dialects is the split of tone D (the tone for obstruent-ending syllables) based on vowel length. This is a hallmark of Kra-Dai languages.²² The only

□□

21 In terms of tonal behaviour, one major difference between the Far Southern and the Southeastern zone is that languages in the Far Southern zone tend to be poor in tone sandhi, whereas languages in the Southeastern zone tend to have complex tone sandhi.

22 Unlike other Southern Píngguà dialects, Southern Píngguà dialects in Nánning and areas to the west split the tone D not by vowel length, but by the sonority of the initial consonant in 20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

other non-Kra-Dai language that I know of with this trait is Kim Mun (Mienic) in Hǎinán (Lǐ 2003: 694–697). This trait in Kim Mun is perhaps due to influence from Hlai, the dominant Kra-Dai language in central Hǎinán.

□□

Middle Chinese, e.g. Nán níng Wèi zǐ lù Píng huà /wət²³/ 域 'region' (< *wik), /wət²/ 活 'live' (< *ywat). See de Sousa (forthcoming).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

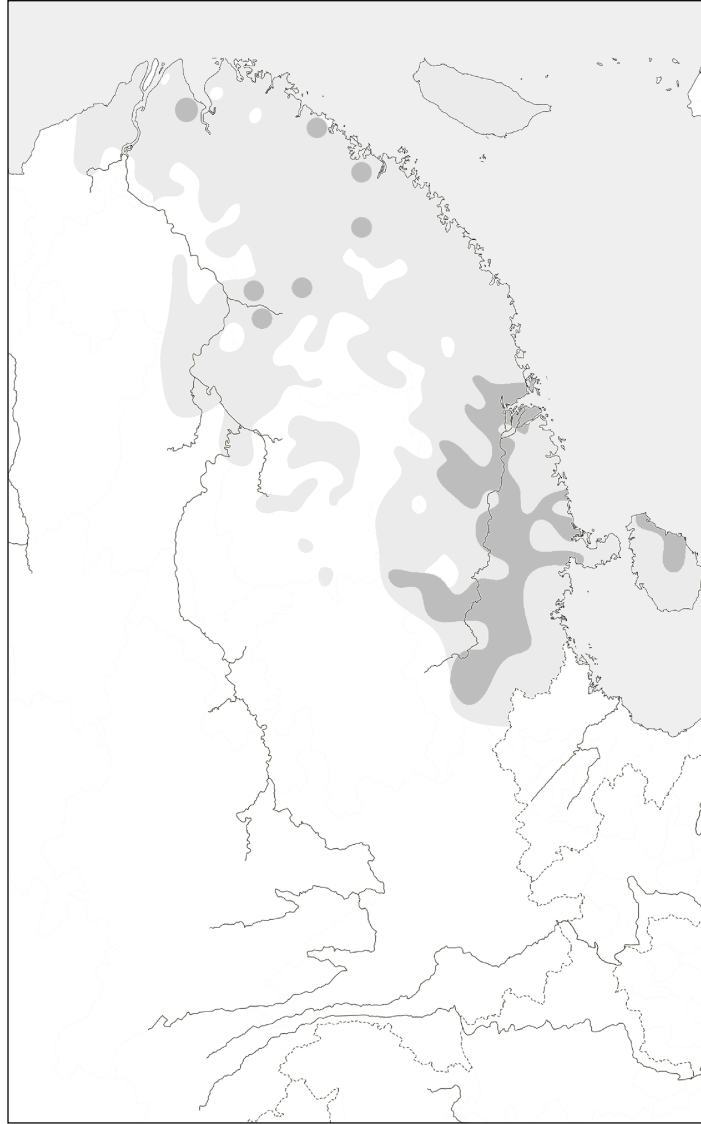


Fig. 1: Areas with nine or more allotones are highlighted in dark gray, areas with six to eight allotones in light gray. Other Sinitic languages have two to five allotones. (Data derived from LACD Map P001).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia : The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

The loss of the voicing contrast (for plosive onsets) has also occurred in many Mon-Khmer languages, which are mostly non-tonal. For instance, in Mon, the old voicing contrast of the onsets is now expressed as a phonational contrast of modal versus breathy. The phonational contrast caused a change in the vowel qualities (e.g. Ferlus 1980, 1984). In Khmer, not only has the onset voicing contrast been lost, but the phonational contrast has also been lost in most dialects. This phonemicization has led to the vowel quality contrasts (e.g. Wayland and Jongman 2002).²³

It is interesting to note that there are languages in MSEA where (onset-related) tone-splitting has not happened, i.e. still at the stage of having 3/1 tones:

- Burmish languages like Burmese, Achang and Xiāndǎo
- Nusu (Loloish)²⁴
- A-Hmyo dialects (Luóbóhé 羅泊河 Miáo, at, e.g. Fúquán 福泉; Western Hmongic; but tone D has partially or totally merged with tone A) (Ratliff 2010: 185; Li 2003: 686–688)

There are also languages where tone-splitting has occurred, but the allo-tones have not been phonemicized, as the original contrast between modal voice and modal voiceless onsets is still intact (i.e. the difference in pitch is still predictable by the phonemic voicing contrast of the onsets).²⁵ These languages include:

□□

23 There are also some tonal languages in China which have split vowel qualities based on tones, presumably through an intermediary stage with phonational difference which has since been lost:

- Mang (Máshān 麻山 Miáo, Western Hmongic; tones B2 and C2 versus others) (Wáng 1985: 107; Ratliff 2010: 196);
- The southern half of the Eastern Mǐn dialects, e.g. Fúzhōu, Fúqīng (tones C2, C1 and D1 versus others; D1 has lower pitch than D2).

The commonality is that tone C developed out of *-h*, which ‘encourages’ breathy phonation, and tone 2, which correlates with voiced onset and lower pitch, which also ‘encourages’ breathy phonation.

24 Other Burmish languages have shown signs of tone-splitting: Zaiwa/Atsi, Maru/Langsu and Lashi. As for Loloish languages, most have departed from the ancestral 3+1 tone system (e.g. Li 2010: 56).

25 The original modal voicing for the onsets may have changed into something like breathy voice, but these onsets are still distinct from the modal voiceless onsets. Dialects of Wú, e.g. Shanghainese, are mostly like this.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

- Wú dialects, including some neighbouring Wú-influenced Mǐn varieties as in:
 - Eastern Mǐn in Cāngnán, Zhèjiāng
 - Southern Mǐn in Guǎngfēng, Jiāngxī
 - Northern Mǐn in Pūchēng, Fújiàn (Zhèngzhāng 1995)²⁶
- Southern Xiāng dialects ('Old Xiāng')
- Xiānghuà and some nearby Mandarin dialects in western Húnán
- A few Northern Gàn dialects, e.g. Wǔníng (Zhū et al. 2009)
- A few Northern Yuè dialects, e.g. Liánshān, Yángshān (Zhèngzhāng 1995)
- A-Hmao dialects ("Northeastern Yunnan" Miao, at, e.g. Shíménkǎn 石門坎 in Wēiníng 威寧; Western Hmongic; it has also developed noun versus non-noun contrasts with tones B2 and C2/D2) (Ratliff 2010: 185; Lǐ 2003: 708)

The phonemicizing of suprasegmental features based on the loss of the original contrast between modal voiced and modal voiceless onsets is the norm in MSEA. This is summarized in Table 2 in Section 4.6.

4.2 Consonantal codas

Many proto languages in East and MSEA are reconstructed with at least six consonantal (i.e. non-glide) codas. For example:

- Pre-Angkorian Khmer (Jacob 1993): *-p -t -c -k -m -n -ŋ -r -l -v -s -h*
- Proto Hmong–Mien (Ratliff 2010): *-p -t -k -m -n -ŋ*
- Proto Tai (Pittayaporn 2009): *-p -t -c -k -m -n (-ŋ) -ŋ -l*
- Middle Chinese (Baxter 1992): *-p -t -k -^wk -m -n -ŋ -^wŋ*

In some languages there is a dramatic loss of coda distinctions. For instance, while Mien has preserved *-p -t -k -m -n -ŋ* (Máo, Méng and Zhèng 1982: 16), Hmong has lost all the plosive codas, and all nasal codas have collapsed into an *-ŋ* or vowel nasalization (Wáng 1985: 18). Most Kra-Dai,

□□

26 One important feature that distinguishes Wú and Huī, which are otherwise very similar to each other, is that Huī dialects have phonemicized the splitting of tones. The inventory of onsets in Huī is similar to Gàn to the west; amongst many similarities, they have both lost the voicing distinction of the Middle Chinese onsets.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

Mienic and Mon-Khmer languages have at least three plosive codas and three nasal codas. In table 2 (Section 4.6), the sampled East and MSEA languages are classified based on two criteria: a) having more than one contrastive plosive coda; and b) having more than one contrastive nasal coda.²⁷ It is the norm in MSEA to have at least two plosive codas and two nasal codas (usually there are at least three each). With the Sinitic languages, LACD Map P121 (Figure 2) shows the distribution of *-m -n -ŋ*, and LACD Map P124 (Figure 3) shows the distribution of *-p -t -k -ʔ -l* in the Sinitic languages.

Having two or more plosive codas is largely confined to the following Sinitic languages in or near Far-Southern China, which is closest to the core of MSEA:

- Southern Mǐn (including Mǐn in Hǎinán Island and Léizhōu Peninsula)
- Yuè
- Southern Pínghuà
- Hakka in Guǎngdōng
- Some Gàn dialects

The same sets of Sinitic languages satisfy the criterion of having *-m* and one other nasal coda. The norm for Sinitic languages is to have just *-n* and/or *-ŋ*, and to have just *-ʔ* or no plosive codas at all. The Sinitic languages in or near the Far-Southern zone are more similar to the core of MSEA with respect to the conservativeness of codas. The number of contrastive codas is summarized in Table 2 in Section 4.6.

4.3 Implosives

Many MSEA languages have the implosive consonants *b̥* and *d̥* (but not *g̥*).²⁸ Examples of languages with *b̥* and *d̥* include Khmer, Vietnamese and Sgaw Karen. Some MSEA languages, e.g. Eastern Kayah Li (Solnit 1997), are said to have non-implosive *b* and *d* (but no *g*, analogous to nearby languages which have *b̥* and *d̥* but no *g̥*). As for the Sinitic languages, neither Middle

□□

²⁷ The syllabic nasals that exist in many Southern Sinitic languages are not included in the criterion of having more than one nasal coda.

²⁸ Implosive *g̥* is cross-linguistically rare. For the velar *g̥*, the voicing that is common for implosives is more difficult to maintain because the distance between the glottis and the oral closure at the velar position is short. Similarly, the plumonic *g̥* is also cross-linguistically rarer than *b̥* and *d̥* (Maddieson 2013a).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

Chinese nor Old Chinese were reconstructed with implosive consonants. However, some modern Sinitic languages have implosives. LACD Map P044 (Figure 4) shows the distribution of implosive onsets in the Sinitic languages.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia : The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.



Fig. 2: Areas with *-m -n -ŋ* highlighted in light gray, and areas with *-m -ŋ* or *-n* in dark gray. Other Sinitic languages have *-n -ŋ*, just one of them, or in rare cases, no nasal codas. (Data derived from LACD Map P121).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

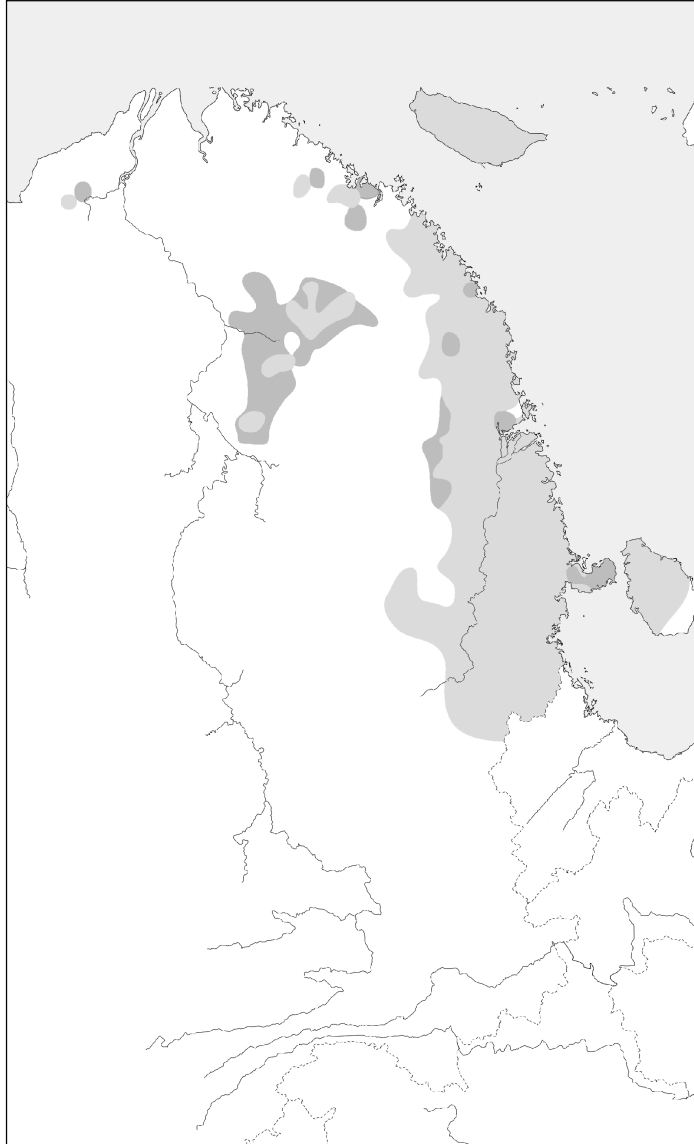


Fig. 3: Areas with three or four non-nasal codas are highlighted in light gray, areas with two non-nasal codas in dark gray. Other Sinitic languages have no non-nasal codas, a single -ʔ, or in rare cases a single -t or -l. (Data derived from LACD Map P124)

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia : The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

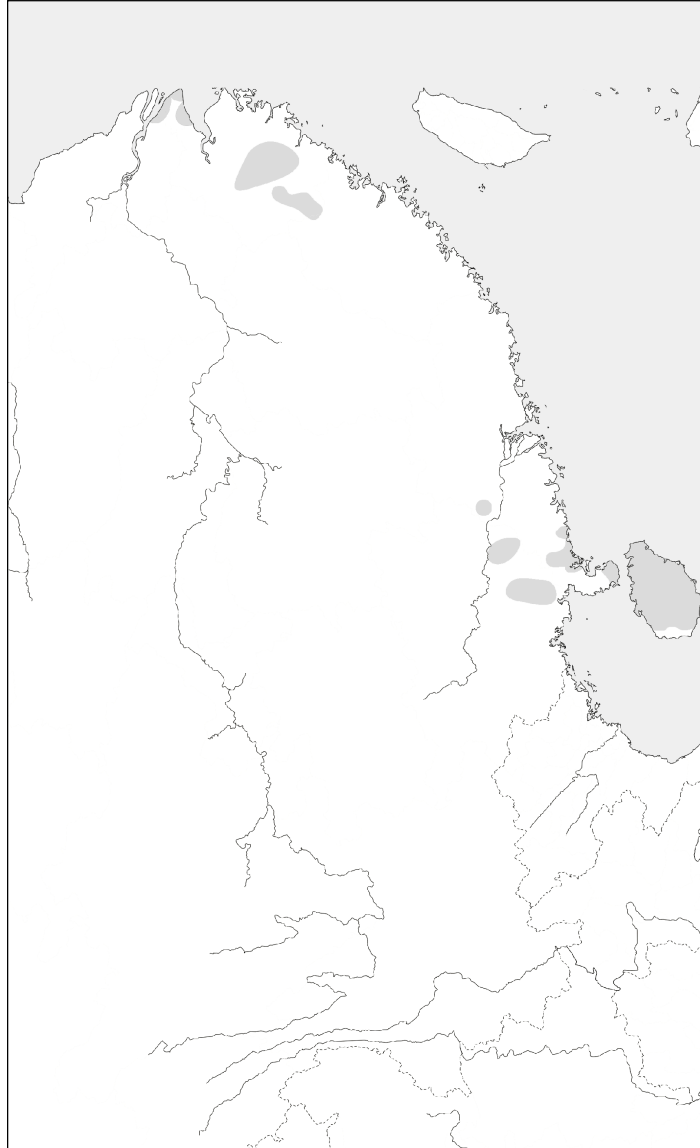


Fig. 4: Areas with implosive consonants are highlighted. Other Sinitic languages have no implosives. (Data derived from LACD Map P044).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

According to this map, implosives are found in:

- Mǐn in Hǎinán Island and Léizhōu Peninsula
- Dǎnzhōu dialect (the Yuè-like language in Northeastern Hǎinán)
- Some of the Gōulòu Yuè dialects near the Guǎngxī–Guǎngdōng border
- Some Southern Wú dialects
- Some Northern Wú dialects around Shànghǎi

The most famous example is Hainanese (i.e. Hǎinán Mǐn), which was very strongly influenced by Ong Be, the lowland Kra-Dai language in northern Hǎinán, which also has *b* and *d*. Across the Hǎinán Strait, some of the Gōulòu Yuè dialects also have implosive *b* and *d*. (However, in some localities they are becoming *p* and *t* respectively.) Further away to the northeast, there are implosives in some of the Wú dialects.

Their origins differ. In Hǎinán and Léizhōu Mǐn, *b* and *d* developed out of **p* and **t* after **b* and **d* lost their voicing and merged into **p* and **t*, whereas in the other Sinitic languages (including the Yuè-like Dǎnzhōu dialect in Hǎinán) *b* and *d* developed out of **p* and **t* when **b* and **d* were still distinct from **p* and **t*. In areas surrounding Shanghai, a new *g* has developed out of voiced **g* (unlike *b* and *d* which developed out of voiceless **p* and **t*).

See Zhū (e.g. 2006b, et al. 2009) on implosives in Sinitic languages, including some newly developed implosives in Northern Gàn dialects and Cháoshàn Mǐn dialects (e.g. Shàntóu/ Swatow). The existence or non-existence of *b d* ~ *b d* is summarized in Table 2 in Section 4.6.

4.4 Front rounded vowels

The vast majority of the world's languages lack front rounded vowels (Maddison 2013b). Most MSEA languages also lack front rounded vowels. Many Sinitic languages in Southern China also lack front rounded vowels. Otherwise, the norm for Sinitic languages is to have front rounded vowels. The following are the main examples of Sinitic languages in Southern China without front rounded vowels:

- Southern Mǐn, including Mǐn of Hǎinán Island and Léizhōu Peninsula
- most Hakka dialects and some neighbouring Southern Gàn dialects
- most Yuè dialects not in the drainage basin of the Pearl River (which entails being somewhat less influenced by Standard Cantonese)

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

- Most Southern Píngguà dialects
- Some Southern Mandarin dialects, especially in Yúnnán and Guìzhōu

Map P117 in LACD (Figure 5) shows the distribution of /y/ (including medial glide /ɥ/) in Sinitic dialects.

4.5 ‘Apical’ vowels

Sinitic languages are well known for their ‘apical vowels’, which are basically syllabic sibilants.²⁹ There are the alveolar [z] and the retroflex [ʐ]; the Sinologist symbols for these are <ɿ> and <ʅ> respectively. There are also the lip-rounded versions of these; the Sinologist symbols for these are <ɥ> and <ʉ>, respectively.

Unlike most Sinitic languages, but like languages in MSEA, many Sinitic languages in or near the Far-Southern zone lack apical vowels. These include most Yuè and Píngguà dialects, most Mǐn dialects, and some Gàn dialects. Map P118 of LACD (Figure 6) shows the distribution of apical vowels in Sinitic dialects.

The existence or non-existence of apical vowels is summarized in Table 2 in Section 4.6.

4.6 Summary of phonological traits

Table 2 summarizes the phonological points raised in Section 4.1 to Section 4.5. The Sinitic dialects in or near the Far-Southern zone (represented by Cantonese and Nánning Píngguà here) and Southern Mǐn show many more phonological traits that are more akin to the core of MSEA than to the other Sinitic languages.

□□

29 Phonetically, the amount of friction in the oral cavity varies between speakers when they produce the apical vowels. Stereotypically, Northerners produce apical vowels with prominent friction, and the friction lasts nearly throughout the duration of the rime. On the other hand, Far-Southerners stereotypically produce the apical vowels in Mandarin with a corresponding approximant or vowel. Most Chinese people pronounce the apical vowels somewhere between these two extremes: starting off with prominent friction, and then the friction weakens towards the end of the rime and becomes a phonetic vowel.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

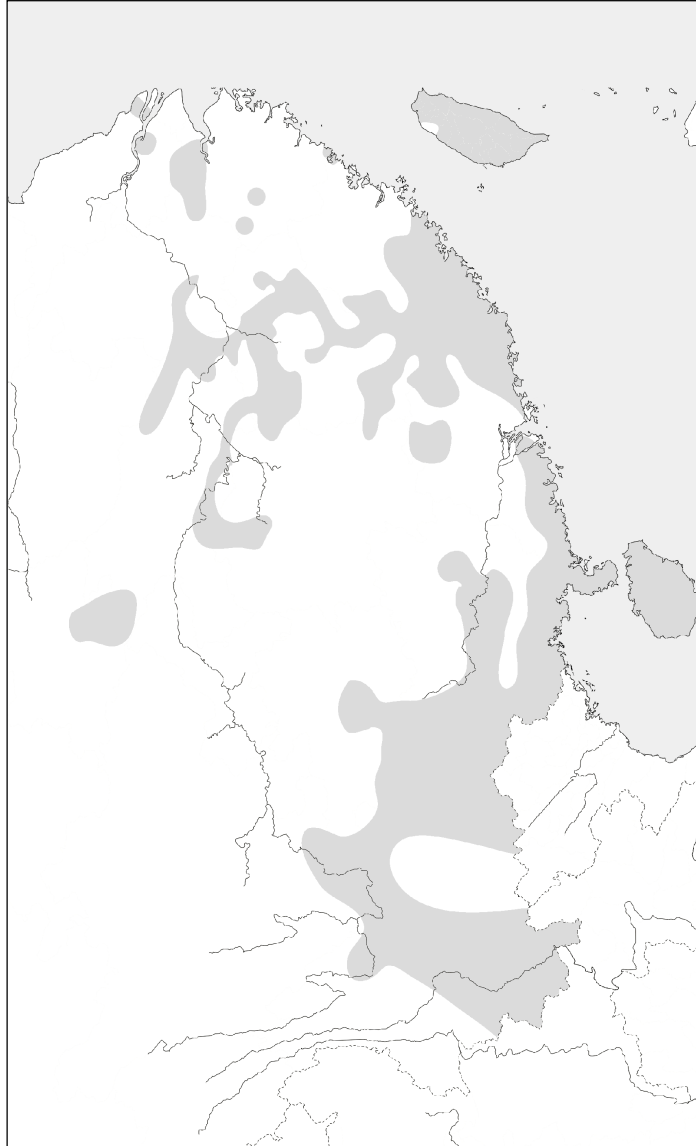


Fig. 5: Areas with neither medial glide ɥ nor vowel y are highlighted. Other Sinitic languages have both ɥ and y, or in rare cases, just one of them. (Data derived from LACD Map P117).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia : The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

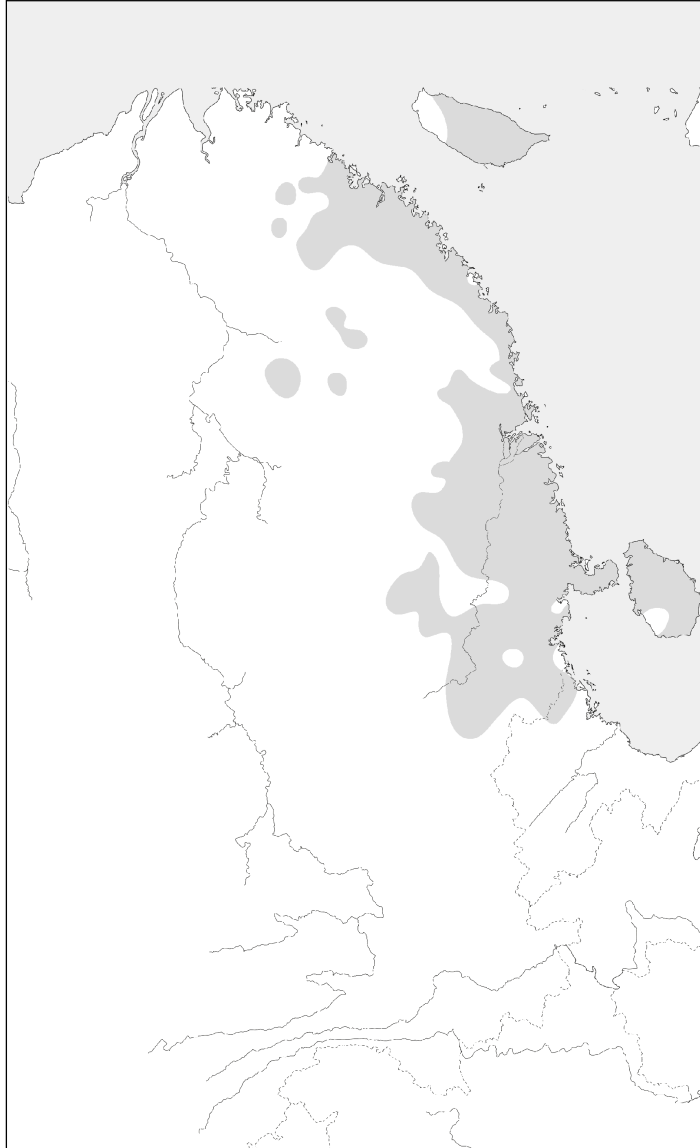


Fig. 6: Areas with no apical vowels are highlighted. Other Sinitic languages have at least one apical vowel. (Data derived from LACD Map P118).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

Table 2: Some phonological features in Sinitic and MSEA languages

	Non Sinitic						FS Sin.	SE Sinitic				C Sinitic		N			
	Thai	Khmer	Vietnamese	E Kayah Li	Burmese	Hmong	Mien	Cantonese	Nánning Pinghuà	Huì'án S Mìn	Fúqīng E Mìn	Fù yáng Wú	Jīxī Huī	'Shāowǔ Mìn-Gàn	Yìchūn Gàn	Gǔzhāng Xiānghuà	Suǐníng C Mandarin
Having "Complex tones" (WALS)	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
phonemicized tonal or phonational contrast from devoicing of onsets	+	(+) ³⁰	+	+	-	+	+	+	+	+	-	+	+	+	+	-	+
More than one contrastive plosive coda	+	+	+	-	-	-	+	+	+	+	-	-	-	-	-	-	-
More than one contrastive nasal coda	+	+	+	-	-	-	+	+	+	+	-	-	-	+	+	-	+
<i>b d ~ b d</i> (but no <i>g ~ g</i>)	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
No rounded front vowels (+ = no)	+	+	+	+	+	+	+	-	+	+	-	-	-	-	-	-	-
No 'apical vowels' (+ = no)	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	-
Total:	7	6	7	5	3	4	6	5	6	6	3	2	2	3	3	2	3

5 Word order

The 'basic' word order in the core of MSEA is SVO. The core MSEA languages are also more strongly left-headed than the usual SVO language (see, e.g. Dryer 2001 on Mon-Khmer word order). The Sinitic languages are also primarily SVO. However, the Sinitic languages are otherwise strongly right headed: noun phrases are strongly right headed, and most adjuncts are

□□

30 Except for some conservative Khmer dialects which have preserved the phonational contrast (e.g. Thung Kabin Khmer in Chanthaburi, Thailand; Wayland and Jongman 2003), all Khmer dialects have lost the original phonational contrast.

31 However, apical vowels exist in Dànánshān Hmong (Wáng 1985: 18), the standard variety of Western Hmongic in China.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

placed before the verb. Contrast the word order in the following sentences from Northern Zhuang (Tai) and Cantonese (Sinitic).

(2) Northern Zhuang (SVO order)

de gai byaek youhcaiq gai noh
 3SG sell vegetable as:well sell meat
 ‘S/he sells vegetables and sells meat.’
 (Wéi and Qín 2006: 198)

(3) Head noun left of most modifiers

go oij [duz vaiz go caij laem
u x
 CLF sugar_cane CLF buffalo 1s ste fall
 G p
hen roen haen raek lo
 side road that brea FP
 k

‘The sugar cane that my buffalo trampled on the side of the road snapped.’

(Wéi and Qín 2006: 251)

(4) Standard Cantonese (SVO order)

佢 賣 菜 又 賣 肉
kʰəy¹³ mai²² tʃɔi³³ jeu²² mai²² jɔk²
 3SG sell vegetable as:well sell meat
 ‘S/he sells vegetable and sells meat.’

(5) Head noun right of modifiers

我 隻 牛 喺 路 邊 踩 𨋖
[ŋɔ¹³ tsɛk³ ŋeu¹¹ hei²⁵ lou²² pin⁵⁵ tsʰai²⁵ lem³³]
 1SG CLF bovine at road side step fall
 嗰 碌 蔗 斷 咗
ko²⁵ luk⁵ tsɛ³³ tʰyn²⁴-
tsɔ²⁵

DEM CLF cane break-PFV

‘The sugar cane that my buffalo trampled on the side of the road snapped.’

This mix of SVO word order and strong right-headedness has created some extraordinarily rare co-occurrences of word order traits in the Sinitic
 20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
 Do not quote or cite this draft.

languages. For instance, the co-occurrence of VO and Relative clause–Noun is nearly unique to the Sinitic languages (WALS feature 96A).³² The Sinitic languages are the only VO languages with obliques predominantly placed in front of the verb in WALS (feature 84A).³³ Having the Adjective–Noun word order (feature 81A) for SVO languages (feature 87A) is also very rare in the region.³⁴

Looking at the word order typological profiles of the languages in the vicinity of the Sinitic languages provides hints as to why the Sinitic languages developed such an unusual mixture of VO order and strong right-headed traits. The Sinitic languages had the most interactions with the following three neighbouring word order areas:

Area A

Area A is the verb-medial core-MSEA zone to the south. The prototypical MSEA languages are SVO and more left-headed than the average SVO languages. Included in this zone are the Hmong-Mien, Kra-Dai, Mon-Khmer and Chamic languages. In the following examples, the clauses are verb-medial, and the modified constituents are generally to the left of the modifiers.

□□

32 Of the 879 languages sampled in WALS feature 96A, five have the co-occurrence of VO and Rel–N. Cantonese, Hakka and Mandarin are Sinitic. Bai is strongly influenced by Sinitic languages. Amis is also geographically close-by, but this co-occurrence in Amis is probably independent of Chinese (Comrie 2008b). As Comrie (2008b: 729–730) points out, having Rel–N order in SVO languages might aid processing when the object is relativized, as having a SV relative clause in front of the relativized object head resembles the normal SVO word order (Yip and Matthews 2007). There are indeed cases like Pwo Karen where relativized objects can have a prenominal relative clause, and relativized subjects must have a postnominal relative clause (Kato 2003: 641), resembling normal SVO word order in both cases (with the relative clauses considered externally headed in both cases).

33 Of the 500 languages sampled in WALS feature 84A, only the three Sinitic languages sampled have the word order of XVO (where X is an oblique).

34 Based on WALS feature 81A (SVO) and 87A (Adjective–Noun), there are 347 SVO languages with the Noun–Adjective order, and 66 SVO languages (including the Sinitic and Bai) with Adjective–Noun word order. This latter co-existence is mostly concentrated in Europe (20 languages) and Central Africa (15 languages). On the other hand, in Asia, including Western Austronesia, there are only two languages other than Sinitic and Bai which are marked as SVO and Adjective–Noun in WALS: Kashmiri and Palauan. However, the status of both being SVO is questionable. Kashmiri is verb-second (e.g. Wali and Koul 1996, Koul and Wali 2006). With Palauan, the slot in front of the verb can only be occupied by a subject agreement marker; subject nominals are placed after the object (i.e., VOS; Georgopoulos 1986). This leaves the Sinitic languages and Bai as the only SVO and Adjective–Noun languages in Asia.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

- (6) Utsat (Chamic; Chinese influenced)
ʔa¹¹thai¹¹ se¹¹ phai³³sian¹¹ ho¹¹lien¹¹ ʔa¹¹kai³³ sa³³ ta¹¹ se⁵⁵,
 litl.sister CLF very feel:sorry old.man MOD one CLF
kian³³ ʔa¹¹kai³³ ni³³ sa³³ ta¹¹ se⁵⁵ ten³² pa³³,
 know old.man this MOD one CLF stomach hungry
 ‘The little sister was very sorry for the old man, and knew that the old man was hungry,’ (Zhèng 1997: 238)
 (*phai³³sian¹¹ ho¹¹lien¹¹* 非常可憐 are Chinese loanwords in Chinese word order.)
- (7) Green Hmong (Hmong-Mien)
kuv nyam tug txivneej kws ncaws pob
 1SG like CLF man REL kick ball
hab tug txivneej kws moog rua Fresno
 and CLF man REL go to Fresno
 ‘I like the man who plays soccer and the man who went to Fresno.’
 (Li 1989: 120)

Area B (and Area A~B)

Area B is the verb-final Tibeto-Burman zone to the west. These languages are SOV, they are generally right-headed, but they also have some left-headed traits (e.g. Tibetan and Burmese are SOV and have N–Num and N–Adj word order). Having N–Adj is in fact the norm for SOV languages cross-linguistically (Dryer 2013). In the following examples, clauses are verb-final, adpositions are placed at the right edge, and the modified constituents are to the right of some modifiers, and to the left of some modifiers.

- (8) Burmese
thu di hsei: thau' me
 3 this medicine drink IRR
 ‘He’s going to take this medicine.’
 (Soe 1999: 132)
- (9) *thu. le' nyi'=pa' ne. nga kou la tou. te*
 3GE hand dirty with 1 OBJ com touch RL
 N e S
 ‘(He) touched me with his dirty hands.’
 (Soe 1999: 256)

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

There are also languages which are transitional between area A and area B. Some Tibeto-Burman languages are exceptionally SVO. They, like the Sinitic languages, exhibit interesting mixtures of properties associated with VO and OV orders. These SVO Tibeto-Burman languages include the Karen languages, Bái languages, and Mru (Peterson 2005).³⁵

(10) Eastern Kayah Li (Karenic)

phremò méthl phrekhū sí nɿ
 woman look:see man CLF two
 'Some women saw two men.'
 (Solnit 1997: 181)

(11) *ʔa khɛ̃ təlwá sɔklɿ nɛ̃ sɔkhō*

3 paddle pass boat PREP snag
 'He paddled the boat past the snag (fallen log).'

(Solnit 1997: 159)

Area C

Area C is the verb-final North Asia zone to the north. These languages are SOV and strongly right-headed. In and near China are the following families of SOV languages: Turkic, Mongolic, Tungusic, Korean and Japanese-Ryūkyūan. The historical Tokharian languages also fit this typological profile.³⁶ In the following examples, clauses are verb-final, and modified constituents are always to the right of the modifiers.

□□

35 Tibeto-Burman languages that have SVO word order are often assumed to have acquired SVO word order under the influence of neighbouring SVO languages. Mru is an interesting case because it is totally surrounded by verb-final languages (Chittagonian, Rakhine, and Kuki-Chin languages). It is also spoken very far away from verb-medial languages like the Khasic or Palaungic languages, and there seems to be no Mon-Khmer lexical borrowings in Mru (Löffler 1966). See more discussions in Djamouri, Paul and Whitman (2007).

36 Other than the three typological areas discussed here, there are also the following typological areas in and around China that the Sinitic languages have less contact with: a) languages of the Formosan–Philippine area, which are mostly verb initial; b) languages of the Indic area, which are verb final and strongly-right headed, except Kashmiri and several other Dardic languages which is verb-second; and c) languages of the Iranian area, with Sarikoli and Wakhi represented in China (Gawarjon 1985). These two Pamiri languages are verb final and more strongly right-headed than the other Iranian languages, but they still have the Iranian trait of having prepositions (although they also have some Uyghur-like postpositions).
 30540510 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
 Do not quote or cite this draft.

'tall person'
(Enfield 2007: 93)

- (15) *khaw5 niaw3*
rice sticky
'sticky rice'
(Enfield 2007: 93)

- (16) Khmer
civeut ti: pi: rabawh knjom
life place two of me
'my second life'
(Haiman 2011: 168)

- (17) Eastern Kayah Li
?iswí nɿ bēlɔ̃ du
curry two bowl big
'two big bowls of curry'
(Solnit 1997: 180)

In the periphery of MSEA, Burmese, which is verb final, has some post-verbal modifiers, like the nominalized stative verb *a-thi* 'new' and stative verb *hklei*: 'small' in the following example. (Attributive nouns like *thi'tha*: 'wood' precede the head noun.)

- (18) Burmese
thi'tha: ein a-thi' hkalei:
wooden house new small
'small new wooden house'
(Myint Soe 1999: 44)

Looking into the history of Chinese, noun phrases were already mostly right headed in Pre-Archaic and Archaic periods.

(19) Pre-Archaic Chinese (14th to 11th century BCE)³⁷

上甲 惠 王 報 用 五 伐
shàngjiǎ huì wáng bào yòng wǔ fá
 shangjia FOC king bao:sacrifice use five human:victim
 十 小 宰
shí xiǎo zǎi
 ten little sacrificial:sheep

‘As for (the ancestor) Shangjia, it must be the king who addresses (him) with a *bao* sacrifice by using five human victims and ten little sacrificial sheep.’

(Djamouri 2001: 162; Jiágǔwén Héjǐ 924)

(20) Early Archaic Chinese

天 不 庸 釋 于 文 王 受 命
tiān bū yōng shì yú [[wén wáng shòu] mìng]
 heaven not then relinquish to [[wen king receive] destiny]

‘Then Heaven will not relinquish [the destiny which King Wen received].’

(Aldridge, 2013: 47; Shàngshū, Jūnshì 君奭; approx 8th century BCE)

(21) 非 時 伯夷 播 刑 之 迪?

fēi [[shí bóyí bō] xíng] zhī dí?
 not.be [[then boyi promulgate] law] GEN guide

‘Is it not the laws promulgated by Boyi which guide (you)?’

(Aldridge, 2013: 47; Shàngshū, Lǚxíng 呂刑; approx 8th century BCE)

However, there were some post-nominal modifiers in the earliest stages of Chinese. SVO languages typically have some pre-nominal and some post-nominal modifiers, and the earlier stages of Chinese had more post-nominal modifiers than the modern Sinitic languages.

□□

37 As is the convention in the West and most of China, historical Chinese texts are transcribed and pronounced in modern Mandarin pronunciation. The pronunciation of the characters in Pre-Archaic Chinese (fourteenth to eleventh century BCE) is earlier than the earliest reconstructable phonological form of Chinese (Old Chinese: tenth to seventh century BCE) anyway.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

- (22) Pre-Archaic Chinese (14th to 11th century BCE)

子 央 歲 于 丁
 zǐ yāng suì yú dīng
 prince yang immolate to Ding

‘The prince Yang [will] immolate something for the ancestor Ding.’
 (Djamouri 2001: 146; Jiágǔwén Héjì 3018)

Numerals, in particular, were placed variously in front of or after the head noun.

- (23) 獲 唯 鳥 七

huò wéi niǎo qī
 capture COP bird seven

‘The catch is seven birds.’

(Djamouri 2001: 151; Jīnzhāng suǒ cáng Jiágǔ Búcí 742)

(Numerals were more often prenominal than postnominal in Pre-Archaic Chinese.)

The earliest classifier-like words more often follow, rather than precede, the head noun.

- (24) Pre-Medieval Chinese

分 與 文君 僮 百 人
 fēn yǔ wénjūn tóng bǎi rén

distribute give wenjun slave hundred people

‘(He) distributed a hundred slaves to Wenjun.’

(Chappell and Peyraube 2007; Shǐjì, Sīmǎ Xiāngrú Lièzhuǎn 司馬相如列傳, approx 1st century BCE)

- (25) Early Medieval Chinese

時 跋 跋提 國 送 獅子 兒 兩 頭 與
 shí bá bá tí guó sòng shīzi ér liǎng tóu yǔ
 time ? Bactria country offer lion child two CLF^{head} give
 乾陀羅 王

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

gāntuólúó wáng

gandhāra king

'At that time, the kingdom of Bactria offered two lion cubs to the king of Gandhāra.'

(Chappell and Peyraube 2007; Luòyáng Qíélánjì 5 洛陽伽藍記 5; 6th century CE)

These post-nominal classifier-like words in earlier stages of Chinese were argued to be not part of the noun phrase of the preceding noun (e.g. Peyraube 1988). Indeed, it can also be argued that the post-nominal classifiers do not form a phrase with the preceding noun in some MSEA languages. For example, in Lao, a phrase can often intervene between a [num + clf] phrase and the preceding noun which it attributes semantically.

(26) Lao

kuu3 sùù4 paa3 sòòng3 too3

1SG buy fish two CLF

'I bought two fish.'

(Enfield 2007: 120)

(27) *kuu3 sùù4 paa3 juu1 talaat5 sòòng3 too3*

1SG buy fish be.at market two CLF

'I bought fish at the market, two (of them).' (= 'I bought two fish at the market')

(Enfield 2007: 120)

Looking at the modern Sinitic languages, their noun phrases are even more strongly right-headed than the ones in older stages of Chinese.

(28) Nánning Píngguà

我 個 對 舊 皮 鞋

ŋa¹³ kə⁵⁵ tɔi⁵⁵ kəu²² pəi¹¹ hai¹¹

1SG DEM pair old leather shoe

'My pair of old leather shoes.'

Nevertheless, there are typically some non-productive left-headed compounds in the Southern Sinitic languages, e.g. Cantonese 魚生 *jy¹¹ saŋ⁵⁵* (fish raw) 'raw fish', 菜乾 *tshɔi³³ kən⁵⁵* (vegetable dry) 'dried vegetable', 人客 *jen¹¹ hak³³* (person guest) 'guest', 熊人 *hɔŋ¹¹ jen^{11>25}* (bear person) 'brown bear (child's word)'. (See also, for example, the many left-headed compounds in 20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

Wēnzhōu Wú (Zhèngzhāng 2008: 232.) More productive than these fixed compounds are the sex affixes for animals. The general trend is for the Northern Sinitic languages to have sex prefixes, resembling the right-headed word order in North Asia, and the Southern Sinitic languages to have sex suffixes, resembling the left-headed word order in MSEA. (Nánning Píngguà is a major exception for being in the Far-Southern zone, but having sex prefixes predominantly.) Some Sinitic dialects in the centre are somewhat mixed; for instance, some dialects have a prefix for one sex and a suffix for the other sex, or a prefix for one animal and a suffix for another animal.

Standard Mandarin (prefixes)

- (29) 公豬 *gōng-zhū* (male-pig) 'boar'
 (30) 母豬 *mǔ-zhū* (female-pig) 'sow'

Xiānghuà (prefixes and suffixes)

- (31) 〇豬 *əiŋ²⁵-tiəu⁵⁵* (male-pig) 'boar'
 (32) 豬娘 *tiəu⁵⁵-ŋjē⁵⁵* (pig-female) 'sow'

Fùyáng Wú (prefixes and suffixes)

- (33) 雄雞 *'hioŋ-təi* (male-fowl) 'rooster'
 (34) 雞娘 *təi-'niǎ* (fowl-female) 'hen'

Shàowǔ Mǐn-Gàn (suffixes)

- (35) 雞公 *kɛi²¹-kuŋ²¹* (fowl-male) 'rooster'
 (36) 雞嫲 *kɛi²¹-ma²²* (fowl-female) 'hen'

Fūqīng Eastern Mǐn (suffixes)

- (37) 雞公 *kiɛ³²-kuŋ⁵³* (fowl-male) 'rooster'
 (38) 雞母 *kiɛ³²-mɔ⁵³* (fowl-female) 'hen'

Cantonese (suffixes)

- (39) 雞公 *kɛi⁵⁵-kɔŋ⁵⁵* (fowl-male) 'rooster'
 (40) 雞𨾏 *kɛi⁵⁵-na²⁵* (fowl-female) 'hen'

Nánning Píngguà (prefixes)³⁸

- (41) 公雞 *kɔŋ⁵³-kɛi⁵³* (male-fowl) 'rooster'
 (42) 母雞 *mu¹³-kɛi⁵³* (female-fowl) 'hen'

□□

38 Píngguà dialects to the west also have gender prefixes, e.g. Chóngzuǒ (Lǐ and Zhū 2009: 177).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

The following table summarizes the noun phrase features discussed in Section 5.1. In general, the languages in the core of MSEA have rather strongly left-headed noun phrases, whereas the modern Sinitic languages have strongly right-headed noun phrases. The Southern Sinitic languages have marginally more nominal left-headedness in having some morphologically left-headed words.

Table 3: Left-headedness on the noun phrase level in some Sinitic and MSEA languages

	Non Sinitic			FS Sin.	SE Sinitic	C Sinitic	N										
	Thai	Khmer	Vietnamese	E Kayah Li	Burmese	Hmong	Mien	Cantonese	Nánning Pínghuà	Huǐ'ān S Mǐn	Fúqīng E Mǐn	Fúyáng Wú	Jīxī Huī	'Shàowǔ Mǐn-Gàn	Yīchūn Gàn	Gūzhàng Xiānghuà	Suǐníng C Mandarin
N – Genitive	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
N – “Adjective” (e.g. chicken – big)	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-
N – Noun (e.g. egg – chicken)	+	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-
N – Sex (e.g. chicken – male)	+	+	+	+	+	+	+	+	-	+	+	±	±	+	+	±	-
N – Demonstrative	+	+	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-
N – Relative clause	+	+	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-
Total:	6	6	6	3	2	5	2	1	0	1	1	½	½	1	1	½	0

5.2 Word order in clauses

The core MSEA languages are SVO, and modifiers usually follow the head. The Sinitic languages are also said to be primarily SVO. However, these languages require the preposing of objects to a pre-verbal position in some situations. In addition, other than the VO word order, the Sinitic languages are strongly right-headed.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

This rare combination of SVO word order and strong right-headedness in the modern Sinitic languages, and the fact that the vast majority of Tibeto-Burman languages, i.e., the relatives of the Sinitic languages, are verb-final, has led to the common assumption of Chinese having more verb-final traits the further one goes back into the history of Chinese (Li and Thompson 1974: 208, LaPolla 1994). However, looking at the written records of Chinese up till fourteenth century BCE, the opposite was true: the further one goes back into the history of written Chinese, the more verb-medial traits there were (Peyraube 1997; Djamouri, Paul, and Whitman 2007). In other words, the exceptionally rare combination of SVO and right-headedness in Chinese has been stable for at least thirty-four centuries (from fourteenth century BCE to twentieth century CE), and Chinese developed from a strange SVO language into a group of even stranger SVO languages, typologically speaking.

First of all, Pre-Archaic Chinese was clearly a SVO language: looking at Pre-Archaic Chinese texts (Shang Dynasty oracle bone script), 93.8% of clauses with two place predicates were (S)VO in Djamouri's corpus (2001: 146); OV order only occurred in specific syntactic environments.³⁹ Pre-Archaic and Archaic Chinese also had *wh*-movement, which is a trait not uncommon for VO languages, but rare for OV languages (e.g. Dryer 1991). Modern Sinitic languages have most obliques placed in front of the verb, which is extremely rare for VO languages. (In WALS, the modern Sinitic languages are the only VO languages that predominantly place oblique phrases before the verb (WALS feature 84A).) However, Pre-Archaic Chinese is a relatively normal VO language, in that it usually places obliques after the object (i.e. VOX word order).⁴⁰

(43) Pre-Archaic Chinese

呼 多 犬 网 鹿 于 麓
hū duō quǎn wǎng lù yú nóng
 order numerous dog.officer net deer at Nong

□□

39 In Pre-Archaic Chinese and Archaic Chinese, OV order only occurred in: a) cleft constructions: {COP ... O V} (the copula was obligatory in Pre-Archaic Chinese, but became optional in the Early Archaic period); b) negative sentences with an accusative pronoun: {NEG O V} (in Pre-Archaic Chinese this was restricted to the negator 不 *bù* (Djamouri, Paul and Whitman 2007: 4), but in Archaic Chinese this applies to other negators as well); and c) *wh*-questions; the non-subject question word is placed between the subject and the verb: {S Q V?}. See Aldridge (2013).

40 Other than the post-object position, another common position for locative phrases, for temporal phrases in particular, is the pre-subject position (Djamouri 2001: 147–148).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

‘Call upon the many dog-officers to net deer at Nong.’
(Djamouri, Paul, and Whitman 2007: 3; *Jiágǔwén Héjì* 10976 recto.)

The same VOX word order is also the norm in the core of MSEA. The following are some examples.

(44) Lao
phen1 lin5 phaj4 juu1 talaat5
3POL play cards be.at market
‘She is playing cards at the market.’
(Enfield 2007: 390)

(45) Khmer
knjom tradaw: sra:j krama: pi: cangkeh
I struggle untie scarf from waist
‘I struggle to untie the scarf from my waist.’
(Haiman 2011: 204)

In contrast to Pre-Archaic Chinese, which is a relatively normal SVO language, two related tendencies developed amongst the modern Sinitic languages (e.g. Zhāng 2010, Liú 2012, Bisang 2012):

- the Sinitic languages accept postverbal constituents less readily
- in many Sinitic languages, the association of postverbal constituents with new information became stronger⁴¹

This created many more verb-final sentences in the modern Sinitic languages than older stages of Chinese. These traits are relatively weak in the Far-Southern Sinitic languages, Cantonese for instance; the Far-Southern Sinitic languages are relatively close to the core of MSEA, in both a geographical sense, and also in a typological sense, in that the Far-Southern Sinitic languages have the most verb-medial traits amongst the Sinitic languages. The Northern Sinitic languages have many verb-final traits; the Far-Western Central Plains Mandarin dialects even have postpositions and are

□□

41 For Mandarin, Li (2011) characterizes postverbal constituents as primarily conveying new information. There are also accounts which characterize postverbal constituents in Mandarin as focused (LaPolla 1995) or indefinite (Li and Thompson 1975). While the information status account seems to model the situation in Mandarin well, in other Sinitic languages definiteness may be the primary motivating factor. More studies are needed on the variation in word order amongst the Sinitic languages.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

predominately SOV. The Northern Sinitic languages have been under the influence of verb-final Altaic languages. Nevertheless, putting the aforementioned SOV Mandarin dialects aside, the Sinitic languages with most verb-final sentences are not the Northern Sinitic languages, but the Southeastern Sinitic languages, which are not known to have significant contacts with verb-final languages.⁴² It is rare for more than one constituent to occur after the verb. As an example of an often-verb-final Southeastern Sinitic language, M. Qián (2008) summarizes the following syntactic environments where sentences have to be verb final in Níngbō Wú (with my reinterpretation and with the help of the description of the tense and aspect system of Níngbō Wú in N.R. Qián 2008):

- Sentences with a post-verbal tense-aspect marker (e.g. present perfective, past perfective, durative, simultaneous, experiential; these markers are often grammaticalized from locative words)
- Some Irrealis sentences, e.g.:
 - Negative sentences (S – O – neg – V)
 - Yes-no questions (S – O – V – Q)
 - Rhetorical questions (S – O – V – Q)
 - Imperative sentences (except that [num-clf] phrase and verbal complements can occur post-verbally)
- Emphatic possessive sentences (S – O – possess – emph)
- ‘To’ (e.g. I place go) and ‘from’ (e.g. I place from go)
- Transitive sentences with an object which is definite

Contrast this with a Far-Southern Sinitic language like Cantonese, where all of these sentences above would normally be in SVO order, similar to a canonical MSEA language.

□□

42 The reason for this is unknown to me. Perhaps this is an independent development. It is known that in SVO languages, there is a correlation between the preverbal position and definiteness (see Section 5.2.3), and perhaps the Southeastern Sinitic languages further grammaticalized this on their own accord.

Before Southeastern China was Sinicized, the indigenous people in the area spoke Kra-Dai, Hmong-Mien, and perhaps also Austroasiatic languages, none of which are known to have SOV word order. Whether there were SOV-speaking indigenous people in the area or not is not known to me. There had been some, typologically speaking, relatively insignificant contacts with SOV languages from the east across the sea: the colonization of Southern-Mín- and Hakka-speaking Taiwan by Japan, and the historical link between the Eastern-Mín-speaking Fúzhōu and the Ryūkyūan Kingdom.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

In the following subsections, I will discuss the situations in which non-subject constituents have to be preverbal in Sinitic languages. I will show that the Northern and Southeastern Sinitic languages have more instances of verb-final sentences, whereas the Far-Southern Sinitic languages have far fewer instances of verb-final sentences, being closer to the core of MSEA. The following word order traits will be discussed:

- Position of adverbials and adpositions (Section 5.2.1)
- Position of modifiers of verbs (Section 5.2.2)
- Position of objects (Section 5.2.3)
 - The object marking construction (Section 5.2.3.1)
 - Preverbal and Postverbal definite objects (Section 5.2.3.2)
 - Word order in clauses with three place predicates (Section 5.2.3.3)

A summary of Section 5.2 is presented in Section 5.2.4.

5.2.1 Position of adverbials and adpositions

Modern Sinitic languages allow post-verbal constituents less readily than Archaic Chinese and MSEA languages. To my knowledge, the only modern Sinitic language that, like Archaic Chinese, commonly has adverbials after the object is Jīxī Huī.

(46) Jīxī Huī

我 看 電影 是 電影園 (裏)
 a^{55} $kʰa^{324>35}$ $tʰe^{223}$ ia^{55} se^{55} $tʰe^{223}$ $ia^{55>53}$ ye^{223} (ni)
 1SG see film at cinema (in)
 'I watched a film in the cinema.'⁴³

(47) 爾 ○ 柴 (是) 哪 ○ 啊 ?

o^{55} $tsoʔ^{32>35}$ sa^{32} (se⁵⁵) $na^{55>53}$ xa^{324} $a?$
 2SG chop firewood at where Q
 'Where do/did you chop firewood?'

□□

43 是 se^{55} is a locative preposition in Jīxī Huī. The copula is also 是 se^{55} .

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

Otherwise, it is probably universal amongst modern Sinitic languages that most adverbials are placed in front of the main verb, especially for temporal phrases. The following are some examples.

(48) Nánning Píng huà

我 太早 住 屋頭 看了 一 出 戲
 ŋa¹³ tai²² tʃau³³ tʃəi²² ŋk³ tɕu¹¹ han²⁵ -lə³³ et³ tʃət³ həi²⁵
 1SG just:now at home watch- one CLF film
 PFV

'I watched a film at home just now.'

(49) Xiāng huà

我 朝頭 ○ 三 個 糞
 u²⁵ tiau⁵⁵ ta zieu¹³ so⁵⁵ keu³³ tɕi¹³
 1SG morning eat three CLF bun

'I ate three buns this morning.'

(50) Standard Mandarin

我 明天 在 站台上 等 你
 wǒ míngtiān zài zhàntái shàng děng nǐ
 1SG tomorrow at platform on wait 2SG
 'I will wait for you at the platform tomorrow.'

MSEA languages, on the other hand, usually have many adverbials which can be placed after the object.

(51) Vietnamese

bố cháu đã từng dạy học ở Ha-oai
 father 1SG ANT EXP teach study in Hawaii
 'My dad has taught in Hawaii.'
 (Nguyễn 1997: 158)

(52) Thai

sua kaw ca aw pay boricàak phrûnii
 clothes old will take go donate tomorrow
 'I'll give away the old clothes tomorrow.'
 (Smyth 2002: 117)

While most adverbials are placed in front of the verb, most Sinitic languages have some location phrases that are placed after the verb (as arguments or adjuncts, depending on the verb). This is especially the case with destinations.

(53) Cantonese

我 今日 去 台北
 ŋɔ²⁵ kem⁵⁵jet² həy³³ tʰɔi¹¹pek⁵
 1SG today go Taipei
 'I am going to Taipei today.'

(54) Fúqīng Eastern Mǐn

我 今晏 去 北京
 ŋua³² kiŋ⁵³naŋ²¹ kʰyɔ²¹ peŋ³kiŋ⁵³
 1SG today go Beijing
 'I am going to Beijing today.'

However, some Sinitic languages require even destinations to be placed before the main verb. This is the norm in Wú and Huī in the Southeastern Zone, the Sinitic dialects in the Northern Zone, and some in the Central zone. The destination precedes the verb, and the destination is at least preceded by a preposition.⁴⁴

(55) Jīxī Huī

到 績溪 去
 tɛ³²⁴ tseŋ^{32>35}tɛ^h²¹ kʰe³²⁴
 to Jīxī go
 'Going to Jīxī.'

(56) Xiānghuà

你 到 何○ 去?
 ŋ²⁵ tau³³ uo¹³ŋ⁴¹ kʰəu³³?
 2SG to where go

□□

44 The constituents translated as 'to' are grammaticalized from verbs; as main verbs, 到 is 'arrive', and 走 in Wēnzhōu is 'go'. However, the 'to' in these examples are no longer verbs. For instance, they cannot take any verbal morphology.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

'Where are you going?'

(57) Pínglǐ Central Mandarin

你 到 哪兒 去 耶? 我 到 城 裏頭 去
 ɲ⁴⁴ tau²³ lar⁴⁴⁵ tɕh²³ iɛ? ɲo⁴⁴ tau²³ tɕhən⁵² li⁴⁴⁵ t'ou tɕh²⁴

2SG to where go Q 1SG to city in go

'Where are you going? I am going to the city.'

(Zhōu 2009: 408)

(58) Wēnzhōu Wú (Southern Wú)

我 走 溫州 去
 ɲ³⁴ tsau^{45>0} ʔjy^{33>11} tɕəu³³ k'he^{42>0}

1SG to Wenzhou go

'I am going to Wenzhou.' (Zhèngzhāng 2008: 340)

In Northern Wú dialects, the preposition is usually elided (discussed below), resulting in what appears to be a SOV sentence.

(59) Fùyáng Wú (Northern Wú)⁴⁵

我 今朝 (到) 上海 去
 ɲɿ 'kintɕɔ ('tɔ) zɔŋhɛ tɕhi

1SG today to Shanghai go

'I am going to Shanghai today.' (It is more common to omit 'tɔ 'to'.)

The Sinitic languages have both prepositions and postpositions. SVO languages usually have prepositions. Postpositions are rarer for SVO languages. However, having postpositions in a SVO language is itself not too surprising, if the postposition is grammaticalized from a noun, and when genitives occur in front of the noun. So, to indicate location, instead of having a left headed structure like the following from Northern Zhuang:

(60) Northern Zhuang

[youq [gwnz [taiz]]]

at above table

'on the table'

□□

45 A proper analysis of the tonal system in Fùyáng Wú is yet to be done. There are two or three contrastive word melodies (and various allo-melodies).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

Sinitic languages would have a general locative preposition, and a postposition which signifies a semantically narrower locative relation. In Sinitic languages, the locative postposition, which is grammaticalized from a noun, is usually no longer a free noun. For instance, in the following example, 上 *teŋ*²² ‘above’ is not a noun meaning ‘top’.

- (61) Nánning Píngguà
 住 檯 上
 [tsəi²² [[tai¹¹] teŋ²²]]
 at table above
 ‘on the table’

Similar structures exist in Karenic languages, which also have mixed VO-associated and OV-associated typological profiles like the Sinitic languages. However, in Eastern Kayah Li at least, the postnominal locative word is still a noun.

- (62) Eastern Kayah Li
dʁ' lɛ̃ kũ
 at ravine interior
 ‘in the ravine’
 (Solnit 2007: 209)
- (63) *dʁ' pjā kũ*
 at bag interior
 ‘in the bag’
 (Solnit 2007: 209)
- (64) *dʁ' hi lɛ̃*
 at house bottom
 ‘under the house’
 (Solnit 2007: 211)
- (65) *dʁ' dɔ̃ lɛ̃*
 at village bottom
 ‘below (downhill from) the village’
 (Solnit 2007: 211)

What is surprising is that the (newer) locative postposition has become obligatory in some Sinitic dialects. This is especially the case in Wú dialects 20140919 *Draft of*: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

(e.g. Liú 2003; 2012: 11–12). Looking at some less-unusual SVO languages first, the locative postposition is usually optional in Cantonese and Mandarin.

(66) Cantonese

掛 喺 客廳 (道)
kwa³³ hei²⁵ hak³tɛŋ⁵⁵ (tou²²)
 hang at living.room at
 'hung up in the living room'

(67) Mandarin

掛 在 客廳 (裏)
guà zài kètīng (lǐ)
 hang at living.room in
 'hung up in the living room'

On the other hand, the postposition is compulsory in most Wú dialects (Liú 2012: 12).

(68) Sūzhōu Wú

掛 勒 客廳 *(裏)
*ko⁵² ləŋ⁵⁵ kʰaŋ⁵⁵tʰin²³ *(lǐ⁴⁴)*
 hang at living.room in
 'hung up in the living room'
 (Lǐ 1998: 164)

Whereas the preposition is often optional in Northern Wú dialects.

(69) Níngbō Wú (Preposition usually omitted for preverbal adverbials)

賊骨頭 (來) 厠坑間 裏 幽 該
 thief (at) toilet in hide FP
 'The thief hid in the toilet'
 (Liú 2003: 272)

(70) 老師 (來該) 黑板 上 寫 字
 teacher (at) black:board on write word
 'The teacher wrote on the blackboard'
 (Liú 2003: 272)

In fact the preposition is often optional, or even used as a postposition in some Northern Wú dialects.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

Prepositions made into postpositions

- (71) 圖書館 裏 來該
 library in at
 ‘at the library’
 (Liú 2003: 272)

In Níngbō Wú (and most other Northern Wú dialects), ‘go to’ is usually expressed with no adposition, whereas ‘come from’ is usually expressed with a postposition ‘from’. The Northern Wú dialects (especially the ones spoken outside of Shanghai) in general show many verb-final typological traits, while SVO word order is still commonly used.

- (72) Níngbō Wú
 囡囡 幼兒班 去
 baby kindergarten go
 ‘Baby goes to kindergarten.’
 (M. Qián 2008: 136)

- (73) 我 學校 介 來
 1SG school from come
 ‘I came from the school.’
 (M. Qián 2008: 136)⁴⁶

5.2.2 Position of adverbials

Adverbials are usually placed in front of the verb.

- (74) Shanghainese (Wú)
 搵個 人 討飯 能介個 樣子 立辣 依答

□□

⁴⁶ M. Qián (2008: 136) describes 介 as a postposition meaning ‘from’. However, Zhū et al. (1996), the Níngbō dictionary, only lists 介 /ka⁴⁴/ as being a demonstrative meaning ‘like this’ or a particle meaning ‘-like’ (Zhū et al. 1996: 40–41). I would like to thank my colleague Xūpíng Lǐ for questioning the status of 介 as a postposition.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

geq- njin thaovae nenkaxeq xiangtsir liq-laq iitaq
geq
 this-CLF person beggar like appearance stand- there
 PROG

'The man stood there like a beggar.'
 (Zhu 2006a: 155)⁴⁷

(75) Standard Cantonese

佢 慢慢 行
kʰeɿ¹³ man²² man²⁵ haŋ¹¹
 3SG slowly walk
 'S/he walks slowly.'

(76) Xiānghuà

你 快 ○手
nǐ²⁵ kʰuə³³ tsau²⁵ ɕiəw²⁵
 2SG quick move:hand
 'Hurry up and get moving.'

(77) Standard Mandarin

你 先 吃 吧 多 吃 一 點
nǐ xiān chī ba duō chī yīdiǎn
 2SG first eat FP more eat a:bit
 'Eat first. Eat a bit more.'

However, many Southern Sinitic dialects (primarily Wú, Gàn, Hakka, Yuè, Píngguà, Hǎinán Mǐn) have a few adverbs which are placed after the verb (either immediately after the verb, or at the end of the clause).

(78) Fùyáng Wú

杭州 到 快 噍。
hãtsɿ 'tɔ 'kʰua diɛ

□□

47 Wú languages have tonal domains that are longer than a syllable. In Shanghainese, except for toneless syllables, there are two contrastive tonal melodies. Zhu (2006a) notates the 'marked' melody with a grave accent.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

Hángzhōu arrive soon COS

'We are arriving in Hángzhōu soon.'

(This 快 'k^hua may be a prospective marker. 快 'k^hua meaning 'fast' is placed in front of the verb.)

(79) Yíchūn Gàn

(再) 去 幾 個 湊
 (tsæ⁴⁴) tɕʰiɛ⁴⁴ tɕi⁵³ kɔ⁴⁴ tʂʰɛu⁴⁴
 again go few CLF more
 'Send a few more people.'

(80) 食 多 發積

tɕʰiʔ⁵ to³⁴ faʔ⁵-
 tɕiʔ⁵
 eat more bit-DIM
 'Eat a bit more.'

(81) 你 食 飯 先

ŋi³⁴ tɕiaʔ⁵ fan²¹³ sɿɛn³⁴
 2SG eat rice first
 'You eat your meal first.'

(82) Hakka

坐 一 下 添
 tʂʰo²⁴ it² ha⁵⁵ tʂiam²⁴
 sit one CLF more
 'Sit a bit more.'
 (Lo 1988: 301–302)

(83) 著 少 一 領 衫

tʂok² sɿɛu³¹ it² liaŋ¹¹ sam¹³
 wear less one CLF clothes
 'Wear one piece of clothing less.'
 (Lo 1988: 303)

(84) Standard Cantonese

食 埋 雪糕 添 啦
 sik² mai¹¹ syt³kou⁵⁵ tim⁵⁵ la⁵⁵
 eat as_well ice:cream in_addition FP

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

'Have ice cream too!'

- (85) 打 多 兩 行 字
ta²⁵ t₂⁵⁵ lœŋ¹³ hœŋ¹¹ tsi²²
 hit more two line word
 'Type two more lines.'

- (86) 我 行 先 啦。
ŋɔ¹³ haŋ¹¹ sin⁵⁵ la³³
 1SG go first COS
 'I am going now.'
 (See, e.g., Peyraube 1996, who discusses the post-verbal adverbs in Cantonese.)

MSEA languages usually have adverbials after the verb.

- (87) Northern Zhuang
gou bæ gonq
 1SG go first
 'I am going now.'

- (88) *gou gwn vanj hœux dem*
 1SG eat bowl rice in_addition
 'I eat another bowl of rice.'
 (Wéi and Qín 2006: 208)

(The word *dem* itself is perhaps a Chinese loan, c.f. Cantonese 添 *tim*⁵⁵ 'add'.)

- (89) Thai
raw pay thîaw mœaŋ thay bɔ̀y bɔ̀y.
 1PL go trip country Thai often
 'We visit Thailand often.'
 (Smyth 2002: 104)

- (90) Green Hmong
tuam moog rua suavteb hab
 Tuam go to China too

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

'Tuam went to China too.' (Li 1989: 121)

- (91) Khmer
knjom kampung raut lee:ng ja:ng sa'ba:j
 I engage.in run play kind happy
 'I was running along happily.'
 (Haiman 2011: 216)

- (92) *knjom skoal koat cbah nah*
 I recognize 3 clear very
 'I recognized him very clearly.'
 (Haiman 2011: 216)

5.2.3 Position of objects

Not only are adverbials mostly placed in front of the verb, objects are also sometimes placed in front of the verb in the Sinitic languages. Although the Sinitic languages could be said to be SVO in general, constituents that can occur postverbally are restricted. With relatively few restrictions are the Far-Southern Sinitic languages like Cantonese; Far-Southern Sinitic languages are relatively free to have two or more constituents after the main verb. At the other extreme are the Southeastern Sinitic languages, where it is rare to have more than one constituent after the verb. Other Sinitic languages, like Mandarin, are somewhat in between these two extremes.

In addition, some Sinitic languages require old information to be placed in front of the verb. This causes even more objects to be preposed to a preverbal position. This is strongly the case in the Southeastern Sinitic languages. Having old information in preverbal position is also strongly preferred in the Northern Sinitic languages, Standard Mandarin for instance,⁴⁸ but the requirement is not as strong as in the Southeastern Sinitic languages. At the other extreme are the Far-Southern Sinitic languages, where there is no grammatical requirement for old information to occur pre-verbally. Closely correlating with old information is definiteness. Although it is known that in SVO languages there are correlations between the pre-verbal position and

□□
⁴⁸ Li (2011) characterizes the post-verbal position in Mandarin as new information. Others have characterized the postverbal position in Mandarin as indefinite (Li and Thompson 1974b) or 'focal' (LaPolla 1995).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
 Do not quote or cite this draft.

definiteness, and the post-verbal positions and indefiniteness (Keenan and Comrie 1977), it is rare for the correlation to be as strong as in the Southeastern Sinitic languages, where definite noun phrases (which usually express old information) are grammatically required to appear pre-verbally.

There are three types of constructions that can be used to prepose an object to a pre-verbal position:

- topicalization (the surface order could, grammatically speaking, freely alternate between SOV and OSV)
- passivization (both the undergoer and actor phrases are pre-verbal)
- object marking (OM) construction

The syntax of these construction varies amongst the Sinitic languages. I will discuss briefly the object marking construction first in Section 5.2.3.1. The interaction between old information status and the preverbal position is discussed in Section 5.2.3.2, and word order in clauses with three place predicates is discussed in Section 5.2.3.3. Discussions on topicalization and passivization are interspersed among other discussions in Section 5.2.3.2 and Section 5.2.3.3.

5.2.3.1 The object marking construction

The object marking construction (OM) is also known as the ‘disposal’ construction or pre-transitive construction. The object marker is most commonly grammaticalized from a verb meaning ‘to take’ or ‘to grab hold of’, and the most common syntactic configuration is {subject – OM – object – verb}. (There are other grammatical pathways, and other configurations, see Chappell (2006, in press).) The object marking construction in Mandarin is well discussed (Li and Thompson 1981: Section 15, Sybesma 1992, Ding 2007, lemmolo & Arcodia 2014, amongst many others). In Mandarin, the object marking construction is used primarily to highlight the change of state or change of location of the undergoer. Sometimes an object-marked sentence and its SVO counterpart are both grammatical. Internet search results indicate that with the following two examples, the object-marked construction is more prevalent than the SVO counterpart, but both are frequently used.

(93) Standard Mandarin

關上 門 了
guān- mén le
shàng

20140919 *Draft of:* de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia : The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

close-up door COS
 ‘(Someone) locked the door’
 (“關上門了” on Google: 1,690,000 results; accessed 3 Nov 2012)

- (94) 把門 關上 了
bǎ mén guān- le.
shàng
 OM door close-up COS
 ‘(Someone) locked the door.’
 (“把門關上了” on Google: 1,970,000 results; accessed 3 Nov 2012)

In Mandarin, the *bǎ*-marked object is usually definite, but not necessarily. Old information objects are usually preposed by the object marking construction, or topicalization. An innovation in Mandarin is that the object marking construction can be used with intransitive predicates, in which case the S argument is marked by the ‘object’ marker (see Chappell 2013).

The Far-Southern Sinitic languages require the preposing of objects far less often. The object marking construction is absent in many Far-Southern Sinitic dialects, for instance Chōngzuǒ Píngguà (Lǐ and Zhū 2009: 193, Liáng and Lín 2009: 322) and Nánníng Cantonese (Lín and Qín 2008: 346–348). Some other Far-Southern Sinitic dialects have object marking constructions, but their usage is restricted and infrequent (e.g. Cheung 1992 on Standard Cantonese). In the case of Hainanese, the object marking construction is restricted to inanimates (Lee 2009). (However, they have the non-grammaticalized ‘take’ serial verb construction; see below.) The following is a demonstration of how the OM construction is basically used in Cantonese for sentences comparable to the Mandarin examples above.

- (95) Standard Cantonese
 門咗 門
san⁵⁵- mun¹¹
tsɔ²⁵
 close- door
 PFV
 ‘(Someone) closed the door(s).’ or ‘They (shops etc.) are closed.’
 (Google search of the string “門咗門”: 11,000 results; 3 Nov 2012)

- (96) 門咗 [度/ 道] 門
san⁵⁵- [tɔu²²/ tɔu²²] mun¹¹
tsɔ²⁵

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
 Do not quote or cite this draft.

close- CLF CLF door

PFV

'(Someone) closed the door.'

(Google search of the string "門咗度門": 1,410 results; "門咗道門": 277 results; 3 Nov 2012)

- (97) ? 將 (度/ 道) 門 門
tsæŋ (*tou*²² *tou*²²) *mun*¹¹ *san*⁵⁵
₅₅ /
 OM CLF CLF door close

(Google search of the string "將門門": 0 results; "將度門門": 9 results, "將道門門": 3 results; 3 Nov 2012)⁴⁹

The syntax of the object marking constructions varies greatly amongst the

Sinitic languages. Mandarin dialects towards the northwest (Western Central Plains Mandarin, Northwestern Mandarin) and the Southeastern Sinitic languages in general have fewer constraints with their object marking constructions than Standard Mandarin. For instance, Standard Mandarin and Cantonese do not allow the object marking construction to be used with negative predicates. However, this construction is commonly found in Mandarin spoken towards the northwest.

- (98) Dungan (Western Central Plains Mandarin in Kyrgyzstan/ Kazakhstan)

ба гу кан бу жян ли,

*pa*²⁴ *kou*⁵¹ *kʰæ*⁻⁴⁴ + *pu*²⁴ + *tɕiæ*⁻⁴⁴ *li*

om dog look+ NEG+ achieve COS

'[He] could not see the dog anymore,' (Lín 2003: 312)

- (99) ба та бу кэщин сы ли ма?

*pa*²⁴ *tʰa*⁵¹ *pu*²⁴ *kʰɛ*²⁴ *ɕiŋ*²⁴ *sz*⁵¹ *li* *ma*

OM 3SG NEG happy die COS Q

'Wouldn't it be so unhappy?' (*lit.* 'unhappy to death') (Lín 2003: 313)

□□

49 Using other classifiers like 對 *tøy*³³ and 隻 *tɕk*³ yielded negligible numbers of search results (less than 10).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

(Similar structures exist in Western Central Mandarin dialects in China as well; see, e.g., Bié 2005.)

Similarly, Standard Mandarin and Cantonese do not allow the object marking construction to be used with monosyllabic predicates. However, such constructions are commonly found in the Southeastern Sinitic languages.

(100) Fùyáng Wú

伊 ○ 我 打
hi kʰəʔ ŋɿ tʰæ̃
 3SG OM 1SG hit
 ‘S/he hit me.’

(101) Taiwanese Southern Mǐn

goan² kiaⁿ² ka⁷ goa² chim¹
 1SG:GEN son OM 1SG kiss
 ‘My son kissed me.’
 (Lee 2009: 480)

On the other hand, Hǎinán Mǐn, a Far-Southern Sinitic language, would use a normal SVO sentence in this situation, as the object marking construction cannot be used with animates:

(102) Hǎinán Mǐn, a.k.a. Hainanese

i⁴⁴ soi²¹ gua²¹
 3SG kiss 1SG
 ‘He kissed me.’
 (Lee 2009: 480)

Similar object marking constructions also exist in many Hmong-Mien languages. Unlike Sinitic languages like Mandarin and Cantonese where the object markers are no longer used as lexical verbs, in White Hmong the object marker is synchronically still used as a main verb meaning ‘take’. Nonetheless, as shown in the example below, the protagonist is clearly not physically handling the undergoer marked by *muab* ‘take’, testifying that *muab* ‘take’ has acquired a grammatical function.

(103) White Hmong

nws muab pojniam nrauj lawm
 3SG take woman divorce PRF

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
 Do not quote or cite this draft.

'He has divorced his wife.'

(Jarkey 1991: 249; quoting Heimbach 1979:174)

The object marking construction in most Sinitic languages, including Mandarin and Cantonese, came from the Medieval Chinese 'take' serial verb construction, where the verb 'take' has not yet been grammaticalized. (The grammaticalization of the 'take' verb began when the coreferential pronoun, e.g. the pronoun 之 *zhī* 3SG in the example below, became optional (Peyraube 1996: 169–170).)

(104) Medieval Chinese

船者 乃 將 此 蟾 以 油 熬 之
chuánzhě nǎi jiāng cǐ chán yǐ yóu áo zhī
 boat:person then take this toad with oil fry 3SG
 'Then the boatman took the toad and fried it.'
 (Chappell 2006; quoting Peyraube 1988, 1996)

Similar 'take' serial verb constructions exist in the MSEA languages. The choice between the 'take' and 'non-take' construction in the MSEA languages, including the Far-Southern Sinitic languages, is usually a stylistic choice in how the event is presented, rather than a grammatical preference or requirement as the other Sinitic languages to the north have with their object marking constructions. (The object of 'take' is usually old information, but it is not that old information must occur in a 'take' construction, unlike many non-Far-Southern Sinitic languages where old information is strongly preferred to be expressed preverbally).

(105) Lao

man2 thim5 ngen2
 3 discard money
 'She discarded (the) money.'

(106) *man2 qaw3 ngen2 thim5*

3 take money discard
 'She took the money (and) discarded (it).'

(Enfield 2007: 381)

(107) Vietnamese

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

tôi tặng cho bạn một miếng gà rán
 1 gift DAT friend one CLF chicken fried
 'I gave you a piece of fried chicken.'

- (108) tôi lấy một miếng gà rán tặng cho bạn
 1 take one CLF chicken fried gift DAT friend
 'I took a piece of fried chicken (and) gave it to you.'
 (John Phan p.c.)

The Far-Southern Sinitic languages also often employ the MSEA-type of ungrammaticalized 'take' construction.

- (109) Nánning Píngguà
 'Non-take' construction:
 佢 一 拋 個 隻 煎餅 呢
*kəi*¹³ *ət*³ *pʰau*⁵³ *ə*⁵⁵ *tʃət*³ *tʃin*⁵³ *pən*³³ *nɛ*⁵⁵
 3 once throw DEM CLF pan:cake TOP
 就 跌落 大象 隻 煎鍋
*tʃəu*²² *tʃit*³+*lek*²³ *tai*²² *tʃɛŋ*²² *tʃət*³ *tʃin*⁵³ *ku*⁵³
 then fall+descend elephant CLF frying:pan
 'He [the mouse] threw the pancake, and it fell on the elephant's frying pan.'

- (110) 'Take' construction:
 佢 抓 燒餅 來 一 拋
*kəi*¹³ *ŋa*⁵³ *tʃiu*⁵³ *pən*³³ *ləi*¹¹ *ət*³ *pʰau*⁵³
 3 take pan:cake come once throw
 燒餅 就 跌落 地下
*tʃiu*⁵³ *pən*³³ *tʃəu*²² *tʃit*³+*lek*²³ *təi*²² *ja*²²
 pan:cake then fall+descend ground
 'He [the mouse] took the pancake and threw it, and the pancake fell on the ground.'

5.2.3.2 Preverbal and Postverbal objects

The Southeastern languages strongly require old information to appear before the main verb. New information noun phrases are usually, but not necessarily, placed after the main verb. In the following example, definiteness corresponds with old information and indefiniteness corresponds with new information. It is ungrammatical for the definite object to occur post-verbally, and very strange for the indefinite object to occur pre-verbally.

(111) Fúqīng Eastern Mǐn

老板 買○ 蜀 架 車
lɔ³²peŋ⁵³ mɛ³²⁻ θoŋ² ka²¹ tɕhia⁵³
lau²¹
 boss buy-PFV one CLF car
 'The boss bought a car.'

(112) 許 蜀 架 車 老板 買○
hy³² θoŋ² ka²¹ tɕhia⁵³ lɔ³²peŋ⁵³ mɛ³²⁻
lau²¹

that one CLF car boss buy-PFV
 'The boss bought the car.'

(113) Fùyáng Wú

個 老板 買得 部 車子
kr 'lɔpæ ma-lə bu 'ts^hotsz
 CLF boss buy- CLF car
 PFV
 'The boss bought a (/the) car.'
 (Li and Bisang 2012: 336)

(114) 個 老板 部 車子 買得回來 嘍
kr 'lɔpæ bu 'ts^hotsz ma-lə-'uɛ-lɛ die
 CLF boss CLF car buy-pfv-return-hither COS
 'The boss bought the car.'

(115) 我 去 放 兩 件 衣裳 得 大 衣櫃 裏
ŋɿ tɕhi fã 'niã dzi izã lə da idzy ni
 1SG go put several CLF clothes to big closet inside

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

'I put several items of clothing into the big closet.'

- (116) 我 兩 件 衣裳 去 放 放 得 大 衣櫃 裏
 ŋɿ 'niã dzi izã tɛ'hi fã fã lə da idzy ni
 1SG several CLF clothes go put put to big closet inside
 'I put the several items of clothing into the big closet.'

The following data from Jīxī Huī show that new information need not occur in post-verbal position. Here we have to make a distinction between two different and independent types of givenness (i.e. old information) versus newness (i.e. new information): *referential* givenness/newness, which relates to the old and new information status of objects in the external world or preceding discourse, and *relational* givenness/newness, which relates to the information structure within a sentence (Gundel 1988, 1998). It seems that in Jīxī Huī at least, and perhaps in all Southeastern Sinitic languages, it is *referential* givenness, and not *relational* givenness, that governs the syntactic position of object phrases. As an example, the following three sentences describe three different scenarios of buying a book. In all cases, the book referred to is both specific and definite. The *relational* givenness and newness of a noun phrase is expressed by the optionality versus obligatoriness of a pre-classifier modifier (the demonstrative in this case), respectively. The *referential* givenness and newness of an object noun phrase determines whether it is placed before or after the main verb. In example (117) below, where the book is mentioned in preceding discourse, the book is both relationally and referentially old, and so the demonstrative is optional (relationally old), and it has to be in a preverbal position (referentially old). In example (118) below, where the speaker points at a book, the *relational* newness of the book is indicated by the obligatory demonstrative. However, the book is *referentially* old: it is used to refer to something already known to the speaker, and it is immediately identifiable by the addressee, in the sense that the speaker is pointing to an exemplar of the book that the speaker already owns. Due to the *relational* newness of the object, the demonstrative is obligatory; due to the *referential* givenness of the object, the object is placed pre-verbally. Example (119) where the speaker is telling the shop assistant that s/he intends to buy a book, involves new information in both senses, and hence the demonstrative is obligatory, and the object is post-verbal.

(117) Jīxī Huī

Old information

(爾) 本 書 我 今朝 買 仿

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
 Do not quote or cite this draft.

On the other hand, in the Far-Southern zone, there is no grammatical correlation between the syntactic position of an object noun phrase and givenness/ definiteness. Far-Southern Sinitic languages, Cantonese for instance, readily accept post-verbal definite noun phrases. In fact, it is often strange to prepose an object using an object marking construction. As for the Central Transitional zone, it is transitional between the Far-Southern and the Northern zones in terms of how much they dis-prefer having post-verbal definite noun phrases. The Shàowǔ Mǐn-Gàn examples below are from the Central Transitional zone.

(122) Shàowǔ Mǐn-Gàn

老板 買了 蜀 架 車
lau⁵⁵ pan²¹ miɛ⁵³-ə ɛi⁵ ka³⁵ tɕhja²¹
 boss buy- one CLF car
 PFV

'The boss bought a car.'

(123) 老板 買了 ○ 蜀 架 車

lau⁵⁵ pan²¹ miɛ⁵³-ə tɕioŋ⁵³ ɛi⁵ ka³⁵ tɕhja²¹
 boss buy- this one CLF car
 PFV

'The boss bought this car.'

(124) Very Strange:

?? 老板 拿 ○ 蜀 架 車 買了
 ?? *lau⁵⁵ pan²¹ na²² tɕioŋ⁵³ ɛi⁵ ka³⁵ tɕhja²¹ miɛ⁵³-ə*
 boss OM this one CLF car bought-
 PFV

'The boss bought this car.'

(125) Cantonese

個 老板 買咗 架 車
ko³³ mai¹³- ka³³ ts'hɛ⁵⁵
lou¹³ pan²⁵ tsɔ²⁵
 CLF boss buy- CLF car
 PFV

'The boss bought the/a car.'

(Li and Bisang 2012: 336)

("買咗架車" on Google: 43,900 results; 13 Nov 2012)

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

(126) Very strange:

?? 個 老板 將 架 車 買咗
 ?? kɔ³³ lou¹³ pan²⁵ tsoŋ⁵⁵ ka³³ tsʰɛ⁵⁵ ma¹³-
 tsɔ²⁵
 CLF boss om CLF car buy-PFV

'The boss bought the car.'

('將架車買' on Google: 4 results; accessed 13 Nov 2012)⁵⁰

The Far-Southern Sinitic languages are like the other MSEA languages in not having grammaticalized the correlation between the givenness/definiteness and the syntactic position of an object. Below are examples of definite noun phrases existing in preverbal and postverbal positions in Green Hmong and Ong Be (Kra-Tai).

(127) Green Hmong

khi tug dlev ces tug miv lug
 tie CLF dog [and.then] CLF cat come
 'Tie up the dog and subsequently the cat will come!'
 (Li 1989: 122)

(128) Ong Be

lai³³ vən⁵⁵ hə³³ sai⁵⁵ tsu³³ bian³³ tuan⁵⁵ hu⁵⁵ uk⁵ mia⁵⁵,
 exist day one rich:man release goat CLF out come
 ma¹³ hu⁵⁵ [...] huk³ tuan⁵⁵ hu⁵⁵ dai¹³ voi³³.
 dog CLF make goat CLF die FP

'[There was a rich man who kept a goat...] One day the rich man released the goat, the dog [...] caused the goat to die.'

(Liú 2009: 97)

5.2.3.3 Word order in clauses with three place predicates

MSEA languages in general have fewer instances of double object constructions. For example, Enfield (e.g. 2007: 355–382) argues that there are no real double object constructions in Lao. Some ways to avoid having two unmarked

□□

50 In both Cantonese and Shàowǔ Mìn-Gàn, the 'acquiring' meaning of 'buy' conflicts with the 'disposal' meaning of the object marking construction. Replacing these sentences with 'sell' would make the object marking construction more acceptable.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

'normal' objects after the main verb in Lao are eliding an object, topicalizing an object, putting them in a serial verb construction (e.g. the 'take' serial verb construction), incorporating the patient into the verb, or making one of the objects an oblique object. The main point is that the prohibition is only towards having two unmarked objects after the verb; it is not a prohibition towards having more than one constituent, as having an oblique object after an unmarked object is often an option.

(129) Lao

Noun incorporation (not 'real' double object construction)

laaw2 [thaa2 sii3] hùan2 lang3 nii4

3SG.FAM apply paint house CLF DEM

'She painted (i.e., 'applied paint (to)') this house.'

(Enfield 2007: 357)

(130) * *laaw2 [thaa2 [sii3 lùm5]] hùan2 lang3 nii4*

3SG.FAM apply paint shiny house CLF DEM

(intended meaning: 'She applied shiny paint to this house. '; Enfield 2007: 357)

(131) Topicalization

hùan2 lang3 nii4 laaw2 thaa2 sii3 lùm5

house CLF DEM 3SG.FAM apply paint shiny

'This house, she applied shiny paint (to).'

(Enfield 2007: 358)

(132) *sii3 lùm5 laaw2 thaa2 hùan2 lang3 nii4*

paint shiny 3SG.FAM apply house CLF DEM

'Shiny paint, she applied (to) this house.'

(Enfield 2007: 358)

(133) Serial verb construction

laaw2 qaw3 sii3 lùm5 thaa2 hùan2 lang3 nii4

3SG.FAM take paint shiny apply house CLF DEM

'She took shiny paint (and) applied (it to) this house.'

(Enfield 2007: 358)

(134) Oblique strategy

laaw2 thaa2 hùan2 lang3 nii4 duaj4 sii3 lùm5

3SG.FAM apply house CLF DEM with paint shiny

20140919 *Draft of:* de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

'She applied this house with shiny paint.'

(Enfield 2007: 358)

Khmer also has restrictions towards having two unmarked objects after the verb. The following is an example of this being resolved by a 'take' serial verb construction.

(135) Khmer

* *kɔ̀ət ha:l khaoʔa:v thɲay*
 he expose clothes sun

(136) *kɔ̀ət yɔ:k khaoʔa:v tr̩u ha:l thɲay*

he take clothes go expose sun

'He put the clothes out in the sun.'

(Bisang 2012: 12)

The syntax of three-place constructions varies considerably across Sinitic languages. The Southeastern Sinitic languages have a dis-preference of having two phrases after the verb; one of the objects has to be placed in front of the verb somehow.

(137) Hui'ān Southern Mǐn

Theme topicalized

伊 一 叢 筆 與 我

i¹ tsit^{8>4} tsan^{2>4} pei⁷ k^ho⁵ ua

3SG one CLF pen give 1SG

'S/he gave me a/one pen.'

(This is the most preferred word order; S – V – IO – DO order is also possible, but not often used. The agent is often omitted.)

In fact, Southern Mǐn's preferences for having definite objects in front of the main verb is so strong that the definite object is often expressed twice in front of the main verb: the definite object is topicalized, and then it is (optionally) repeated by a resumptive pronoun supported by an object marker, as shown in the following example.

(138) 我 冊 共 伊 ○ 咧 桌 咧

ua³ ts^he⁷ ka^{5>4} i¹ hio^{5>4} le^{7>8} tɔ⁷ le⁷

1SG book OM 3SG put at table LOC

'I put the book on the table.'

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

The following are examples from another Southeastern Sinitic language.

(139) Fùyáng Wú

* 伊 親得 我 一 口
 * *hi tʰin-lə ŋɿ iəʔ kʰiu*
 3SG kiss- 1SG one CLF_{mouth}
 PFV

(親 'tʰin' 'kiss' is a three-place predicate in Fùyáng Wú)

(140) Passivized

我 撥 伊 親得 一 口
ŋɿ pəʔ hi tʰin-lə iəʔ kʰiu
 1SG PASS 3SG kiss- one CL_{mouth}
 PFV

'I was kissed by him/her once.'

(141) Object marking construction

伊 ○ 我 親得 一 口
hi kʰəʔ ŋɿ tʰin-lə iəʔ kʰiu
 3SG OM 1SG kiss- one CL_{mouth}
 PFV

'S/he kissed me once.'

In Xiānghuà, which is spoken in the Central Transitional zone, the most commonly used ditransitive construction involves a preposition-marked indirect object placed in front of the main verb.

(142) Xiānghuà⁵¹

就 跟 它 放 到 ○ 裏
tʰiəu²⁵ kai⁵⁵ tʰa⁵⁵ frŋ³³ tau³³ pi¹³ la²⁵
 then OM 3SG place to jar in
 '[...] then put it in the jar.'

(143) ○ 跟 我 得 件 衣

zɿ³³ kai⁵⁵ u²⁵ tɿ³³ tʰiə²⁵ i⁵⁵
 3SG DAT 1SG give CLF clothes

□□

51 The word 跟 *kai⁵⁵* has many functions in Xiānghuà, amongst them object marker and dative marker. See, e.g., Chappell, Peyraube, and Wu (2011), Chappell (forthcoming).

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

'He gave me a shirt.'

On the other hand, the Far-Southern Sinitic languages, similar to the non-Sinitic MSEA languages to the south, are relatively freer in having two constituents (of any sort) after the main verb. (Although these Sinitic and non-Sinitic MSEA languages are still not totally free in having two constituents after the main verb, as seen in the case of Lao discussed earlier in this section.)

(144) Cantonese⁵²

阿華 錫咗 我 一 啖
*a*³³*wa*¹¹ *sək*³- *ŋɔ*¹³ *jet*⁵ *tam*²²
*tsɔ*²⁵

Ah.Wah kiss-PFV 1SG one CLF
 'Ah Wah kissed me once.'

(145) 佢 畀咗 啲 藥 我

□□

52 Since Hashimoto (1976), the variation in the order of the (non-topicalized) T and R arguments in double object constructions is often cited as an example of the 'north–south' divide within the Sinitic family. Mandarin has the cross-linguistically more common V R T word order, whereas Cantonese has the cross-linguistically rarer V T R order. The rarer word order in Cantonese is attributed to 'Taicization'; Thai also has (or appears to have) the rarer V T R order.

In reality, the variation in ditransitive constructions amongst Sinitic languages is much more complex than Hashimoto's (1976) generalization. Firstly, Cantonese only uses the V T R order for 'give'-type verbs; other double-object verbs use the V R T word order, e.g. *kau*³³ 'teach', *pun*³³ 'sentence' (i.e. sentence [convict] [penalty]). Secondly, it is not the case that all Southern Sinitic languages use the V T R word order. For instance, Southern Min only uses the V R T word order like Mandarin. Nánning Pínghuà also only uses the V R T word order (although some speakers accept the V T R word order, under the influence of Nánning Cantonese). Thirdly, the V T R word order in Cantonese has probably nothing to do with influences from nearby non-Sinitic languages. In Northern Zhuàng, the V R T construction is the default order, whereas the V T R order can only be used when the R phrase is very short. In Vietnamese, only the V R T construction is used. In fact, the V T R word order in Cantonese is a relatively recent development; the V T R word order is developed from a serial verb construction with the second verb elided: 'give' T 'pass' R > 'give' T R (Chin 2011). In fact, the serial verb 'give-pass' construction is still commonly heard in Cantonese films from the 'black-and-white' era. Somewhat similarly, the V T R 'give' construction in Thai can be thought of as having the R-marking preposition *kĕĕ* omitted: 'give' T (*kĕĕ*) R (see Thepkanjana 2008). See Zhāng (2011) for a very-thorough diachronic and synchronic account of the development of ditransitives amongst Sinitic languages.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

*kʰəy*¹³ *peɪ*²⁵- *ti*⁵⁵ *jœk*³ *ŋɔ*¹³
*tsɔ*²⁵
 3SG give-PFV CLF:MASS medicine 1SG
 ‘S/he gave me the/some medicine.’

- (146) 個 阿婆 收埋咗 嗰 五百萬 喺
*kɔ*³³ *a*³³ *pʰɔ*¹¹ *səu*⁵⁵ *mai*¹¹- *kɔ*²⁵ *ŋ*¹³ *pak*³ *man*²² *hei*²⁵
*tsɔ*²⁵
 CLF o.woman hide-PFV that five.mil at
 櫃桶 底
*kweɪ*²² *tʰɔŋ*²⁵ *teɪ*²⁵
 drawer under
 ‘The old woman hid the five million [units of currency] under the table.’

- (147) Hainanese (Hǎinán Mǐn)
 我 分 蜀 ○ 冊 (至) 伊
*gua*²¹ *bun*⁴⁴ *dziak*³³ *bu*²¹ *seɯ*⁵⁵ *i*⁴⁴
 (ti¹¹)
 1SG give one CLF book to 3SG
 ‘I gave a book to him.’
 (Lee 2011: 502-503)

The following are examples showing other MSEA languages readily allowing two constituents (either bare or full) after the main verb.

- (148) Ong Be
*beu*³³ *juə*³³ *nə*²¹ *hiu*⁵⁵ (*ɟu*³³) *hau*⁵⁵
 deliver clothes that CLF to 1SG
 ‘Pass me that shirt/ Pass that shirt to me.’ (Liú 2009: 35)

- (149) Khmer
əɔj *cee:k* *cru:k* *muaj* *camnuan*
 give banana pig one amount
 ‘[G]ive the pig some bananas.’
 (Haiman 2011: 207)

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
 Do not quote or cite this draft.

- (150) *aoj cee:k muaj camnuan dawl cru:k*
 give banana one bunch towards pig
 '[G]ive a bunch of bananas to the pig.'
 (Haiman 2011: 207)

5.2.4 Summary of word order in clauses

The following table summarizes the clause-level word order traits discussed in this Section 5.2.

Table 4: Left headedness on the clause level (for the most-common construction in each category). (± adpositions: both preposition and postposition. other ±: both orders are prevalent)

	Non Sinitic			FS Sin.				SE Sinitic			C Sinitic		N				
	Thai	Khmer	Vietnamese	E Kayah Li	Burmese	Hmong	Mien	Cantonese	Nánning Pinghuà	Hui'an S Min	Fúqīng E Min	Fúyáng Wú	Jīxī Huī	'Shaowǔ Mǐn-Gàn	Yichūn Gàn	Gǔzhāng Xiānghuà	Suǐníng C Mandarin
VO: bought the car	+	+	+	+	-	+	+	+	+	-	-	-	-	+	+	+	±
VO: bought a car	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+
VO: hit me	+	+	+	+	-	+	+	+	+	±	+	-	+	+	+	+	+
VO(P)O: give me the book	+	+	+	+	-	+	+	+	+	-	+	-	-	+	+	-	-
VO(P)O: give me a book	+	+	+	+	-	+	+	+	+	-	+	+	+	+	+	+	+
VOPO(P): put the book on table	+	+	+	+	-	+	+	+	+	-	-	-	-	-	-	-	-
go destination	+	+	+	+	-	+	+	+	+	+	+	-	-	+	+	-	-
at inside location	+	+	+	±	-	+	±	±	±	±	±	±	±	±	±	±	±
VOX: play ball location	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

	Non Sinitic				FS Sin.				SE Sinitic				C Sinitic				N
go first	+	+	+	+	-	+	+	+	-	-	-	-	- ⁵³	-	+	-	-
eat more	+	+	+	+	+	+	?	+	+	-	-	±	±	-	+	-	-
Total:	11	11	11	10 ^½	1	10	8 ^½	9 ^½	8 ^½	3	5 ^½	3	3 ^½	6 ^½	8 ^½	4 ^½	3 ^½

In Table 4 (and all the feature tables above), the absolute values of the total score have little significance, as the criteria are hand-picked to demonstrate some of the word order differences amongst the Sinitic languages. Nevertheless, the relative scores amongst the Sinitic languages do show the relative difference in left-headedness on the clause level amongst the Sinitic languages. The Far-Southern Sinitic languages (represented by Cantonese and Píngguà here) have relatively more left-headed traits, as they are close to the core of MSEA. The Northern Sinitic languages (represented by Mandarin here) have more right-headed traits, as they are influenced by North Asia. However, the Southeastern Sinitic languages (represented by Southern Mǐn, Eastern Mǐn, Wú and Huī here) are also have many right-headed traits; a proper explanation for this is not yet known to me (see footnote 42).

6 Conclusions and discussion

In this paper I have discussed some of the phonological and word order traits in the Sinitic languages. The Far-Southern Sinitic languages are the most similar to the core of MSEA: highly tonal, conservative with codas, and relatively normal SVO languages. In terms of word order, some left-headed word order traits in the Far-Southern Sinitic languages arise from influence from the core of MSEA. However, not all left-headed traits are influences from the core of MSEA. For instance, not having restrictions on multiple constituents after the main verb in the Far-Southern Sinitic languages could simply mean that they are relatively normal SVO languages, and that they are less influenced by the other Sinitic languages to the north, where this restriction exists. We have also seen that other than the Mandarin dialects that are SOV, the Sinitic languages with the most OV-associated traits are not the Northern Sinitic languages, but the Southeastern Sinitic languages. The strong prevalence of verb-final clauses in the Southeastern Sinitic languages is probably an inter-

□□
53 However, for very old speakers, the word order is 'go first' (+) rather than 'first go' (-).
 20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
 Do not quote or cite this draft.

nal development. It cannot be a direct influence from North Asia, as North Asia is so far away, and the Central and Northern Sinitic languages in between are in general not as strongly right-headed as the Southeastern Sinitic languages.

The summary tables (Tables 2 to 4 above) sometimes show the Far-Southern Sinitic languages as having higher scores of ‘MSEA-ness’ than other Sino-Tibetan languages like Burmese and Southern Mǐn. Their high scores do not indicate that the Far-Southern Sinitic languages are more MSEA-like than these other Sino-Tibetan languages: the scores only indicate that the Far-Southern Sinitic languages have some traits that are more MSEA-like than Burmese and Southern Mǐn. There are many other typological traits, for instance lexical patterns and grammaticalization pathways, which would better illustrate the strong link between the core of MSEA and languages like Burmese and Southern Mǐn (see, e.g., Matisoff 1991, 2001a). What this paper is trying to argue is that, just as there are linguistic criteria which firmly place Burmese in the MSEA linguistic area, there are also many criteria which firmly place the Far-Southern Sinitic languages in the MSEA linguistic area. The Burmish languages and the Far-Southern Sinitic languages are both at the periphery of the MSEA linguistic area, but neither are as ‘fringe’ as, e.g., Mandarin. Some studies on the MSEA linguistic area leave out the languages in China. This is unwise, as the centres of diversity for the Kra-Dai and Hmong-Mien families are still in Southern China, and the Southern Sinitic languages also have many MSEA linguistic traits. Studies of the MSEA linguistic area would benefit immensely if the Southern Sinitic languages, the Far-Southern Sinitic languages in particular, are included in the MSEA linguistic area.

Abbreviations

1	first person	LOC	locative
2	second person	MASS	mass (i.e. part or more than one)
3	third person	MOD	modifier marker
ACC	accusative	N	non-
ANT	anterior	NEG	negative
CLF	classifier	OBJ	object
CONT	continuous	OM	object marker (in obj-marking construction)
COP	copula	PASS	passive

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.
Do not quote or cite this draft.

COS	change of state	PFV	perfective
DAT	dative	PL	plural
DEM	demonstrative	POL	polite
DIM	diminutive	PREP	preposition
EXP	experiential	PRF	perfect
FAM	familiar	PROG	progressive
FOC	focus	PST	past
FP	final particle	Q	question
GEN	genitive	REL	relative clause
IMP	imperative	RLS	realis
IRR	irrealis	SG	singular

Acknowledgements

I would like to thank the team members of the ERC Sinotype project and their language consultants for providing some of the Sinitic data. I would also like to thank John Phan for providing data from Vietnamese, Martha Ratliff and Lisa Ginsburg for providing data from Hmong, and other participants at the 'Mainland Southeast Asian Languages: The State of the Art in 2012' Workshop at the Max Planck Institute for Evolutionary Anthropology, Leipzig for their comments. The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013): ERC Advanced Grant agreement No. 230388: 'The hybrid syntactic typology of Sinitic languages' (2009-2013), and partial funding from the Netherlands Organization for Scientific Research grant 'Human olfaction at the intersection of language, culture and biology'.

References

- Abulimit, Azgöl. 2006. 阿孜古丽阿布力米提 Àzīgūlī Ābùlīmǐtī. *Wéiwúěryǔ Jīchǔ Jiàochéng* 维吾尔语基础教程 [Uyghur Basic Course]. Běijīng: The Central University of Nationalities Press.
- Aldridge, Edith. 2013. Survey of Chinese Historical Syntax Part I: Pre-Archaic and Archaic Chinese. *Language and Linguistic Compass* 7(1). 39–57.
- Bauer, Robert S. 1996. Identifying the Tai Substratum in Cantonese. In *Pan-Asiatic Linguistics. Proceedings of the Fourth International Symposium on Languages and Linguistics V, 1806-1844*. Bangkok: Mahidol University.
- Baxter, William H. 1992. *A Handbook of Old Chinese Phonology*. Berlin: Mouton de Gruyter.
- 20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. *Do not quote or cite this draft.*

- Baxter, William H. & Laurent Sagart. 2014. *Old Chinese: A new reconstruction*. Oxford: Oxford University Press.
- Bié, Mǐn'gē 别敏鸽. 2005. Guānzhōng fāngyán tèshū “bǎ”zìjù tànyuán 关中方言特殊“把”字句探源 [Searching the origins of the special Bǎ construction in Guānzhōng dialect]. *Journal of Huainan Normal University* 2005(1). 127–128.
- Bisang, Walter. 2006. Southeast Asia as a linguistic area. In Keith Brown (Editor in Chief), *Encyclopedia of Languages & Linguistics* (Volume 11, 2nd edition), 587–595. Oxford: Elsevier.
- Bisang, Walter. 2012. Chinese from a typological perspective. Paper presented at the Fourth International Conference on Sinology. Taipei: Academia Sinica.
- Bó, Wénzé 薄文泽. 2004. Càijiāhuà Gàikuàng 蔡家话概况 [Sketches of the Càijiā Language]. *Mínzú Yǔwén* 2004(2). 68–81.
- Cáo, Zhìyún 曹志耘 et al. (eds.). 2008. *Hànyǔ Fāngyán Dìtújí 汉语方言地图集 Linguistic Atlas of the Chinese Dialects*. Běijīng: The Commercial Press.
- Chappell, Hilary. 2006. From Eurocentrism to Sinocentrism: The case of disposal constructions in Sinitic languages. In Felix Ameka, Alan Dench & Nicholas Evans (eds.), *Catching Language: the Standing Challenge of Grammar Writing*. 441–486. Berlin: Mouton de Gruyter.
- Chappell, Hilary. 2012. The role of language contact and hybridization in Sinitic languages. Paper presented at the Fourth International Conference on Sinology. Taipei: Academia Sinica.
- Chappell, Hilary. 2013. Pan-Sinitic object marking: Morphology and syntax. In Guangshun Cao, Hilary Chappell, Redouane Djamouri & Thekla Wiebusch (eds.), *Breaking down the barriers: Interdisciplinary studies in Chinese linguistics and beyond*. (Language and Linguistics Monograph Series 50), 785–816. Taipei: Academia Sinica.
- Chappell, Hilary. forthcoming. *A Grammar of Waxiang, a Sinitic language of Northwestern Hunan*. Berlin: De Gruyter Mouton.
- Chappell, Hilary & Alain Peyraube. 2007. The diachronic syntax of the dative construction from Medieval Chinese to early Southern Min (Sinitic). Talk presented at Conference on Ditransitive Constructions. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Chappell, Hilary, Alain Peyraube & Yunji Wu. 2011. A comitative source for object markers in Sinitic languages: 跟 *kai*⁶⁵ in Waxiang and 共 *kang*⁷ in Southern Min. *Journal of East Asian Linguistics* 20(4). 291–338.
- Chén, Hǎilún 陈海伦 & Cūnhàn Liú 刘村汉. 2009. *Yuèyǔ Pínghuà Tǔhuà Fāngyīn Zìhuì — Dièrbīn: Guībēi Guidōng jí Zhōubiān Pínghuà Tǔhuà Bùfèn 粤语平话土话方音字汇 — 第二编: 桂北桂东及周边平话土话部分* [Dialectal Pronunciation of Characters in Yuè, Pínghuà and Patois — 2nd Volume: Pínghuà and Patois in Northern Guǎngxī, Eastern Guǎngxī and Surrounding Areas]. Shànghǎi: Shanghai Education Publishing House.
- Cheung, Samuel Hung-nin. 1992. The pretransitive in Cantonese. In *中國境內語言暨語言學 Chinese Languages and Linguistics* 1. 241–303.
- Chin, Andy Chi-on. 2011. Grammaticalization of the Cantonese double object verb [pei³⁵] 界 in typological and areal perspectives. *Language and Linguistics* 12(3). 529–563.
- Chirkova, Katia 齊卡佳 Qí Kǎjiā. 2008. Báimǎ Zàngyǔ wéi Dīzú shuō zhīyǐ 白馬藏族為氏族說質疑 The Baima Tibetans and the Di people of Chinese historical records: Challenging the link. *Bulletin of Chinese Linguistics* 3(1). 167–180.
- Comrie, Bernard. 2007. Areal typology of Mainland Southeast Asia: what we learn from the WALs maps [Special issue]. *Manusya* 13. 18–47.
- 20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

- Comrie, Bernard. 2008a. The areal typology of Chinese: between North and Southeast Asia. In Redouane Djamouri, Barbara Meisterernst & Rint Sybesma (eds.), *Chinese Linguistics in Leipzig, Chinese Linguistics in Europe 2*, 1–21. Paris: Centre de recherches linguistiques sur l'Asie Orientale, École des hautes études en sciences sociales.
- Comrie, Bernard. 2008b. Prenominal relative clauses in verb-object languages. *Language and Linguistics* 9(4). 723–733.
- Dahl, Östen. 2008. An exercise in a *posteriori* language sampling. *Sprachtypologie und Universalienforschung* 61(3). 208–220.
- Dede, Keith. 2007. The origin of the anti-ergative [xa] in Huangshui Chinese. *Language and Linguistics* 8(4). 863–881.
- Ding, Picus Sizhi. 2007. Studies on Bǎ Resultative Construction: A Comprehensive Approach to Mandarin Bǎ Sentences (Lincom Studies in Asian Linguistics 62). Munich: Lincom.
- Djamouri, Redouane. 2001. Markers of predication in Shang bone inscriptions. In Hilary Chappell (ed.), *Sinitic Grammar: Synchronic and Diachronic Perspectives*, 143–171. Oxford: Oxford University Press.
- Djamouri, Redouane. forthcoming. Tǎngwàng Language.
- Djamouri, Redouane, Waltraud Paul & John Whitman. 2007. Reconstructing VO constituent order for Proto-Sino-Tibetan. Paper presented at The 18th International Conference on Historical Linguistics. Montréal: Université du Québec à Montréal.
- Dryer, Matthew S. 1991. SVO languages and the OV/VO typology. *Journal of Linguistics* 27. 443–482.
- Dryer, Matthew S. 2001. Mon-Khmer Word Order from a Crosslinguistic Perspective. In Karen L. Adams & Thomas John Hudak (eds.), *Papers from the Sixth Annual Meeting of the Southeast Asian Linguistics Society 1996*, 83–99. Tempe: Arizona State University.
- Dryer, Matthew S. 2013. Relationship between the Order of Object and Verb and the Order of Adjective and Noun. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/97> (accessed 24 March 2014).
- Dryer, Matthew & Martin Haspelmath (eds.). 2011. *The World Atlas of Language Structures Online*. Munich: Max Planck Digital Library. <http://wals.info/>
- Enfield, N. J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34. 181–206.
- Enfield, N. J. 2007. *A Grammar of Lao*. Berlin & New York: Mouton de Gruyter.
- Enfield, N. J. 2011. Linguistic diversity in Mainland Southeast Asia. In N. J. Enfield (ed.), *Dynamics of Human Diversity*, 63–79. Canberra: Pacific Linguistics.
- Ferlus, Michel. 1980. Formation des registres et mutations consonantiques dans les langues Mon-Khmer. *Mon-Khmer Studies* 8. 1–76.
- Ferlus, Michel. 1984. Essai de phonétique historique de môn. *Mon-Khmer Studies* 12. 1–90.
- 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.
- Gawarjon 高尔鞞 Gāo Ērqiāng. 1985. *Tǎjīkěyǔ Jiǎnzhi* 塔吉克语简志 [Sketch of the Tajik Language]. Běijīng: The Ethnic Publishing House.
- Georgopoulos, Carol. 1986. Palauan as a VOS language. In Paul Geraghty, Lois Carrington & Stephen A. Wurm (eds.), *FOCAL I: Papers from the Fourth International Conference 20140919 Draft of*: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton.

Do not quote or cite this draft.

- on *Austronesian Linguistics* (Pacific Linguistics C-93), 187–198. Canberra: Australian National University.
- Gundel, Jeanette K. 1988. Universals of topic-comment structure. In Michael Hammond, Edith A. Moravcsik & Jessica Wirth. *Studies in Syntactic Typology*. Amsterdam; Philadelphia: John Benjamins.
- Gundel, Jeanette K. 1998. Centering theory and the givenness hierarchy: a proposed synthesis. In Marilyn A. Walker, Aravind K. Joshi & Ellen F. Prince (eds.). *Centering Theory in Discourse*. Oxford: Oxford University Press.
- Haiman, John. 2011. *Cambodian: Khmer* (London Oriental and African Language Library 16). Amsterdam, Philadelphia: John Benjamins.
- Hashimoto, Mantaro. 1976. Language diffusion on the Asian continent: problems of typological diversity in Sino-Tibetan. *Computational Analyses of Asian and African Languages* 3. 49–65.
- Hashimoto, Mantarō 橋本萬太郎. 1978. *Gengo Ruikei Chili Ron 言語類型地理論* [Thesis on Geography of Language Typology]. Tōkyō: Koubundou 弘文堂.
- Hashimoto, Mantaro. 1986. The Altaicization of Northern Chinese. In McCoy, John & Timothy Light (eds.), *Contributions to Sino-Tibetan Studies* (Cornell Linguistic Contributions V), 76–97. Leiden: Brill.
- Haspelmath, Martin, Matthew Dryer, David Gil & Bernard Comrie (eds.). 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Haudricourt, André-Georges. 1954. De l'origine des tons en Vietnamien. *Journal Asiatique* 242. 69–82.
- Heimbach, Ernest E. 1979. *White Hmong – English Dictionary* (1979 revised edition) (Linguistic Series IV, Data Paper 75). Ithaca & New York: Cornell University.
- Iemmolo, Giorgio & Giorgio Francesco Arcodia. 2014. Differential object marking and identifiability of the referent: a study of Mandarin Chinese. *Linguistics* 52(2): 315–334. doi 10.1515/ling-2013-0064.
- Jacob, Judith M. 1993. An examination of the vowels and final consonants in correspondences between Pre-Angkor and Modern Khmer. In David A. Smyth (ed.), *Cambodian Linguistics, Literature and History*, 87–102. London: University of London.
- Janhunen, Juha, Marja Peltomaa, Erika Sandman & Xiawu Dongzhou. 2008. *Wutun*. Munich: Lincom.
- Jarkey, Nerida. 1991. Serial verbs in White Hmong: a functional approach. Sydney: University of Sydney dissertation.
- Kato, Atsuhiko. 2003. Pwo Karen. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan Languages*, 632–648. London & New York: Routledge.
- Keenan Edward L. & Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8. 63–99.
- Koul, Omkar N. & Kashi Wali. 2006. *Modern Kashmiri Grammar*. Springfield, VA: Dunwoody Press.
- Lee, Yeon-Ju & Sagart, Laurent. 2008. No limits to borrowing: the case of Bai and Chinese. *Diachronica* 25(3). 357–389.
- LaPolla, Randy J. 1994. On the change to verb-medial order in Proto-Chinese: evidence from Tibeto-Burman. In Hajime Kitamura, Tatsuo Nishida & Yasuhiko Nagano (eds.), *Current Issues in Sino-Tibetan Linguistics*, 98–104. Ōsaka: Organizing Committee of the 26th International Conference on Sino-Tibetan Languages and Linguistics.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

- LaPolla, Randy J. 1995. Pragmatic relations and word order in Chinese. In Pamela Downing & Michael Noonan (eds.), *Word Order in Discourse*, 297–329. Amsterdam & Philadelphia: John Benjamins.
- Lee, Hui-chi. 2009. On the object marker *BUE* in Hainan Min. *Language and Linguistics* 10(3). 471–487.
- Lee, Hui-chi. 2011. Double object construction in Hainan Min. *Language and Linguistics* 12(3). 501–527.
- Lee, Yoen-Ju & Laurent Sagart. 2008. No limits to borrowing: the case of Bai and Chinese. *Diachronica* 25(3). 357–385.
- Li, Chao. 2011. Postverbal constituents in Mandarin Chinese. In Zhuo Jing-Schmidt (ed.), *Proceedings of the 23rd North American Conference on Chinese Linguistics* (Volume 2), 30–37. Eugene: University of Oregon.
- Li, Charles N. 1989. The origin and function of switch reference in Green Hmong. In Leiv Egil Breivik & Ernst Håkon Jahr (eds.), *Language Change: Contributions to the study of its causes*. (Trends in Linguistics Studies and Monographs 43), 115–129. Berlin & New York: Mouton de Gruyter.
- Li, Charles N. & Sandy Thompson. 1974. An explanation of word order change SVO → SOV. *Foundations of Language* 12. 201–214.
- Li, Charles N. & Sandy Thompson. 1975. The semantic function of word order: a case study in Mandarin. In Charles N. Li (ed.), *Word Order and Word Order Change*, 164–195. Austin: University of Texas Press.
- Li, Charles N. & Sandy Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, Los Angeles & London: University of California Press.
- Lǐ, Jīnfāng 李锦芳. 1999. *Bùyāngyǔ Yánjiū* 布央语研究 [Studies of the Bùyāng Language]. Běijīng: The Central University of Nationalities Press.
- Lǐ, Jīnfāng 李锦芳. 2002. *Dòng-Tái Yǔyán yǔ Wénhuà* 侗台语言与文化 [Kam-Tai Language and Culture]. Běijīng: The Ethnic Publishing House.
- Lǐ, Lán 李蓝. 2004. *Húnán Chéngbù Qīngyī Miáo Rénhuà* 湖南城步青衣苗人话 [The People's Speech of Blue Dress Miáo in Chéngbù, Húnán]. Běijīng: China Social Sciences Press.
- Lǐ, Liánjìn 李连进 & Yàn'è Zhū 朱艳娥. 2009. *Guǎngxī Chóngzuǒ Jiāngzhōu Zhèyuánhuà Bǐjiào Yánjiū* 广西崇左江州蔗园话比较研究 [Comparative Studies of the Sugarcane Field Dialect in Jiāngzhōu, Chóngzhō, Guǎngxī]. Guilín: Guangxi Normal University Press.
- Lǐ, Yǒngsui 李永燧. 2010. Miǎn-Yíyǔ: yīzhǒng shēngdiào zúyǔ 缅彝语：一种声调语言 [Lolo-Burmese: a type of tonal proto language]. In Zhāomíng Dài 戴昭铭 & Mǎtísǒfū 马提索夫 (James Matisoff) (eds.), *Hànzàngyǔ Yánjiū Sishínián: Dì Sishíjiè Guójì Hànzàng Yǔyán jì Yǔyánxué Huìyì Lùnwénjí* 第40届国际汉藏语言暨语言学会议论文集 [Forty years of Sino-Tibetan Research: Proceedings of the 40th International Conference on Sino-Tibetan Languages and Linguistics], 47–57. Harbin: Heilongjiang University Press.
- Lǐ, Xiǎofán 李小凡. 1998. *Sūzhōu Fāngyán Yǔfǎ Yánjiū* 苏州方言语法研究 [Research on the Syntax of the Sūzhōu Dialect]. Běijīng: Peking University Press.
- Li, Xuping & Walter Bisang. 2012. Classifiers in Sinitic languages: from individuation to definiteness marking. *Lingua* 112. 335–355.
- Lǐ, Yúnbīng 李雲兵. 2003. Miáo-Yáoyǔ shēngdiào wèntí 苗瑶语声调问题 [The question of tones in Miáo-Yáo languages]. *Language and Linguistics* 4(4). 683–712.
- Liáng, Mǐn 梁敏. 2002. *Biāohuà Yánjiū* 标话研究 [Studies on the Biāo Language]. Běijīng: The Central University of Nationalities Press.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

- Liáng, Wěihuá 梁伟华 & Yi Lín 林亦. 2009. *Guǎngxī Chóngzuǒ Xīnhé Zhèyuánhuà Yánjiū* 广西崇左新和蔗园话研究 [Studies of the Sugarcane Field Dialect in Xīnhé, Chóngzhǒ, Guǎngxī]. Guilín: Guangxi Normal University Press.
- Lín, Tāo 林涛. 2003. *Zhōngyà Dōnggānyǔ Yánjiū* 中亚东干语研究 *Жуня Дунганйү Янжю* [Studies of the Dungan Language of Central Asia]. Hong Kong: Hong Kong Educational Press.
- Lín, Yi 林亦 & Fèngyǔ Qín 覃凤余. 2008. *Guǎngxī Nánning Báihuà Yánjiū* 广西南宁白话研究 [Studies of Nánning Cantonese of Guǎngxī]. Guilín: Guangxi Normal University Press.
- Liú, Dānqīng 刘丹青. 2003. *Yǔxù Lèixíngxué yǔ Jiècí Lǐlùn* 语序类型学与介词理论 [Typology of Word Order and Theories on Adpositions]. Běijīng: The Commercial Press.
- Liú, Dānqīng 刘丹青. 2012. *Gǔjīn Hànyǔ de jùfǎ lèixíng yǎnbiàn: kuā-fāngyán de kùcáng lèixíngxué shìjiǎo* 古今汉语的句法类型演变：跨方言的库藏类型学视角 [Changes in the syntactic typology in diachronic and synchronic Chinese: viewpoints from pan-dialectal inventory typology]. Paper presented at the Fourth International Conference on Sinology. Taipei: Academia Sinica.
- Liú, Jiànsān 刘剑三. 2009. *Lín'gāoyǔ Huàyǔ Cáiliàojí* 临高语语料集 [Corpus Material Collection of the Lín'gāo Language]. Běijīng: The Central University of Nationalities Press.
- Liú, Lǐlǐ 刘俐李. 1993. *Xīnjiāng Hànyǔ fāngyán de xíngchéng* 新疆汉语方言的形成 [The formation of Chinese dialects in Xīnjiāng]. *Fāngyán* 1993(4). 265–274.
- Lo, Seo-Gim 羅肇錦. 1988. *Kèyǔ Yǔfǎ* 客語語法 [Grammar of the Hakka Language]. Taipei: Taiwan Student Book Co.
- Löffler, Lorenz G. 1966. The contribution of Mru to Sino-Tibetan linguistics. *Zeitschrift der Morgenländischen Gesellschaft* 116(1). 118–159.
- Ian Maddieson. 2013a. Voicing and Gaps in Plosive Systems. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/5> (accessed 31 March 2014).
- Ian Maddieson. 2013b. Front Rounded Vowels. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/11> (accessed 31 March 2014).
- Máo, Zōngwǔ 毛宗武 & Cháo jí Méng 蒙朝吉. 1986. *Shēyǔ Jiǎnzhi* 畲语简志 [Sketch of the Shē Language]. Běijīng: The Ethnic Publishing House.
- Máo, Zōngwǔ 毛宗武, Cháo jí Méng 蒙朝吉 & Zōngzé Zhèng 郑宗泽. 1982. *Yáo zú Yǔyán Jiǎnzhi* 瑶族语言简志 [Sketch of the Language of the Yáo Nationality]. Běijīng: The Ethnic Publishing House.
- Matisoff, James. 1991. Areal and universal dimensions of grammaticalization in Lahu. In Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to Grammaticalization* (Volume 2), 383–453. Amsterdam, Philadelphia: John Benjamins.
- Matisoff, James. 2001a. Genetic vs. contact relationship: prosodic diffusibility in South-East Asian languages. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*, 291–327. Oxford: Oxford University Press.
- Matisoff, James. 2001b. On the genetic position of Bai within Tibeto-Burman. Paper presented at the 34th International Conference on Sino-Tibetan Languages and Linguistics. Kūnmíng: Yunnan Nationalities Institute.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

- Matisoff, James. 2003. Southeast Asian languages. In William Frawley & Bernard Comrie (eds.), *International Encyclopedia of Linguistics* (2nd edition, Volume IV), 126–130. New York & Oxford: Oxford University Press.
- Matthews, Stephen & Virginia Yip. 2011. *Cantonese — A Comprehensive Grammar* (2nd edition). Abingdon[on-Thames] & New York: Routledge.
- Nguyễn, Đình-Hoà. 1997. *Vietnamese: Tiếng Việt Không Sơn Phần* (London Oriental and African Language Library 9). Amsterdam; Philadelphia: John Benjamins.
- Norman, Jerry. 1988. *Chinese*. Cambridge: Cambridge University Press.
- Norman, Jerry. 2003. The Chinese dialects: phonology. In Graham Thurgood & Randy J. LaPolla (eds.), *The Sino-Tibetan Languages*. (Routledge Language Family Series), 72–83. London & New York: Routledge.
- Norman, Jerry & Tsu-lin Mei. 1976. The Austroasiatics in Ancient South China: some lexical evidence. *Monumenta Serica* 32. 274–301.
- Ostapirat, Weera. 2000. Proto-Kra. *Linguistics of the Tibeto-Burman Area* 23(1). 1–251.
- Ostapirat, Weera. 2005. Kra-Dai and Austronesian: notes on phonological correspondences and vocabulary distribution. In Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds.), *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, 107–131. London & New York: Routledge-Curzon.
- Ōuyáng, Juéyà 欧阳觉亚. 1989. Hànyǔ Yuèfāngyán lǐ de Gǔ Yuèyǔ chéngfèn 汉语粤方言里的古越语成分 [Elements of the Ancient Yuè Language in the Yuè Dialect of Chinese]. In Shūxiāng Lǚ 吕叔湘 (ed.), *Yǔyán Wénzì Xuéshù Lùnwénjí — Qìngzhù Wáng Lì Xiānshēng Xuéshù Huódòng Wǔshí Zhōunián* 语言文字学术论文集 — 庆祝王力先生学术活动五十周年 [Academic Proceedings of Language and Writing — Celebrating the Fiftieth Anniversary of Professor Wáng Lì's Academic Activities]. Běijīng: Knowledge Publishing House.
- Peterson, David A. 2005. Initial observations on Bangladesh Mru morphosyntax. Paper presented at the 38th International Conference on Sino-Tibetan Language and Linguistics. Xiàmén: Xiamen University.
- Peyraube, Alain. 1988. *Syntaxe Diachronique du Chinois: Évolution des Constructions Datives du XIVe Siècle av. J.-C. au XVIIIe Siècle*. Paris: Collège de France, Institut des Hautes Études Chinoises.
- Peyraube, Alain. 1996. Le Cantonais est-il du chinois? *Perspectives Chinoises* 34. 26–29.
- Peyraube, Alain. 1997. On word order in Archaic Chinese. *Cahiers de Linguistique - Asie Orientale* 26(1). 3–20.
- Pittayaporn, Pittayawat. 2009. The phonology of Proto-Tai. PhD dissertation. Cornell University.
- Pulleyblank, Edwin George. 1963. The consonantal system of Old Chinese (Part II). *Asia Major* 9. 206–265.
- Qián, Méng 钱萌. 2008. Níngbō fāngyán zhōng de SOV jù 宁波方言中的SOV句 [SOV sentences in Níngbō dialect]. In Linguistic Association of Shanghai and Chinese Linguistic Association of Hong Kong (eds.), *Wúyǔ Yánjiū: Disìjì Guójì Wúfāngyán Xuéshù Yántǎohuì Lùnwénjí* 吴语研究：第四届国际吴方言学术研讨会论文集 [Studies of Wú Language: Proceedings of the Fourth International Conference on Wú Dialects], 134–140. Shànghǎi: Shanghai Education Publishing House.
- Qián, Nǎiróng 钱乃荣. 2008. Níngbō fāngyán de shítài 宁波方言的时态 [Tense and aspect in Níngbō dialect]. In Linguistic Association of Shanghai and Chinese Linguistic Association of Hong Kong (eds.), *Wúyǔ Yánjiū: Disìjì Guójì Wúfāngyán Xuéshù Yántǎohuì*
- 20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

- Lùnwénjī* 吴语研究：第四届国际吴方言学术研讨会论文集 [Studies of Wú Language: Proceedings of the Fourth International Conference on Wú Dialects], 121–126. Shànghǎi: Shanghai Education Publishing House.
- Ratliff, Martha. 1998. Ho Ne (She) is Hmongic: one final argument. *Linguistics of the Tibeto-Burman Area* 21(2). 97–109.
- Ratliff, Martha. 2010. *Hmong-Mien Language History* (Pacific Linguistics 613). Canberra: Australian National University.
- Sagart, Laurent. 2002. Gan, Hakka and the formation of Chinese dialects. In Dah-an Ho (ed.), *Dialect Variations in Chinese, Papers from the Third International Conference on Sinology, Linguistics Section*, 129–153. Taipei: Academia Sinica.
- Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2). 411–444.
- Sagart, Laurent. 2008. The expansion of *Setaria* farmers in East Asia: a linguistic and archaeological model. In Alicia Sanchez-Mazas, Roger Blench, Malcolm D. Ross, Iliia Peiros & Marie Lin (eds.), *Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics*, 133–157. London & New York: Routledge.
- Sagart, Laurent. 2011. Classifying Chinese dialects/Sinitic languages on shared innovations. Talk given at Centre de recherches linguistiques sur l'Asie orientale, Norgent sur Marne.
- Shěn, Míng 沈明. 2006. Jīnyǔ de fēnqū (gǎo) 晋语的分区 (稿) [The classification of Jin (draft)]. *Fāngyán* 2006(4). 343–356.
- Simpson, Andrew. 2005. Classifiers and DP Structure in Southeast Asia. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford Handbook of Comparative Syntax*, 806–838. Oxford & New York: Oxford University Press.
- Smyth, David. 2002. *Thai: An Essential Grammar*. London & New York: Routledge.
- Soe, Myint. 1999. A grammar of Burmese. Eugene, OR: University of Oregon dissertation.
- Solnit, David B. 1997. *Eastern Kayah Li: Grammar, Texts, Glossary*. Honolulu: University of Hawai'i Press.
- Sousa, Hilário de. forthcoming. *A Grammar of Nanning Pinghua*.
- Starostin, Sergej. 1995. The historical position of Bai. *Moskovskiy Lingvisticheskiy Zhurnal* *Московский Лингвистический Журнал* [Moscow Journal of Linguistics] 1. 174–190.
- Sun, Hongkai. 2003. Is Baima a dialect or vernacular of Tibetan. *Cahiers de Linguistique – Asie Orientale* 32(1). 61–81.
- Sybesma, Rint P. E. 1992. Causatives and accomplishments: the case of Chinese ba. Leiden: Leiden University dissertation.
- Thepkanjana, Kingkarn. 2008. Ditransitive Constructions in Thai. In Andrej Malchukov, Martin Haspelmath & Bernard Comrie (eds.), *Studies in Ditransitive Constructions: A Comparative Handbook*, 409–426. Berlin & New York: De Gruyter Mouton.
- Thurgood, Graham. 1993. Phan Rang Cham and Utsat: tonogenetic themes and variants. In Jerold A. Edmondson & Kenneth J. Gregerson (eds.), *Tonality in Austronesian Languages* (Oceanic Linguistics Special Publication 24), 91–106. Honolulu: University of Hawai'i Press.
- Vovin, Alexander. 2000. Did the Xiongnu speak a Yeniseian language? *Central Asiatic Journal* 44(1). 87–104.
- Wali, Kashi & Omkar N. Koul. 1996. *Kashmiri: A Cognitive-Descriptive Grammar*. London, New York: Routledge.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

- Wang, Feng. 2006. *Comparison of languages in contact: the distillation method and the case of Bai* (Language and Linguistics Monograph Series B: Frontiers in Linguistics III). Taipei: Institute of Linguistics, Academia Sinica.
- Wāng, Fēng 汪鋒. 2012. *Yǔyán Jiēchū yǔ Yǔyán Bǐjiào: yǐ Báiyǔ wéi Lì* 语言接触与语言比较:以白语为例 [Language Contact and Language Comparison: using Bai as an example]. Běijīng: The Commercial Press.
- Wáng, Fùshì 王辅世. 1985. *Miáoyǔ Jiǎnzhi* 苗语简志 [Sketch of the Miao Language]. Běijīng: The Ethnic Publishing House.
- Wáng, Hóngjūn 王洪君. 2009. Jiāngù yǎnbiàn, tuīpíng hé céngcì de Hànyǔ fāngyán lìshǐ guānxì móxíng 兼顾演变、推平和层次的汉语方言历史关系模型 [A historical relationship model of Chinese dialects which encompasses change, leveling and stratification]. *Fāngyán* 2009 (4). 204–218.
- Wayland, Ratre P. & Allard Jongman. 2002. Registrogenesis in Khmer: a phonetic account. *Mon-Khmer Studies* 32: 101–115.
- Wayland, Ratre P. & Allard Jongman. 2003. Acoustic correlates of breathy and clear vowels: the case of Khmer. *Journal of Phonetics* 31. 181–201.
- Wéi, Jǐngyún 韦景云 & Xiǎoháng Qín 覃晓航. 2006. *Zhuàngyǔ Tōnglùn* 壮语通论 [Common Theories on the Zhuang Language]. Běijīng: The Central University of Nationalities Press.
- Wéi, Qīngwěn 韦庆稳. 1981. Yuèréngé yǔ Zhuàngyǔ de guānxì chūtàn 越人歌与壮语的关系初探 [First discussions on the relationship between Yuèréngé and the Zhuang language]. In Editorial board of Mínzú Yǔwén 民族语文编辑部 (eds.), *Mínzú Yǔwén Lùnjí* 民族语文论集 [Essays in Ethnic Languages]. Běijīng: China Social Sciences 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, Wwei Huang, Ranjan Deka, Bing Su, Ranajit Chakraborty & Li Jin. 2004. Genetic Evidence Supports Demic Diffusion of Han Culture. *Nature* 431: 302–305.
- Wǔ, Yúnjī 伍云姬 & Ruiqīng Shěn 沈瑞清. 2010. *Xiāngxī Gǔzhàng Wǎxiānghuà Diàochá Bàogào* 瓦乡话调查报告 [Survey Report on the Wǎxiāng Language in Gǔzhàng, Western Húnán]. Shànghǎi: Shanghai Educational Publishing House.
- Wurm, Stephan Adolphe, Rong Li, Theo Baumann & Mei W. Lee (eds.), 1987. *Language Atlas of China*. Hong Kong: Longman.
- Xióng, Zhèngguāng 熊正辉 & Zhèn-xīng Zhāng 张振兴. 2008. Hànyǔ fāngyán de fēnqū 汉语方言的分区 [Classification of the Chinese dialects]. *Fāngyán* 2008(2). 97–108.
- Yip, Virginia & Stephen Matthews. 2007. Relative clauses in Cantonese-English bilingual children: typological challenges and processing motivations. *Studies in Second Language Acquisition* 29(2). 277–300.
- Yóu, Wénliáng 游文良. 2002. *Shèzú Yǔyán* 畬族语言 [The language of the Shē Nationality]. Fúzhōu: Fujian People's Publishing House.
- Zhāng, Mǐn 張敏. 2010. “Dòng hòu xiànzhì” de qūyù tuīyí jí qí shízhí “動後限制”的區域推移及其實質 The postverbal constraint as a geographical continuum. Talk presented at the International Symposium for Comparative and Typological Research on Languages of China, The Hong Kong University of Science and Technology.
- Zhāng, Mǐn 張敏. 2011. Hànyǔ fāngyán shuāngjīwù jiégòu nánběi chāyí de chéngyīn: lèixíngxué yánjiū yǐnfā de xīn wèntí 漢語方言雙及物結構南北差異的成因：類型學研究引發的新問題 Revisiting the alignment
- 20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.

- typology of ditransitive constructions in Chinese dialects. *Bulletin of Chinese Linguistics* 4(2). 87–270.
- Zhāng, Zhènxīng 张振兴 & Zhèngguāng Xióng 熊正辉 et al. (eds.). in press. *Zhōngguó Yǔyán Dìtújí* 中国语言地图集 [Atlas of Languages of China], new edition. Běijīng: The Commercial Press.
- Zhèng, Yíqīng 郑贻青. 1997. *Huíhuīhuà* 回辉话 [The Huíhuí Language]. Shànghǎi: Shanghai Far East Publishers.
- Zhèngzhāng, Shāngfāng. 1991. Decipherment of Yue-Ren-Ge (song of the Yue boatman). *Cahiers de Linguistique — Asie Orientale* 20(2). 159–168.
- Zhèngzhāng, Shāngfāng 郑张尚芳. 1995. Gàn, Mǐn, Yuèyǔ lǐ gǔquánzhuó shēngmǔ jīndú zhuóyīn de fāngyán 赣、闽、粤语里古全浊声母今读浊音的方言 [The Gàn, Mǐn and Yuè dialects where Middle Chinese voiced obstruent onsets remained voiced]. In Tsu-Lin Mei 梅祖麟 (ed.), *Wúyǔ hé Mínyǔ de Bǐjiào Yánjiū* 吴语和闽语的比较研究 *Comparative Studies among Wu and Min Dialects*: 13–17. Shànghǎi: Shanghai Education Publishing House.
- Zhèngzhāng, Shāngfāng 郑张尚芳. 1998. Gǔ Yuèyǔ zhāng 古越语章 [Section on the Ancient Yuè language]. In Dǒng, Chǔpíng 董楚平 & Yǒngpíng Jīn 金永平 (eds.), *Wǔ–Yuè Wénhuà Tōngzhì* 吴越文化通志 [Annals of Wǔ and Yuè Cultures]. Shànghǎi: Shanghai People Publishing House.
- Zhèngzhāng, Shāngfāng 郑张尚芳. 1999. Bǎiyǔ shì Hàn–Bái Yǔzú de yīzhī dúlì yǔyán 白语是汉白语族的一支独立语言 [Bái is an independent language in the Sinitic–Bái language family]. In Shí, Fēng 石锋 & Wúyún Pān 潘悟云 (eds.), *Zhōngguó Yǔyánxué de Xīn Tuōzhǎn* 中国语言学的新拓展 [New Expansions in Linguistics in China], 19–73. Hong Kong: City University of Hong Kong Press.
- Zhèngzhāng, Shāngfāng 郑张尚芳. 2008. *Wēnzhōu Fāngyán Zhì* 温州方言志 [Records of the Wēnzhōu Dialect]. Běijīng: Zhonghua Book Company.
- Zhèngzhāng, Shāngfāng 郑张尚芳. 2010. Càijiāhuà Bǎiyǔ kuānxi jí cígēn bǐjiào 蔡家话白语关系及词根比较 [The relationship between Càijiā language and Bái language, and comparison of word roots]. In Wúyún Pān 潘悟云 & Zhōngwēi Shěn 沈钟伟 (eds.), *Yánjiū zhī Lè* 研究之乐 *The Joy of Research II*: 389–400. Shànghǎi: Shanghai Educational Publishing House.
- Zhōng, Jīnwén 钟进文. 2007. Gān–Qīng Dìqū Tèyǒu Mínzú Yǔyán Wénhuà de Qūyù Tèzhēng 甘青地区特有民族语言文化的区域特征 Area Features of Minorities Language and Culture Unique to Gansu and Qinghai Provinces. Běijīng: The Central University of Nationalities Press.
- Zhōu, Zhèng 周政. 2009. Pínglì Fāngyán Diàochá Yánjiū 平利方言调查研究 [Investigative Study on the Pínglì Dialect]. Běijīng: Zhonghua Book Company.
- Zhu, Xiaonong. 2006a. A Grammar of Shanghai Wu. Munich: Lincom Europa.
- Zhū, Xiǎonóng 朱晓农. 2006b. Nèibào yīn 内爆音 On Implosives. *Fāngyán* 2006(1). 16–21.
- Zhū, Xiǎonóng 朱晓农, Zémín Liú 刘泽民 & Fùqióng Xú 徐馥琼. 2009. Zìfā xīnshēng de nèibào yīn — láizì Gànyǔ, Mínyǔ, Hānyǔ, Wúyǔ de dìyīshǒu cáiliào 自发生新的内爆音 — 来自赣语、闽语、哈尼语、吴语的第一手材料 [Recently developed implosives: new findings from Gàn, Mǐn, Hāni and Wú]. *Fāngyán* 2009(1). 10–17.
- Zhū, Zhāngnián 朱彰年, Gōngmù Xuē 薛恭穆, Wéihuī Wāng 汪维辉, Zhìfēng Zhōu 周志锋 (eds.), 1996. *Níngbō Fāngyán Cídiǎn* 宁波方言词典 [Dictionary of the Níngbō Dialect]. Shànghǎi: Chinese Dictionary Publishing House.

20140919 Draft of: de Sousa, Hilário. 2015. The Far Southern Sinitic Languages as part of Mainland Southeast Asia. In Enfield, N.J. & Comrie, Bernard (eds.), *Languages of Mainland Southeast Asia: The state of the art* (Pacific Linguistics 649), 356–439. Berlin: De Gruyter Mouton. Do not quote or cite this draft.