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

- Lexical tonal and/or phonational contrasts (except that most Khmer dialects lost their phonational contrasts; languages which are primarily tonal often have five or more tonemes);
- Analytic morphological profile with many sesquisyllabic or monosyllabic words;
- Strong left-headedness, including prepositions and SVO word order.

The Sino-Tibetan languages, like Burmese and Mandarin, are said to be periphery to the MSEA linguistic area. The periphery languages have fewer traits that are typical to MSEA. For instance, Burmese is SOV and right-headed in general, but it has some left-headed traits like post-nominal adjectives (‘stative verbs’) and numerals. Mandarin is SVO and has prepositions, but it is otherwise strongly right-headed. These two languages also have fewer lexical tones.

This paper aims at discussing some of the phonological and word order typological traits amongst the Sinitic languages, and comparing them with the MSEA typological canon. While none of the Sinitic languages could be considered to be in the core of the MSEA language area, the Far Southern Sinitic languages, namely Yuè, Pínghuà, the Sinitic dialects of Hāinán and Léizhōu, and perhaps also Hakka in Guǎngdōng (largely corresponding to Chappell (2012, in press)’s ‘Southern Zone’) are less ‘fringe’ than the other Sinitic languages from the point of view of the MSEA linguistic area. Studies on the MSEA linguistic area would benefit from considering the Far Southern Sinitic languages (and perhaps also some neighbouring Sinitic varieties like Southern Mǐn) as part of the MSEA linguistic area.

The rest of the paper is structured as follow. In section 1, we will present a brief overview of the Sinitic languages, primarily on their history and the genealogical relationships within and beyond the Sinitic language family. In section 2, we will 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 3, we will discuss some of the MSEA-like phonological traits in the Sinitic languages. In section 4, we will discuss the variation in word order amongst the Sinitic languages. A conclusion will be presented in Section 5.

1. The Sinitic languages

The Sinitic languages are the descendents of the historical Chinese language. The periodisation of the Chinese language differs amongst linguists, with historical syntactician often favouring terms like ‘Archaic Chinese’ and ‘Medival Chinese’, while historical phonologists often favouring terms like ‘Old Chinese’ and ‘Middle Chinese’. The earliest stage of the Chinese language with written record is Pre-Archaic Chinese, which is represented by the Shāng Dynasty oracle bone script (fourteenth to eleventh century BCE). The earliest phonologically reconstructable form of Chinese is Old Chinese, which was 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 with that of the Romance languages (e.g. Norman 2003: 82). Around the same time that Vulgar Latin was spread by Roman conquests, Common Chinese was spread by the expansions of the Qin and Han Empires (221 BCE – 220 CE). Based on lexical and phonological innovations, Sagart (2011) dates the most recent common ancestor of the modern Sinitic languages to about third or second century BCE, with Wàixiāng 瓦鄉 being the earliest branch. The Sinitic languages are often called ‘Chinese dialects’. The term ‘dialect’ is a translation of the Chinese term 方言 (Mandarin fāngyán), which literally means ‘regional speech’. The Chinese term fāngyán is semantically wider than the Western notion of ‘dialect’, and readily includes what in the West would be considered separate languages of the same language family.

The Language Atlas of China (Zhāng et al. in press; Würm & Lǐ et al. 1987) classifies the Sinitic languages into ten main dialect groups, based primarily on phonological criteria. Each dialect group includes a number of dialects that are not mutually intelligible with each other. The ten main dialects groups are (Xióng and Zhāng 2008):

- Jin 晉;
- Mandarin 官話;

---

1 Historical syntacticians and phonologists of Archaic/Old Chinese deal with morphology in different ways. Historical syntacticians tend to only look at the syntax and morphology of the strings of characters in texts. Historical phonologists of Old Chinese often look into the linguistic properties of the many single-consonant affixes that probably existed in Old Chinese, but are not necessarily indicated in the writing system, e.g. *gʰɑŋ ‘king’, *gʰɑŋ-s ‘be king’ (Baxter and Sagart n.d.).

2 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.
- Wú 吳;
- Huī 徽;
- Gàn 赣;
- Xiāng 湘;
- Mín 民;
- Hakka (or Kèjiā 客家);
- Yuè 粤; and
- Pínghuà 平話.

There are also other smaller Sinitic varieties which fall outside this ten-group classification:
- The patois (tūhuà 土話) of Southern Húnán (Xiāngnán Tūhuà 湘南土話), Northern Guǎngdōng (Yuèběi Tūhuà 粵北土話) and Eastern Guāngxī (Guìdōng Tūhuà 桂東土話);³
- The Dānzhōu 儋州 dialect in Northeastern Hǎinán (somewhat Yuè-like, with influence from other Sinitic and non-Sinitic languages in lowland Hǎinán);
- The Wǎxiāng 瓦鄉 dialect in western Húnán (e.g. Wǔ and Shēn 2010, Chappell forthcoming);
- The Sinitic first language of Blue Dress Miáo people in Southwestern Húnán and neighbouring Northern Guāngxī (Qīngyì Miáo Rénhuà 青衣苗人話; Lǐ 2004); and
- The Sinitic first language of the Shē 畬 people (Yóu 2002) (somewhat Hakka-like).

Map

Externally, the Sinitic language family is a member of the larger Sino-Tibetan language family. There are (at least) two sets 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 Cāijiā 蔡家 language (Bó 2004) on the Yúnnán–Guīzhōu border. Zhèngzhāng (2010) argues that Cāijiā is a sister of Bái (and hence also genealogically related to Sinitic in his theory). Sagart (2011) considers Cāijiā (or at least the Sinitic layer in Cāijiā if Cāijiā turns out not to be a Sinitic language) a sister of Wǎxiāng. Wǔ and Shēn (2010: 30–42) point out the lexical similarities amongst Wǎxiāng,

³ In the first edition of the Language Atlas of China (Wūrm & 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 Guāngxī Patois’ is not actually used in the Language Atlas of China; this term is increasingly popular in referring to the Patois in Eastern Guāngxī in the Hēzhōu 會州 area (e.g. Chén and Lǚ 2009). These patois are considered a type of Northern Pínghuà in the Atlas. However, they are better viewed as a geographical continuation of the neighbouring Patois of Southern Húnán.
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 language, their boundaries follow the boundaries of the historical prefectures or counties to some degree. For instance, although nowadays the language area of Huī Chinese is split amongst the three modern provinces of Ānhui, Zhējiāng, and Jiāngxi, it corresponds largely to the historical prefectures of Huīzhōu 徽州 and Yánzhōu 廣州. 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 Chinese is largely confined within the drainage basin of Xiāng 湘 and Zi 资 Rivers (both tributaries of the Yangtze). Terrain is one major factor that caused the diversity of the Sinitic languages to be concentrated in Southern China, rather than Northern China, where the Chinese language originates. Southern China is mountainous, and the linguistic diversity is relatively high. 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 very diverse; correlating with this fact is the unevenness of the terrain of that area, which is not part of the North China Plain. Then there is the complicated migration history. In the case of Mandarin, Mandarin expanded outward from the North China Plain area rapidly within the last few centuries. Towards the northeast, the ban on Hàn Chinese people settling in Manchuria began to relax in 1860. Towards the northwest, Northern Xinjīāng Mandarin formed in about 1780 (Liū 1993:4). Towards the southwest, Mandarin speakers arrived during the Míng Dynasty (1368–1644).

The Sinitic languages are also notable for that the bulk of their speakers have been united under unified single regimes for most of their history. Chinese people in general recognise the hegemony of the Common Chinese language, of which the latest stage is Standard Mandarin. Even when China is not unified, people use varieties of the same Common Chinese language as a lingua franca. The concept of there being a Common Chinese language began as early as the Western Zhou dynasty (11th century BCE – 771 BCE); it 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. The diversity amongst the modern Sinitic languages is partially caused by them having preserved linguistic material from various historical stages of Common Chinese. For instance, out of the major branches of Sinitic, only Mín retained a phonological layer from

---

4 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.
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 wiped out all phonological diversity amongst the Sinitic languages other than Mín. The tree model is ill-fitted to the Sinitic languages, as some 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 has been diffusion amongst the various non-standard Sinitic languages (e.g. the influence of Cantonese on Hakka and Mín in Guǎngdōng Province), making the classification of the Sinitic languages a notoriously difficult job.

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. Hashimoto (1978) and (1986) are the first major works that discuss the 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 and Tibeto-Burman peoples for more than one thousand years, intermittently, during the last two millennia. The most influential dynasties were Mongolic (e.g. the Khitan Liáo Dynasty, 907–1125 CE) or Tungusic (e.g. the Jurchen Jin Dynasty, 1115–1234 CE). The latest North Asian dynasties, the Mongol Yuán Dynasty (1279–1368 CE) and the Manchu Qīng Dynasty (1644–1912 CE), governed the entirety of China rather than just Northern China. 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 Xīxiá Dynasty, 1038–1227 CE), and people of other Northern or Western ethnicities. 5 Northern Chinese was influenced greatly by the North Asian languages, Mongolic and Tungusic languages in particular, due to the North Asian languages being politically powerful, and also that many of the North Asian people shifted into speaking

5 During the Sixteen Kingdoms period (304–439 CE), there were various polities headed by the Di 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) 耷 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 Gào Yún 高雲/ Ko Un 고운 (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 Lūshān 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.
Chinese. For instance, under Altaic influence, in Mandarin and Jin there are less tones, less 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 post-positional case markers and often SOV, namely the Far-Western Central Mandarin dialects of Línxià (a.k.a. Hézhōu 河州) and Xīníng 西寧 areas (e.g. Dede 2007), and the Tángwàng 唐汪 language (Djamouri forthcoming).6

Huàngshuǐ Mandarin (Xīníng area)
1. 狗肉 哈 吃了
   dog  meat [OBJ] eat PFT
   ‘The dog ate the meat.’ (Dede 2007: 867)7

The situation with Southern China was the opposite: Chinese people cause disturbance to the Southern non-Sinitic people more often than the reverse. Before the arrival of Chinese people, in Southern China there were Kra-Dai,8 Hmong-Mien, Austronesian and Austroasiatic-speaking people.9 China first set up administrative bases in the Pearl River region and in the lower Red River regions during Qin Dynasty (221–207 BCE). From then onwards, the vast majority of migration movements for Chinese people had 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 had many wars with North Asians). The southward migration of Chinese people caused the southward migration of the Southern indigenous people deeper into Southeast Asia. Some of the indigenous population of Southern China were absorbed into the migrant Chinese

6 Nearby there is also the mixed language Wǔtún 五屯 (e.g. Janhunen, Peltomaa, Sandman and Dongzhou 2007) of which the vocabulary is over 50% Mandarin, and the grammar is mostly Tibetan. The phonology and lexicon in Wǔtún is not as obviously Sinitic-like as Tángwàng. See Zhōng (2007) on the language contact situation in this area.
7 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: a) it is a reflex of the same character in older stages of Chinese; b) it is homophonous with that character but the morphemes etymologically different; or c) they have the same meaning but the morphemes are etymologically different, and they are not even homophonous.
8 Kra-Dai is an increasing-used term (e.g. Ostapirat 2000, 2005; Pittayaporn 2009) for the language family usually called Tai-Kadai.
9 Ostapirat (2005) argues for the close relationship between Kra-Dai and Austronesian, and Sagart (2004) argues that Kra-Dai people were Austronesian back-immigrants from Taiwan. That some conservative Kra languages have sesquisyllabic forms matching the disyllabic forms in Austronesian languages is a strong support for the link between Kra-Dai and Austronesian families. Many Kra tribes have legends of their ancestors coming from the east (i.e. from Guǎngdōng and Fǔjiān type of direction) and having crossed the sea in big boats (Li 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 crossing like the Mouth of the Pearl River.
population. Genetically, it is known that the patrilineage of many Southern Chinese men 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 on the Northern Pínghuà speakers, which concluded that Northern Pínghuà speakers are genetically primarily Southeast Asian on both their patrilineage and matrilineage (Gan et al. 2008).

Linguistically, many Southern Sinitic languages 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 Li 2002: 94–149). Roughly corresponding to modern day Wú-speaking area was the Yuè 越 kingdom (? – 222 BCE), which was probably Kra-Dai-speaking, judging by the transliterated lyrics of the sixth century BCE Song of the Yuè (Yuèréngé 越人歌; Wěi 1981, Zhengzhang 1991) and words in other sources. 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). Mín is argued by Norman and Mei (1976) to have an Austroasiatic substratum (but this theory is criticised by Sagart (2008)). 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, we shall 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.

2. 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):

– Mon-Khmer languages (perhaps not including far-flung ones like Khasic(?));

---

10 Gan et al. (2008) make their claim for Pínghuà people in general. However, all but one of their sampling locations are Northern Pínghuà-speaking.
11 There are competing theories in Vietnam that the language in Yuèréngé (Viet Nhơn Ca in Vietnamese) is Vietic.
12 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 Mín- 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 resembles Yào (≡ Mien) in Northern Guǎngdōng the most, and Ho Ne people do in fact consider themselves Yào, according to Yóu (2002: 8–10).
– Kra-Dai languages;
– Hmong-Mien languages;
– Chamic languages (perhaps not including Achenese);
– Some of the surrounding Sino-Tibetan languages, e.g. Karen, Lolo-Burmese, some nearby Sinitic languages.

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 admitted in Comrie (2007, 2008a)). Most studies on language areas begin by having preconceptions of what linguistic features are common in a linguistic area, and then the geographical extend of the said features are determined. Comrie (2007) takes a different approach. Instead of having a preconceived list of typological features, all features in the World Atlas of Language Structures (WALS; Haspelmath, Dryer, Gil, and Comrie 2005) were examined to see whether there are typological features that distinguish MSEA from other areas. 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 canonically MSEA features. Comrie (2008a) follows similar methods, but concentrates on the Sinitic languages. The Sinitic languages are compared with both MSEA and North Asia in Comrie (2008a). 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 a score of 8 out of 20 for MSEA features (the lowest scored language out of the surveyed languages),¹³ 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 Chinese is typologically between MSEA and North Asia.¹⁴ 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;

¹³ Comrie (2007) has an extra MSEA feature that is not featured in Comrie (2008a): feature 45A ‘Politeness Distinctions in Pronouns’.
¹⁴ Of cause one could also say that the MSEA and North Asian languages are typologically half-like the Sinitic languages. However, I suppose MSEA and North Asia serve as good typological standards of comparison as their word order typological profiles are more normal: the MSEA languages are rather consistently left-headed, while the North Asian languages are very strongly right-headed. The Sinitic languages have typologically unusual profiles of being SVO but otherwise strongly right-headed, as discussed in the rest of this paper.
– 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:
  o Verb – Object order;
  o Preposition – NP order;
  o Noun – Genitive order;
  o Noun – Adjective order;
  o Noun – Demonstrative order;
  o Noun – Numeral order;
  o Noun – Relative clause order;
  o Adjective – Degree word order;

– ‘Topic’ predicative possession construction (“possessor-TOPIC exist possessum);
– Verbal encoding for predicative adjectives; and
– 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 newer 2011 online edition of WALS (Dryer and Haspelmath 2011), and added the following languages: Cantonese, Hakka, Eastern Kayah Li, Hmong Njua, and Mien. Since this paper focuses on the Sinitic languages, it would preferable to include more data from the Sinitic languages. However, since there is a general lack of data from the Sinitic languages in WALS (and in Western Linguistics in general), only two non-Mandarin Sinitic languages with a reasonable amount of data in WALS — Cantonese and Hakka — are included in this exercise. 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 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) for Hmong, 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 scores 14 and 13 respectively, which are closer to the score of 16 achieved by Khmer in the core of MSEA. (Amongst the Hmong-Mien languages, the Hmongic languages are in general less
influenced by Chinese, and are hence typologically more like the core MSEA languages than the Mienic languages (e.g. Ratliff 2010: 239–240).
<table>
<thead>
<tr>
<th>Map</th>
<th>Feature</th>
<th>Thai</th>
<th>Khmer</th>
<th>Vietnamese</th>
<th>Indonesian (EKayahLi)</th>
<th>Burmese</th>
<th>(Hmong Nj)</th>
<th>(Mien)</th>
<th>(Cantonese)</th>
<th>(Hakka)</th>
<th>Mandarin</th>
</tr>
</thead>
<tbody>
<tr>
<td>7A</td>
<td>C°</td>
<td>No</td>
<td>Implosives</td>
<td>Implosives</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>9A</td>
<td>ŋ</td>
<td>Initial</td>
<td>No initial</td>
<td>Initial</td>
<td>Initial</td>
<td>Initial</td>
<td>Initial</td>
<td>No</td>
<td>Initial</td>
<td>No</td>
<td>No initial</td>
</tr>
<tr>
<td>11A</td>
<td>y ø æ</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>/y ø æ/</td>
<td>None</td>
</tr>
<tr>
<td>13A</td>
<td>Tone</td>
<td>No Complex</td>
<td>No Complex</td>
<td>No Complex</td>
<td>No Complex</td>
<td>Complex</td>
<td>Complex</td>
<td>Complex</td>
<td>Complex</td>
<td>Complex</td>
<td>Complex</td>
</tr>
<tr>
<td>33A</td>
<td>Numl Pl.</td>
<td>No</td>
<td>Pl. word</td>
<td>Pl. word</td>
<td>No</td>
<td>No plural</td>
<td>No plural</td>
<td>No plural</td>
<td>No plural</td>
<td>No plural</td>
<td>No plural</td>
</tr>
<tr>
<td>54A</td>
<td>Distrib. Num.</td>
<td>Optional</td>
<td>Obligatory</td>
<td>Obligatory</td>
<td>Optional</td>
<td>Optional</td>
<td>Obligatory</td>
<td>Obligatory</td>
<td>Obligatory</td>
<td>Obligatory</td>
<td>Obligatory</td>
</tr>
<tr>
<td>68A</td>
<td>Perfect</td>
<td>‘Finish’</td>
<td>‘Finish’</td>
<td>Other</td>
<td>‘Finish’</td>
<td>‘Finish’</td>
<td>No perfect</td>
<td>Other</td>
<td>Other</td>
<td>Other</td>
<td>Other</td>
</tr>
<tr>
<td>83A</td>
<td>Obj &amp; Verb</td>
<td>VO</td>
<td>VO</td>
<td>VO</td>
<td>VO</td>
<td>VO</td>
<td>VO</td>
<td>VO</td>
<td>VO</td>
<td>VO</td>
<td>VO</td>
</tr>
<tr>
<td>86A</td>
<td>Gen &amp; N</td>
<td>N-Gen</td>
<td>N-Gen</td>
<td>N-Gen</td>
<td>N-Gen</td>
<td>Gen-N</td>
<td>Gen-N</td>
<td>Gen-N</td>
<td>Gen-N</td>
<td>Gen-N</td>
<td>Gen-N</td>
</tr>
<tr>
<td>87A</td>
<td>Adj &amp; N</td>
<td>N-Adj</td>
<td>N-Adj</td>
<td>N-Adj</td>
<td>N-Adj</td>
<td>N-Adj</td>
<td>N-Adj</td>
<td>N-Adj</td>
<td>N-Adj</td>
<td>N-Adj</td>
<td>Adj-N</td>
</tr>
<tr>
<td>89A</td>
<td>Num &amp; N</td>
<td>N-Num</td>
<td>N-Num</td>
<td>Num-N</td>
<td>Num-N</td>
<td>Num-N</td>
<td>Num-N</td>
<td>Num-N</td>
<td>Num-N</td>
<td>Num-N</td>
<td>Num-N</td>
</tr>
<tr>
<td>90A</td>
<td>Rel &amp; N</td>
<td>N-Rel</td>
<td>N-Rel</td>
<td>N-Rel</td>
<td>N-Rel</td>
<td>Rel-N</td>
<td>Rel-N</td>
<td>Rel-N</td>
<td>Rel-N</td>
<td>Rel-N</td>
<td>Rel-N</td>
</tr>
<tr>
<td>91A</td>
<td>Deg &amp; Adj</td>
<td>Adj-Deg</td>
<td>Adj-Deg</td>
<td>(Deg-Adj)</td>
<td>Dem-Adj</td>
<td>(Adj-Deg)</td>
<td>Deg-Adj</td>
<td>Adj-Deg</td>
<td>Deg-Adj</td>
<td>Deg-Adj</td>
<td>Deg-Adj</td>
</tr>
<tr>
<td>118A</td>
<td>Predicative Adjectives</td>
<td>Verbal</td>
<td>Verbal</td>
<td>Verbal</td>
<td>Verbal</td>
<td>(Verbal)</td>
<td>Verbal</td>
<td>(Verbal)</td>
<td>Verbal</td>
<td>(Verbal)</td>
<td>Verbal</td>
</tr>
<tr>
<td>119A</td>
<td>Nominl Pl. &amp; Loc Pred.</td>
<td>Different</td>
<td>Different</td>
<td>Different</td>
<td>Different</td>
<td>Different</td>
<td>Different</td>
<td>Different</td>
<td>Different</td>
<td>Different</td>
<td>Different</td>
</tr>
</tbody>
</table>

| Total +: | 18 | 16 | 17 | 13 | (14) | 10 | (13) | (11) | (9) | (10) | 7 |

Table 1 Some typological features in Sinitic and MSEA languages (based on Comrie 2008a; added information are put in parentheses)

---

15 Eastern Kayah Li is coded in WALS as N–Rel. This is problematic. Eastern Kayah Li has two constructions that resemble externally headed relative clauses. The one referred to in WALS, called ‘postposed attributive clause’ (Solnit 1997: 253–258), is more like participial construction: the number of arguments that this participial can take is very restricted. What is structurally more like a relative clause is the ‘preposed attributive clause’ (Solnit 1997: 249–252), which is a clause with much less restrictions. However, the unusual trait of this preposed attributive construction in Eastern Kayah Li is that the head must be a classifier (and the coreferential noun can occur within the relative clause).
In the rest of this paper, we shall discuss further some of the phonological and word order issues discussed in Comrie (2007, 2008a) and some other related issues. I shall take this opportunity to introduce Chappell (2012, in press)’s division of the Sinitic languages into four macro-areas (Chappell 2012: 5–6), with my own alterations (due to differences in linguistic criteria used) and simplifications.16

- Northern zone:
  Bēijīng Mandarin, Northern (Jīlǔ) Mandarin, Peninsular (Jiāolíáo) Mandarin, Northeastern Mandarin, Northwestern (Lányín) Mandarin, Central (Zhōngyuán) Mandarin (?), and Jīn.

- Transitional zone:
  Central (Zhōngyuán) Mandarin (?), Southeastern (Jiānghuái) Mandarin, Southwestern Mandarin, Xiāng, Wǎxiāng, Gǎn, and Western Mín.

- Southeastern zone:
  Mín, Wú, Huí.

- Far-Southern zone (≈ Chappell’s “Southern Area”):
  Yuè, Pínghuà, Hakka, and the Mín exclaves in Léizhōu Peninsula and Hǎinán.

Amongst the four zones, the one with most verb-medial traits is unsurprisingly the Far-Southern zone, as many of these languages still have strong interactions with Kra-Dai languages. Expectedly, the Northern zone has a number of verb-final traits, being in contact with the North Asian languages. However, putting the aforementioned Far-Western Central

---

16 Some of the differences between the four typological zones in this paper and Chappell (2012)’s four macro-areas are:
- the term ‘Far-Southern zone’ is used here instead of Chappell’s ‘Southern area’; ‘Southern Sinitic’ has a number of different meanings, including the non-Northern Sinitic languages, or the Southern non-Mandarin Sinitic languages;
- Northern Wú and Huí are included in the same Southeastern zone as Mín and Southern Wú. Northern Wú and Huí are more strongly influenced by Mandarin, and are hence sometimes treated differently from Southern Wú;
- the Mín 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 Mín dialect of Hǎinán (a.k.a. Hainanese) is strongly influenced by the Kra-Dai language Ong-Be (i.e. the lowland indigenous language of Hǎinán), and the Mín dialect of Léizhōu Peninsula is closely related to that of Hǎinán. Yuè and Pínghuà have also been strongly influenced by Kra-Dai languages;
- Western Mín is a Mín dialect that is strongly influenced by Gǎn, and is here included in the same Transition zone as Gǎn, rather than being in the Southeastern Zone together with other Mín dialects.

Chappell (2012)’s division of the Sinitic languages into four macro-areas is a refinement on Norman (1988: §8.1)’s division of the Sinitic languages into the typological zones of North (Mandarin), South (Yuè, Hakka, Mín), and Central (Xiāng, Gǎn, and Wú). The four macro-areas in Chappell (2012) were based on the distribution of the various grammaticalisation pathways of the passive and object marking constructions. However, it is noted that (2012: 6) 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 approximate, due to the paucity of data. The membership within each zone is sometime speculative, again due to the paucity of data.
Mandarin dialects aside, the zone with the most verb-final traits is, surprisingly, the Southeastern zone.

In the rest of this paper, unless specified, data from Sinitic languages are provided by the seven members of the ERC Sinotype project, 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: Gǔzhāng Wāxiāng (fieldnotes);
– Wēiróng Chén: Huiān Southern Mín (first language and field notes);
– Yūjié Chén: Zhōukōu Central Mandarin (first language and field notes);
– Xūpíng Lǐ: Yichūn Gàn (fieldnotes);
  Fūyáng Wú (first language);
– Sing Sing Ngai: Shàowǔ Western Mín (fieldnotes);
  Fúqīng Eastern Mín (heritage language);
  Standard Cantonese (first language);
– Hilário de Sousa: Nánńíng Southern Pínghuà (fieldnotes),
  Standard Cantonese (first language);
– Jiàn Wáng: Jīxi Huī (fieldnotes);
  Suīníng Central Mandarin (first language).

3. Phonology
In this section, we will discuss the following phonological phenomena in the Sinitic languages and the MSEA languages to the south:

– Tones and onsets;
– Codas;
– Implosives; and
– Front rounded vowels.

Some maps from the “Phonetics” volume of the Linguistic Atlas of Chinese Dialects (LACD; Cáo et al. 2008) will be shown. Notation like “Map P117” refers to Map 117 in the Phonetics (“P”) volume of LACD.

3.1 Tones and onsets
Most MSEA languages have phonemic use pitch and/or phonational differences, which are two closely related suprasegmental features. In this paper, ‘tone’ refers to systems where at least pitch contrasts have been phonemicised. Many of these systems also include phonational contrasts. (Languages where only phonational contrasts, but not pitch contrasts, have been phonemicised are not considered to be ‘tonal’.)
Many languages in East and MSEA are tonal. Many language families in this area had a 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 earlier languages with 3+1 tones include:
- Proto Kra-Dai;
- Proto Hmong-Mien;
- Middle Chinese;
- Proto Mín;
- Proto Việt-Mường;
- Proto Bái;
- Proto Lolo-Burmese; and
- 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 other Mon-Khmer languages (Haudricourt 1954).

The Sino-Tibetan languages have Written Tibetan as a reference. (Classical Tibetan was most probably non-tonal; there are still Tibetan dialects in the periphery which are nontonal.) Written Burmese in fact still often marks the high tone with း, which is related to the Indic sign विसर्गः (*-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: basically 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 of these languages have moved beyond this 3+1 tone system. What usually happens is that the voicing contrast of the onset causes the tones to develop a (pitch-wise) higher allotone and a lower allotone. When the contrast between modal voiced and modal voiceless obstruent onsets is lost, the two sets of allotones become separate phonemes. Theoretically a language with 3+1 tones would thus end up with 6+2 tones. However, most languages have other splits and mergers between these tones. For instance, Northern Vietnamese has 6+2 tones, Standard Lao has 5+4, while Central Thai has 5+3 tones. Amongst the Sinitic languages, the Far-Southern languages, being closest to the core of MSEA, have the most tones. The Southeastern languages have a bit fewer tones (a tangent fact is that this is the zone with the most rampant tone sandhi), the Central languages have even fewer tones, and the Northern languages, being closest to North Asia, have the least tones. Prototypically:
- Yüè and Southern Pínghuà dialects have 6+3 tones;
- Mín and Wú dialects have 6+2 or 5+2 tones;
– (The number of tones in Gàn dialects varies hugely);
– Hakka dialects have 4+2 tones;
– Xiāng dialects have 5+1 tones;
– Southeastern Mandarin dialects have 5+1 or 4+1 tones;
– Jin dialects have 4+1 tones;
– Other Mandarin dialects have 4+0 or 3+0 tones.

LACD Map P001 shows the number of tones amongst the Sinitic languages. The languages with the highest number of tones are clearly in Far-Southern China, the area closest to the core of MSEA. Mandarin has the least number of tones, especially Northwestern Mandarin. (Notice that in Chinese linguistics, tones in sonorant-ending syllables and obstruent-ending syllables are counted separately. Other allotones are also counted separately.)

**LACD Map P001**

One trait that the hallmark of Kra-Dai-ness in Yuè and most Southern Pínghuà dialects is the split of tone D (the tone for obstruent-ending syllables) based on vowel length.¹⁷ This split in tone D is near universal amongst Kra-Dai languages, universal in Yuè-dialects, and common in Southern Pínghuà dialects. The only other non-Kra-Dai language that I know of with this trait is Kim Mun (Mienic) in Hǎinán (Lǐ 2003: 694–697), which perhaps is an influence from Hlai.

The losing of voicing contrast (for plosives) has also occured in the many Mon-Khmer languages, which are mostly non-tonal. For instance, in Mon, the old voicing contrast of the onsets is now expressed by a phonational contrast of modal versus breathy. The phonational contrast caused a changed the vowel qualities (e.g. Jenny ___. In Khmer, not only has the onset voicing contrast been lost, the phonational contrast has also been lost in most dialects. This phonemicised the vowel quality contrasts (e.g. Wayland and Jongman 2002).¹⁸

¹⁷ Southern Pínghuà dialects in Nánníng and to the west split the tone D not by vowel length, but by the sonority of the initial consonant in Middle Chinese, e.g. Nánníng Wèizīlù Pínghuà /wɐt²/ 域 ‘region’ (< *wik), /wɐt²/ 話 ‘live’ (< *ywat). See de Sousa (forthcoming).
¹⁸ There are also some tonal languages which have split vowel qualities based on tones, presumably through an intermediary stage with phonational difference which has since been lost:
- Mang (Mǎshān 麻山 Miáo, Western Hmongic; tones B2 and C2 versus others) (Wáng 1985: 107; Ratliff 2010: 196);
- About half of the Eastern Mín 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’ breathiness phonation, and tone 2, which correlates with voiced onset and lower pitch, which also ‘encourages’ breathiness phonation.
It is interesting to note that there are languages in MSEA where the voicing-related tone-splitting has not happened:

- Burmish languages like Burmese, Achang and Xiāndào;
- Nusu (Loloish);19
- A-Hmyo dialects (Luóbóbó 罗泊河 Miáo, at, e.g. Fúquán 福泉; Western Hmongic; tone D has partially or totally merged to tone A) (Ratliff 2010: 185; Lǐ 2003: 686–688).

There are also languages where tone-splitting has occurred, but the tone-splitting has not been phonemicised, as the original contrast between modal voice and modal voiceless onsets is still largely intact. (The original modal voiced onsets may have changed into something else like breathy voiced, but they are still distinct from the modal voiceless onsets.) These languages include:

- Wú dialects, including some neighbouring Wú-influenced Mín varieties as in:
  - Eastern Mín in Cāngnán, Zhèjìāng;
  - Southern Mín in Guǎngfēng, Jiāngxī;
  - Northern Mín in Pūchēng, Fújiàn (Zhèngzhāng 1995);20
- Southern Xiāng dialects (‘Old Xiāng’);
- Wǎxiāng and some nearby Mandarin dialects;
- A few Northern Gàn dialects, e.g. Wǔníng (Zhū et al. 2009);
- A few Northern Yuè dialects, e.g. Liánshān, Yángshān (Zhèngzhāng 1995);

The phonemicising of (further) suprasegmental features based on the lost of the original contrast between modal voiced and modal voiceless onsets seems to be the norm in MSEA. This is summarised in table 2 in §3.2.

3.2 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 -n̂ -ñ -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; and

19 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. Lǐ 2010: 56).

20 One important feature that distinguishes Wū and Huī is that Huī dialects have phonemicised the splitting of tones.
Middle Chinese (Baxter 1992): \[-p -t -k -m -n -\eta -\text{"\(\eta\)}\].

In some languages there is a dramatic loss of coda distinctions. For instance, while Mien has preserved \[-p -t -k -m -n -\eta\] (Máo, Méng and Zhèng 1982: 16), Hmong has lost all the plosive codas, and all nasal codas have collapsed into one \(-\eta\) or vowel nasalisation (Wáng 1985: 18). Most Kra-Dai, Mienic and Mon-Khmer languages have at least three plosive codas and three nasal codas. In the table 2 below I divide the sampled languages based on two criteria: a) having more than one contrastive plosive coda; and b) having more than one contrastive nasal coda. It is the norm in MSEA to have at least two plosive codas and two nasal codas (usually more). With the Sinitic languages, LACD Map P121 shows the distribution of \[-m -n -\eta\] and LACD Map P124 shows the distribution of \[-p -t -k -\text{"\(\eta\)} -l\] in the Sinitic languages.

**LACD Map P121**

**LACD Map P124**

Having two or more plosive codas are largely confined to the following Sinitic languages in or near Far-Southern China, which is closest to the core of MSEA:
- Southern Mín (including Mín in Hǎinán and Léizhōu Peninsula);
- Yuè;
- Southern Pínghuà;
- Hakka in Guǎndōng;
- Some Gàn dialects.

The same sets of Sinitic languages satisfy the criterion of having \(-m\) and one other nasal codas. Many Mandarin dialects have \(-n -\eta\) or just one of these.

The following is a summary of the typological features discussed in §3.1 and §3.2.

<table>
<thead>
<tr>
<th></th>
<th>Thai</th>
<th>Khmer</th>
<th>Vietnamese</th>
<th>E Kayal Li</th>
<th>Burmese</th>
<th>Hmong Njua</th>
<th>Mien</th>
<th>Cantonese</th>
<th>Namng Pinghua</th>
<th>Huîn S. Mín</th>
<th>Fiêng E. Mín</th>
<th>Shlawi W. Mín</th>
<th>Yêchîn Gân</th>
<th>Fâiyâng Wû</th>
<th>Jëst Hûi</th>
<th>Gîwàng Wâşîâng</th>
<th>Suîning C Mandarin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Having “Complex tones” (WALS)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

*P \rightarrow P, phonemicised T or _.

21 The syllabic nasals that exist in many Southern Sinitic languages are not included in the criterion of having more than one nasal coda.
Table 2: Some onset, coda and suprasegmental features in Sinitic and MSEA languages
((+): Most Khmer dialects have lost the original phonation contrast.)

3.3 Implosives

Many MSEA languages have the implosive consonants ɓ and ɗ (but no ɠ, as voicing is more difficult to maintain when the distance between the glottis and the oral closure is short). Examples include Khmer, Vietnamese, Lao, and Thai. Some other MSEA languages are said to have ɓ and ɗ which are not implosives (e.g. Eastern Kayah Li; Solnit 1997). As for the Sinitic languages, neither Middle Chinese nor Old Chinese were reconstructed with implosive consonants. However, some modern Sinitic languages have implosives. LACD Map P044 shows the distribution of implosive onsets in the Sinitic languages.

LACD Map P044

According this map, implosives are found in:

- Mín in Hǎinán and Léizhōu Peninsula;
- Dānzhōu dialect (the Yuè-like language in Northeastern Hǎinán);
- Some of the Gōulōu Yuè dialects near the Guǎngxī–Guāngdōng border;
- Some Southern Wú dialects; and
- Some Northern Wú dialects around Shànghǎi.

The most famous example is Hainanese, which is in general very strongly influenced by Ong Be, the lowland Kra-Dai language in northern Hǎinán. Across the Hainan straight, there are some of the Gōulōu Yuè dialects which have implosive ɓ and ɗ. (However, in some localities they are becoming p and t.) Further away to the northeast, there are implosives in some of the Wú dialects.

Their origins differ. In Hǎinán and Léizhōu Mín, ɓ and ɗ developed out of *p and *t after *b and *d lost their voicing and merged into *p and *t (Southern Mín speakers settled in Hǎinán and Léizhōu relatively late), whereas in the other Sinitic languages (including the Yuè-like Dānzhōu dialect in Hǎinán) ɓ and ɗ developed out of *p and *t when *b and *d
was still distinct. Around Shanghai, a new Ʉ has developed out of *g (unlike ɓ and ɗ which developed out of *p and *t).

There are also some other Southern Sinitic dialects where the Middle Chinese onsets *b and *d underwent sound changes different from *g and other voiced obstruents. For instance, in many patois of Southern Húnán and Northern Quảngđồng, *b *d became voiceless unaspirated, whereas other voiced obstruents became voiceless aspirated (see Qín 2007: 185 for a full list of localities where this happened). This suggests that the Middle Chinese *b and *d onsets, perhaps under the influence of MSEA languages, were pronounced as implosives in these Sinitic dialects, whereas the other voiced obstruents were pronounced as non-glotallic.

See Zhū (e.g. 2006b, et al. 2009) on implosives in Sinitic languages, including some newly developed implosives in Northern Gàn dialects and Chíoshàn Mín dialects (e.g. Shàntóu/Swatow). The existence or non-existence of ɓ ɗ ~ b d is summarised in table 3 in §3.4.

3.4 Front rounded vowels and back unrounded vowels
A common trait in MSEA is the lack of front rounded vowels. This is also generally the case with the following Southern Sinitic languages:
- Southern Mín, including Mín of Hánán and Léizhōu Peninsula;
- most Hakka dialects and some neighbouring Southern Gàn dialects;
- most Yuè dialects not in the Pearl River Drainage Basin;
- Most Southern Pínghuà dialects;
- Some Southern Mandarin dialects, especially in Yúnnán and Guìzhōu.
Map P117 in LACD is a map on the distribution of /y/ in Sinitic dialects.

LACD Map P117

A more useful diagnostic feature of the core of MSEA is perhaps having non-low back unrounded vowels (including [i] that is non contrastive with [u]). Unrounded non-low back vowels are also found in some Sinitic languages.

The following table is a summary of §3.3 and §3.4.
4. Word order

The core MSEA languages are SVO, and they are more strongly left-headed than the usual SVO language (see, e.g. Dryer 2001 on Mon-Khmer word order). The Sinitic languages are also basically SVO. However, the Sinitic languages are otherwise strongly right headed: noun phrases are strongly right headed, and most adjuncts are placed before the verb. Contrast the word order in the following sentences from Northern Zhuang (Tai) and Cantonese (Sinitic).

**Northern Zhuang**

**SVO order**

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

**Head noun left of most modifiers**

3. **go oij [duz vaiz gou caij laemx henz roen] haenx raek lo**
   CL sugar_cane CL buffalo 1SG step fall side road that break FP
   ‘The sugar cane that my buffalo trempled on the side of the road snapped.’
   (Wéi and Qín 2006: 251)

**Standard Cantonese**

**SVO order**

4. **佢 賣 菜 又 賣 肉**
   佢 kʰoê³³  賣 maï²² 又 thi³³ 菜 jeu²² 肉 maï²² 肉 juk²²
3SG sell vegetable as:well sell meat
’S/he sells vegetable and sells meat.’

**Head noun right of modifiers**

5. 我 於 牛 路 上 踏 跤 喂 聒 歌 斷 咽
   [ŋɔ³ tsek⁵ nœu¹] he²⁵ lou²² pin⁵⁵ tsʰar²⁵ lem³⁳] kɔ²⁵ lok⁵ tse³³ tʰyn²⁴-tsə³⁵
1SG CL bovine at road side step fall DEM CL cane break-PFV

‘The sugar cane that my buffalo trempled on the side of the road broke.’

This created some extraordinarily rare co-occurrence of word order in the Sinitic language. For instance, the co-occurrence of the VO order and the Relative clause–Noun order is nearly unique to the Sinitic languages (WALS feature 96A).²² The Sinitic languages are the only VO languages with obliques predominantly placed in front of verb in WALS (feature 84A).²³ Having the Adjective–Noun word order (feature 81A) for SVO languages (feature 87A) is also very rare in the region.²⁴

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 mixed 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.**
The verb-medial 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;

²² 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 quoted in Comrie (2008b: 729–730), having Rel–N order in SVO languages might aid processing when the object is relativised, as having a SV relative clause in front of the relativised object head resembles the normal SVO word order (Yip and Matthews 2007). There are indeed cases like Pwo Karen where relativised objects can have a prenominal relative clause, and relativised subjects must have a postnominal relative clause (Kato 2003: 641), resembling normal SVO word order in both cases.

²³ Of the 500 languages sampled in WALS feature 84A, only the three Sinitic languages sampled have the word order of XVO.

²⁴ 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). In Asia, including Western Austronesia, there are only two languages other than Sinitic and Bái 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 Bái as being the only SVO and Adjective–Noun languages in Asia.
Utsat (Chamic; strongly Chinese influenced)

6. ʔaɪʰtsatʰ ʔaɪʰtaɪ³⁴ ʔaɪʰsiaŋ³⁴ ʔaɪʰhoʱ³⁴liən³⁴ ʔaɪʰkai³³ ʔaɪʰsa³³ ʔaɪ³⁴ ʔaɪʰse⁴⁵,

l.sister CL very feel:sorry old.man MOD one CL

ʔaɪ³³kai³³ ʔaɪ³³ni³³ ʔaɪ³³sa³³ ʔaɪ³³ta³³ ʔaɪ³³se⁵⁵ ʔaɪ³³ten³² ʔaɪ³³pa³³,

know old.man this MOD one CL stomach hungry

‘The little sister was very sorry for the old man, and knew that the old man was hungry,’ (Zhèng 1997: 238)

(ʔaɪ³³siaŋ³⁴ ʔaɪ³³hoʱ³⁴liən³⁴ are Chinese loanwords in Chinese word order: 非常可憐.)

Green Hmong (Hmong-Mien)

7. kuv nyam tug txivneej kws ncaws pob

1SG like CL man REL kick ball

hab tug txivneej kws moog rua Fresno

and CL 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).

The verb final Tibeto-Burman zone to the west. These languages are SOV, primarily rightheaded but not very strongly right-headed (e.g. Tibetan and Burmese are SOV and have N–Num and N–Adj word order). The Tibeto-Burman languages are verb final, except for the following SVO languages, which, like the Sinitic languages, exhibit interesting mix of VO and OV properties (“Area A–B”): the Karen languages, Bái languages, and Mru (Peterson 2005);²⁵

Burmese

8. thu  di  hsei:  thau’  me

3 this medicine drink IRR

‘He’s going to take this medicine.’ (Myint Soe 1999: 132)

9. thu.  le’  nyi’=pa’  ne.  nga.  kou  la  tou.  te

3GEN hand dirty with 1 OBJ come touch RLS

‘(He) touched me with his dirty hands.’ (Myint Soe 1999: 256)

²⁵ 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 Khassic or Palauanic languages, and there seems to be no Mon-Khmer lexical borrowings in Mru (Löffler 1966). See more discussions in Djamouri, Paul and Whitman (2007).
Eastern Kayah Li (Karenic)

10.  \textit{phrem}ɔ̀ mɛ́ th th ʌ phrekhū sī nā
    woman look:see man CL two
    ‘Some women saw two men.’ (Solnit 1997: 181)

11.  ʔa khē təlwā səklā nē səkhō
    3 paddle pass boat PREP snag
    ‘He paddled the boat past the snag (fallen log).’ (Solnit 1997: 159)

Area C.
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.²⁶

Uyghur (Turkic)

12.  sɛn bu kino-ni kør
    2SG this film-ACC see[IMP]
    ‘You watch this film!’ (Abulimit 2006: 239)

13.  top ojna-watqan bala bizniŋ sinip-ta oqu-jdu
    ball play-CONT boy 1PL:GEN class-LOC study-3.NPAST
    ‘The boy who is playing with a ball studies in our class.’ (Abulimit 2006: 324)

The SVO word order in the Sinitic languages resembles that of the verb-medial MSEA zone to the south, while the strong right-headedness in the Sinitic languages resembles that of the verb-final North Asian zone to the north. In fact, the strong right-headedness of the Sinitic languages makes them typologically more similar to the North Asian languages than their relatives — the Tibeto-Burman languages — to the west.

We shall discuss the syntax at the noun phrase level first, and then at the clause level.

4.1  
\textit{Word order in noun phrases}
In or close to the core of MSEA, most modifiers follow the head noun (e.g. Simpson 2005).

²⁶ There are also the following typological zones in and around China that the Sinitic languages had less contact with: a) the Formosan–Philippines languages, which are mostly verb initial; b) the languages of the Indic area, which are verb final and strongly-right headed, except Kashmiri which is verb-second; c) Sarikoli and Wakhi in Western Xinjiang (Gawarjon 1985) are verb final and more strongly right-headed than other Iranian languages, but they still have the Iranian trait of having prepositions (although they also have some Uyghur-like postpositions).
Lao
14. khon2 suung3
   person tall
   ‘tall person’ (Enfield 2007: 93)

15. khaw5 niaw3
   rice sticky
   ‘sticky rice’ (Enfield 2007: 93)

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

Eastern Kayah Li
17. ?iswí nā bēlɔ du
   curry two bowl big
   ‘two big bowls of curry’ (Solnit 1997: 180)

Burmese, which is verb final, has some post-verbal modifiers, like the nominalised stative verb a-thi’ ‘new’ and stative verb hklei: ‘small’ in the following example. (Attributive nouns like thi’tha: ‘wood’ have to precede the head noun.)

Burmese
18. thi’tha: ein a-thi’ hklei:
    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.

Pre-Archaic Chinese (14th to 11th century BCE)27
19. 上甲 惠 王 報 用 五 代

---

27 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.
shàngjiā huì wáng bào yòng wǔ fá
Shangjia FOC king bao:sacrifice use five human:victim

shí xiǎo láo
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 Héjí 924)

Early Archaic Chinese
20. 天 不 庸 説 于 文 王 受 命
tiān bù yòng 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].’
(Aldrige, to appear; Shàngshū, Jūnshì 君奭; approx 8th century BCE)

21. 非 時 伯 夷 播 刑 之 迪?
fēi [shí bò yí bō] xíng zhī dí?
not.be [then Boyi promulgate] law GEN guide
‘Is it not the laws promulgated by Boyi which guide (you)?’
(Aldrige, to appear; 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. Numerals, in particular, were placed variously in front or after the head noun.

Pre-Archaic Chinese (14th to 11th century BCE)
22. 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 Héjí 3018)

23. 獲 唯 鳥 七
huò wéi niǎo qī
capture COP bird seven
‘The catch is seven birds.’
The earliest classifier-like words more often follow rather than precede the head noun.

Pre-Medieval Chinese
24. 分與文君僮百人
fèn yǔ wénjūn tóng bāi rén
distribute give Wenjun slave hundred people
‘(He) distributed a hundred slaves to Wenjun.’
(Chappell and Peyraube 2007; Shìjì, Sīmǎ Xiāngrú Lièzhuàn 司馬相如列傳, approx 1st century BCE)

Early Medieval Chinese
25. 時跋蹕提國送獅子兒兩頭與乾陀羅王
shí bá báí guó sòng shīzǐ ěr liàng tóu yǔ gāntuólú wáng
time ? Bactria country offer lion child two CL HEAD give Gandhāra king
‘At that time, the kingdom of Bactria offered two lion cubs to the king of Gandhāra.’
(Chappell and Peyraube 2007; Luòyáng Qiélánjì 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). However, it can also be argued that the post-nominal classifiers do not form a phrase with the preceding noun in some MSEA languages, e.g. Lao, where a phrase can often intervene between a [NUM + CL] phrase and the preceding noun which it modifies in a semantic sense.

Lao
26. kuu3 sùù4 paa3 sòòng3 too3
1SG buy fish two CL
‘I bought two fish.’ (Enfield 2007: 120)

27. kuu3 sùù4 paa3 juu1 talaat5 sòòng3 too3
1SG buy fish be.at market two CL
‘I bought fish at the market, two (of them).’ (= ‘I bought two fish at the market’) (Enfield 2007: 120)
(“This is a perfectly normal way of saying ‘I bought two fish at the market’, and has nothing of the pragmatically marked quality shown by the first English free translation [...].” (Enfield 2007: 121))
Looking at the modern Sinitic languages, their noun phrases are even more strongly right-headed than the ones in older stages of Chinese.

Nánning Pínghuà
28. 我 個 對 舊 皮 鞋
   ŋa13 kɔ55 tɔi55 kɔ}>22 pɔi11 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 say55 (fish raw) ‘raw fish’, 菜乾 tsʰɔi33 kɔŋ55 (vegetable dry) ‘dried vegetable’, 人客 jɛn11 hak33 (person guest) ‘guest’, 熊人 hʊŋ11 jɛn11>25 (bear person) ‘brown bear (child’s word)’. (See also, e.g., the many left headed compounds in Wênzhōu Wú (Zhèngzhāng 2008: 232)) More productive are the gender affixes for animals. The general trend is for the Northern Sinitic languages to have gender prefixes, resembling the right-headed word order in North Asia, and the Southern Sinitic languages to have gender suffixes (Nánning Pínghuà is a major exception), resembling the left-headed word order in MSEA. Some Sinitic dialects in the middle have prefix for one gender and suffix for the other gender.

Standard Mandarin (prefixes)
29. 公豬 ㄍㄎ-ㄓㄨ  (male-pig) ‘boar’
30. 母豬 ㄆㄨ-ㄓㄨ  (female-pig) ‘sow’

Wǎxiāng (prefixes and suffixes)
31. ○豬 ㄔㄧㄢˇ-ㄊㄡㄨ  (male-pig) ‘boar’
32. 豬娘 ㄊㄧㄡ-ㄋㄧㄤ  (pig-female) ‘sow’

Fùyāng Wú (prefixes and suffixes)
33. 雄雞 ㄏㄧㄠ尪-ㄕㄧ  (male-fowl) ‘rooster’
34. 雞娘 ㄕㄧ-ㄋㄧㄚ  (fowl-female) ‘hen’

Shàowǔ Western Mín (suffixes)
35. 雞公 ㄑㄧ广告服务男-ㄕㄧ  (fowl-male) ‘rooster’
36. 雞娘 ㄑㄧ广告服务-ㄕㄧ  (fowl-female) ‘hen’

Fūqīng Eastern Mín (suffixes)
37. 雞公 kie³²-kuy⁵³ (fowl-male) ‘rooster’
38. 雞母 kie³²-mɔ⁵³ (fowl-female) ‘hen’

Cantonese (suffixes)
39. 雞公 kvi⁵⁵-kuy⁵³ (fowl-male) ‘rooster’
40. 雞乸 kvi⁵⁵-na²⁵ (fowl-female) ‘hen’

Nánníng Pínghuà (prefixes)
41. 公雞 kuy⁵³-mu13 (male-fowl) ‘rooster’
42. 母雞 mui³³-kvi⁵³ (female-fowl) ‘hen’

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

The following table summarises the noun phrase features discussed in this §4.1.

<table>
<thead>
<tr>
<th></th>
<th>Thai</th>
<th>Khmer</th>
<th>Vietnamese</th>
<th>E Kayah Li</th>
<th>Burmese</th>
<th>Hmong Nhua</th>
<th>Mien</th>
<th>Cantonese</th>
<th>Nánníng Pínghuà</th>
<th>Huế S Mín</th>
<th>Fúqing E Mín</th>
<th>Shào Wu W Mín</th>
<th>Yelín Gnán</th>
<th>Fúyang Wú</th>
<th>JXT Huí</th>
<th>Guizhōng Wángshì</th>
<th>Suíníng C Mandarin</th>
</tr>
</thead>
<tbody>
<tr>
<td>N – Genitive</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>N – “Adjective” (e.g. chicken – big)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>N – Noun (e.g. egg – chicken)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>N – Gender (e.g. chicken – male)</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>–</td>
</tr>
<tr>
<td>N – Demonstrative</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>N – Numeral</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>N – Relative clause</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Total:</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>½</td>
<td>½</td>
<td>½</td>
<td>0</td>
</tr>
</tbody>
</table>
4.2 Word order in clauses

On the clause level, there has been the assumption that Chinese had more verb-final traits the further one goes back into the history of Chinese (Li and Thompson 1974a: 208, LaPolla 1994), due to the rare cooccurence of SVO word order and strong right-headedness in the modern Sinitic languages, and also that the vast majority of Tibeto-Burman languages being verb-final. 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 (Djamouri, Paul, and Whitman 2007). 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.28 Pre-Archaic and Archaic Chinese also had wh-movement (to a position between the subject and the verb), which is a trait very rare with OV languages (e.g. Dryer 1991). In WALS, the modern Sinitic languages are the only VO languages that predominantly place oblique phrases in front of the verb (i.e. XVO order; WALS feature 84A). However, Pre-Archaic is a relatively normal VO language for that it usually places obliques after the object (i.e. VOX word order).29

Pre-Archaic Chinese

43. 呼 多 犬 网 鹿 于 焚
   hū duō quàn wǎng lù yú nóng
   order numerous dog.officer net deer at Nong
   ‘Call upon the many dog-officers to net deer at Nong.’
   (Djamouri, Paul, and Whitman 2007: 3; Jiāguōwén Héjí 10976 recto.)

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

Lao

44. phen1 lin5 phaj4 juu1 talaat5

28 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 Aldrige (to apper).

29 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).
3POL play cards be.at market
‘She is playing cards at the market.’ (Enfield 2007: 390)

Khmer
45. 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; and
– in many Sinitic languages, the association of postverbal constituents with new information/ indefiniteness became stronger.30

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 like Cantonese, and are thus relatively close to the core of MSEA in terms of word order. The Northern Sinitic languages understandably have many verb-final traits; there are even the Far-Western Central Mandarin dialects which have postpositions and often SOV. However, putting the SOV Mandarin dialects aside, the Sinitic languages with most verb-final sentences are surprising not the Northern Sinitic languages, but the Southeastern Sinitic languages. It is rare for more than one constituent to occur after the verb, and M. Qián (2008) summerises the following situations 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 gramaticalised from locative words);
– Some Irrealis sentences, e.g.:
  o Negative sentences (S – O – NEG – V);
  o Yes-no questions (S – O – V – Q);
  o Rhetorical questions (S – O – V – Q);
  o Imperative sentences (except that [NUM–CL] phrase and verbal complements can occur post-verbally);

30 For Mandarin, Li (2011) characterises postverbal constituents as primarily conveying new information. There are also accounts which characterise 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.
– Emphatic existential sentences (S – O – exist – EMP);
– ‘Go’ (e.g. I place go);
– ‘From’ (e.g. I place from go);
– Transitive sentences with a definite object.

In the following subsections, we will discuss the various situations where non-subject constituents have to be preverbal in the Sinitic languages. We will be discussing:
– Position of adverbials and adpositions (§4.2.1);
– Position of modifiers of verbs (§4.2.2);
– Position of objects (§4.2.3);
  o The object marking construction (§4.2.3.1);
  o Preverbal and Postverbal definite objects (§4.2.3.2);
  o Word order in clauses with three place predicates (§4.2.3.3).

### 4.2.1 Position of adverbials and adpositions

Modern Sinitic languages in general allow post-verb constituents less readily than Archaic Chinese and MSEA languages. One word order trait that is probably universal amongst the modern Sinitic languages is that most adverbials are placed in front of the main verb, especially for temporal phrases.

**Nǎnníng Pínghuà**

46. 我 太早 住屋頭 看了 一 出 戲
   1SG just:now at home watch-PFV one CL film
   ‘I watched a film at home just now.’

**Wǎxiāng**

47. 我 朝頭 〇 三 個 餅
   1SG morning eat three CL bun
   ‘I ate three buns this morning.’

**Standard Mandarin**

48. 我 明天 在 站台 等 你
    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.

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

Thai
50. sùa bàu ca aw pay boricääk phhrûngnií
   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 (and also some locations).

Cantonese
51. 我 今日 去 台北
   1SG today go Taipei
   ‘I am going to Taipei today.’

Fúqing Eastern Mín
52. 我 今晩 去 北京
   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 case with Wú and many dialects in the Northern Zone. The destination precedes the verb, and the destination is at least preceded by a preposition. In Northern Wú dialects, the preposition is usually elided, resulting in what appears to be a SOV sentence.

Wäxiäng
53. 你 到 何处 去?
   2SG to where go
‘Where are you going?’ (There are also examples with only to tau33 ‘to’.)

Pingli Central Mandarin

54. 你到哪兒去耶？我到城裏頭去

m34 tau23 lar445 tcʰi33 ieʔ go44 tau23 tsʰo32 tɕʰou tɕʰi33

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

‘Where are you going? I am going to the city.’ (Zhōu 2009: 408)

Wēnzhōu Wú (Southern Wú)

55. 我走溫州去

ŋ̍34 tsau45ʔ jy33>11 tɕə u33 kʰ ei42>0

1SG to Wenzhou go

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

Fùyáng Wú (Northern Wú)31

56. 我今朝（到）上海去

ŋɤ kintsɔ (tɔ) zoŋhe 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. Having postpositions in a SVO language is itself not too surprising if the postposition is grammaticalised 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:

Northern Zhuang

57. youq gwnz taiz

at above table

‘On the table’

Sinitic languages have a preposition, and a postposition of which the semantics is more specific. In Sinitic languages, the postnominal locative word is usually no longer a free noun.

Nánning Pínghuà

31 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).
What is surprising in the Sinitic languages is how dominant the locative postposition has become in some Sinitic languages, especially in some Wú dialects (e.g. Liú 2003; 2012: 11–12). Cantonese is a language where such postposition is less obligatory. In Cantonese, an expression like ‘in the library’ usually only requires a preposition.

In Mandarin, having the postposition is common, but not obligatory.
Mandarin

64. 在圖書館 (裏)
    zài tùshūguǎn (lǐ)
    at library in

    ‘At/in the library’

The postposition is compulsory in most Wú dialects (Liú 2012: 12).

Sūzhōu Wú

65. 掛 勒 客廳 *(裏)
    ko52 ləʔ55 kʰaʔtʰin23 *(lǐ)
    hang at living.room in

    ‘Hung up in the living room’ (Lǐ 1998: 164)

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

Níngbō Wú\(^{32}\)

**Preposition usually omitted for preverbal adverbials**

66. 賊骨頭 (來) 屌坑間 裏 幽 該
    thief (at) toilet in hide FP

    ‘The thief hid in the toilet’ (Liú 2003: 272)

67. 老師 (來該) 黑板 未 寫 字
    teacher (at) black:board on write word

    ‘The teacher wrote on the black board’ (Liú 2003: 272)

**Prepositions made into postpositions**

68. 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 adpositions, whereas ‘come from’ is usually expressed with a postposition ‘from’. The

---

\(^{32}\) The Ningbō Wú examples come without romanisation. It is the norm in Chinese linguistic work that only characters are given.
Northern Wú dialects (especially ones outside of Shanghai) are in general rather verb-final like.

Níngbō Wú

69. 因因 幼兒班 去
   baby kindergarten go
   ‘Baby goes to kindergarten.’ (M. Qián 2008: 136)

70. 我 學校 中 來
    1SG school from come
    ‘I came from the school.’ (M. Qián 2008: 136)

4.2.2 Position of modifiers of verbs
Attributors of verbs are overwhelmingly placed in front of the verb.

Shanghainese (Wú)

71. 你個 人 討飯......熊介個......樣子 立辣 依答
    geq-geq njin thaovac nenkaxeq xiangtsir liq-laq itaq
    this-CL person beggar like appearance stand-PROG there
    ‘The man stood there like a beggar.’ (Zhu 2006a: 155)

Standard Cantonese

72. 佢 慢慢 行
    kʰey^13 man^22 man^25 haŋ^4
    3SG slowly walk
    ‘S/he walk slowly.’

Wâxiâng

73. 你 快 ○手
    ne^25 kʰua^13 tsau^25 ci⁵w^5
    2SG quick move:hand
    ‘Hurry up and get moving,’

Standard Mandarin

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

34 Wú languages have tonal domains that are longer than a syllable. In Shanghainese, except for toneless syllables, there are two contrastive tonal melodies. Zhu (2006) notates the ‘marked’ melody with a grave accent.
74. 你 先 吃 吧 多 吃 一點  

nǐ xiān chī ba duō chī yídian  
2SG first eat FP more eat a:bit  
‘Eat first. Eat a bit more.’

(Sometimes adjectives appear after a verb, but they are part of a verbal complement, where the complement predicates the verb.

Standard Cantonese
75. 行 得 慢吞吞  

hài1 tōk⁵ mán²-t⁶en⁵ t⁶en⁵  
walk MANNER slow-IDEO  
≈ ‘(Someone) walks (and the walking is) so slow.’

76. load 到 仲 慢過 龜速  

lou⁵ tou³ tsou² mán²-k⁶3³ k⁶3³ ts⁶ok⁵  
load EXTEND even slow-surpass turtle:speed  
≈ ‘[The mobile internet] loads (to the extend of being) even slower than turtle speed.’ (m.eprice.com.hk/mobile/talk/3149/33856/1/2/0/; 7 Nov 2012)

However, many non-Mandarin Sinitic languages (other than Mǐn in the Southeastern zone including Western Mǐn) have a few adverbs which are placed after the verb (either immediately after the verb, or at the end of the clause).

Fùyáng Wú
77. 杭州 到 快 噠。  

hǎnxai’ ts kʰua dié  
Hángzhōu arrive soon COS  
‘We are arriving in Hángzhōu soon.’  
(This 快 kʰua may be a prospective marker. 快 kʰua meaning ‘fast’ is placed in front of the verb.)

Yíchūn Gàn
78. （再）去 幾 個 湊  

(tsa⁴⁴ tsë¹⁴ tséi³³ k³¹ tsʰei²⁴  
again go few CL more  
‘Send a few more people.’
79. 食 多 發積
*tɕʰiʔ⁵ to³⁴ fəʔ³⁴-tɕʰiʔ⁵
eat more bit-DIM
‘Eat a bit more.’

80. 你 食 飯 先
nt²⁴ tɕiaʔ¹⁷ fan²¹³ sien³⁴
2SG eat rice first
‘You eat your meal first.’

Hakka
81. 坐 一 下 淋
*tsʰo²⁴ it² ha⁵⁵ tʰiam²⁴
sit one CL more
‘Sit a bit more.’ (Lo 1988: 301–302)

82. 著 少 一 領 衫
tsok² seu¹¹ it² lian¹¹ sam¹³
wear less one CL clothes
‘Wear one piece of clothing less.’ (Lo 1988: 303)

Standard Cantonese
83. 食 堆 雪糕 淋 啦
*stk² maj¹¹ sytʰkou⁵⁵ tim⁵⁵ la⁵⁵
eat as_well ice:cream in_addition FP
‘Have ice cream too!’

84. 打 多 兩 行 字
ta²⁵ to⁵⁵ laŋ¹³ høy¹¹ tsᵣ²²
hit more two line word
‘Type two more lines.’

85. 我 行 些 啦。
ŋɔ³¹ høy¹¹ 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.

Northern Zhuang
86. gou bae gong.
   1SG go first
   ‘I am going now.’

87. gou gwn vanj haeux 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 $tʰim^{55}$ ‘add’.)

Thai
88. raw paythiaw muaŋ thay bɔ̀ybɔ̀y.
    1PL go:trip country Thai often
    ‘We visit Thailand often.’ (Smyth 2002: 104)

Green Hmong
89. tuam moog rua suavteb hab
   Tuam go to China too
   ‘Tuam went to China too.’ (Li 1989: 121)

Khmer
90. knjom kampung raut lec:ng ja:ng sa’baj
    I engage.in run play kind happy
    ‘I was running along happily.’ (Haiman 2011: 216)

91. knjom skoal koat cbah nah
    I recognize 3 clear very
    ‘I recognized him very clearly.’ (Haiman 2011: 216)

4.2.3 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. In the Far-Southern zone, languages like Cantonese is relatively free to have two or more constituents after the main verb. In the Southeastern zone, often only one constituent is allowed after the verb, and the postverbal position is strongly associated with new information or
indefiniteness, so much so that definite noun phrases basically have to be preverbal. Other Sinitic languages like Mandarin are somewhat in between these two extremes.\textsuperscript{35} Although it is known that in SVO languages, there are correlations between the pre-verbal position and definiteness, and the post-verbal positions and indefiniteness (Keenan and Comrie 1977), it is rare for the correlation to be as strong as those in the Southeastern Sinitic languages.

There are three main ways in which the object is preposed to a pre-verbal position, and different Sinitic languages have different preferences for which ones they use if the object is to be placed in front of the verb:

- topicalisation; the surface order could, grammatically speaking, freely alternate between SOV and OSV;
- passivisation: \{undergoer – PASS – actor – verb\};
- object marking, or ‘disposal’ construction: \{subject – OM – object – verb\}

The syntax of these construction varies amongst the Sinitic languages. We shall discuss briefly the object marking construction first.

\textbf{4.2.3.1 The object marking construction}

The object marking construction is also known as the ‘disposal’ construction or the pre-transitive construction. The object marker is most commonly grammaticalised from a verb meaning ‘to take’, 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: §15, Sybesma 1992, Ding 2007, amongst many others). In Mandarin it 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 examples, the object-marked construction is more prevalent than the SVO counterpart, but both are frequently used.

\textbf{Standard Mandarin}

\begin{verbatim}
92. 關上門了
guān-shàng mén le
close-up PRF door le
‘(Someone) locked the door’
('關上門了' on Google: 1,690,000 results; 3 Nov 2012)
\end{verbatim}

\textsuperscript{35} Li characterises the post-verbal position in Mandarin as new information. Others have characterised the postverbal position in Mandarin as indefinite (Li and Thompson 1974b) or focal (LaPolla 1995).
In Mandarin, the marked object is usually definite, but not necessarily. An innovation in Mandarin is that the object marking construction can be used with intransitive predicates (see Chappell in press).

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 (e.g. Pínghuà in Chōngzuō (Lǐ and Zhū 2009: 193, Liáng and Lín 2009: 322); Nánníng Cantonese (Lǐ and Qín 2008: 346–348)). Some other Far-Southern Sinitic dialects have object marking constructions, but their usage are restricted and infrequent (e.g. Cheung 1991 on Standard Cantonese). In the case of Hainanese, they are restricted to inanimates (Lee 2009). The following examples from Cantonese and their search frequencies on the internet is a demonstration of the rarity of the object marking construction in Cantonese in comparison with the Mandarin examples above.

Standard Cantonese

94. 門啲 門

\textit{san}^{55}-\textit{tsə}^{25} \textit{mun}^{11}

close-PFV door

‘(Someone) closed the door.’ or ‘They (shops etc.) are closed.’

(Google search of the string "門啲門": 11,000 results; 3 Nov 2012)

95. 門啲 [度 道] 門

\textit{san}^{55}-\textit{tsə}^{25} \{\textit{tou}^{22} / \textit{tou}^{22}\} \textit{mun}^{11}

close-PFV CL CL door

‘(Someone) closed the door.’

(Google search of the string "門啲度門": 1,410 results; "門啲道門": 277 results; 3 Nov 2012)

96. 將 (度 道) 門 門

\textit{tsəŋ}^{55} \{\textit{tou}^{22} / \textit{tou}^{22}\} \textit{mun}^{11} \textit{san}^{55}

OM CL CL door close
The syntax of the object marking construction varies greatly amongst the Sinitic languages. Mandarin towards the northwest (Western Central Mandarin, Northwestern Mandarin) and the Southeastern Sinitic languages in general have less 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 monosyllabic predicates, nor with negative predicates. However, such verb-final sentences can be found in the northwest and southeast Sinitic languages.

Dungan (Western Central Mandarin in Central Asia)

97. ба гу кан бу жян ли,
   \[pa^{24} kou^{51} kʰæ^{44} + pu^{24} + tɕiæ^{44} li\]
   OM dog look + NEG + achieve PRF
   ‘[He] could not see the dog anymore,’ (Lín 2003: 312)

98. ба та бу кэщин сы ли ма?
   \[pa^{24} tʰa^{51} pu^{24} kʰɛ^{44}ɕi^{24} sz^{51} li ma\]
   OM 3SG NEG happy die PRF Q
   ‘Wouldn’t it be so unhappy?’ (Lín 2003: 313)

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

Fùyáng Wú

99. ῆ 〇 我 打
   \[fii kʰɔŋ ɲɛ tə\]
   3SG OM 1SG hit
   ‘S/he hit me.’

Taiwanese Southern Mín

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

36 Using other classifiers like 對 to² and 隻 tsek⁵ yielded negligible number of search results (less than 10).
(On the other hand, Hǎinán Mín, a Far-Southern Sinitic language, would use a normal SVO sentence, as the object marking construction cannot be used with animates:

Hainanese

101. istringstream sstream

3SG kiss 1SG

‘He kissed me.’ (Lee 2009: 480))

Similar object marking construction also exists in many Hmong-Mien languages. Unlike Sinitic languages like Mandarin and Cantonese where the object markers are no longer used as verbs, in White Hmong the object marker is synchronically still used as a main verb meaning to ‘take’.

White Hmong

102. nws muab pojiam nrauj lawm

3SG take woman divorce PRF

‘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 grammaticalised. (The take verb started to be gramaticalised when the last coreferential pronoun, as shown in the following example, became optional; Peyrabe 1996: 169–170.)

Medieval Chinese

103. 船者乃将此蟾以油熬之

boat:person then take this toad with oil fry 3SG

‘Then the boatman took the toad and fried it.’ (Chappell 2006; quoting Peyrabe 1988, 1996)

Similar ‘take’ serial verb constructions exist in MSEA languages. The choice between the ‘take’ and ‘non-take’ construction is usually (but not always) a stylistic choice in the how the event is presented, rather than a grammatical preference or requirement as in most Sinitic languages.

Lao

104. man2 thim5 ngen2

3 discard money
‘She discarded (the) money.’

105.  man2 qau3 ngen2 thim5
3 take money discard
‘She took the money (and) discarded (it).’ (Enfield 2007: 381)

Vietnamese

106.  tôi tằng cho ban một miếng gà rán
1 gift DAT friend one CL chicken fried
‘I gave you a piece of fried chicken.’

107.  tôi lấy một miếng gà rán tằng cho bạn
1 take one CL 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 dialects also often employ the MSEA-type of ungrammaticalised ‘take’ construction.

Nánning Pínghuà

108. 佢 一 拋 個 隻 煎餅 呢
kəj63 ut3 pʰau53 ʂ⁵³ tsɔr3 tʃiŋ53 pʰən3 ne5³
3 once throw DEM CL pan:cake TOP
就 跌落 大象 隻 煎鍋
tʃʊə55 tʃi5 tʃi23 tʃe53 tʃiŋ53 tʃiŋ53 ku53
then fall+descend elephant CL frying:pan
‘He [the mouse] threw the pancake, and it fell on the elephant’s frying pan.’

109. 佢 抓 燒餅 來 一 拋
kəj63 na23 liu53 pʰən3 liː1 tʃi3 pʰau53
3 take pan:cake come once throw
燒餅 就 跌落 地下
liu53 pʰən3 tʃʊə55 tʃi5 tʃi23 tʃa53
pan:cake then fall+descend ground
‘He [the mouse] took the pancake and threw it, and the pancake fell on the ground.’

4.2.3.2 Preverbal and Postverbal definite objects
The Southeastern languages have strong requirements for definite noun phrases to not exist after the main verb. (Indefinite noun phrases are usually, but not necessarily placed after the main verb.)

Fúqing Eastern Mín

110. 老板 買〇 車

* boss buy-PFV car

‘The boss bought a car.’

Fùyáng Wú

112. 個 老板 買得 部 車子

* CL boss buy-PFV CL car

‘The boss bought a(/the) car.’ (Li and Bisang 2012: 336)

113. 個 老板 部 車子 買得回來 喘

* CL boss CL car buy-PFV-return-hither COS

‘The boss bought the car.’

114. 我 去 放 兩 件 衣裳 得 大 衣架 裹

* 1SG go put several CL clothes to big closet inside

‘I put several clothes into the big closet.’

115. 我 兩 件 衣裳 去 放 放 得 大 衣架 裹

* 1SG several CL clothes go put put to big closet inside

‘I put the several clothes into the big closet.’
One counterexample in the Southeast is Wēnzhōu Wú, where \([\text{CL} + \text{N}]\) noun phrases mark definiteness by changing the tone of the classifier to tone D (tone 7 or 8), and such definite noun phrases can occur postverbally.

Wēnzhōu Wú

116. ŋ̍4ɕi3 ma4 paŋ7 si3
    I want buy \text{CLvolume} book
    ‘I want to buy a book.’ (Cheng and Sybesma 2005: 266)

117. ŋ̍4ɕi3 ma4 paŋ3 si3
    I want buy \text{CLvolume} book
    ‘I want to buy the book.’ (Cheng and Sybesma 2005: 266)

Classifiers which are already underlying tone 7 or 8 do not change their tones, and they can be interpreted as either definite or indefinite.

118. ŋ̍4ɕi3 ha7 y7/ lie7 thuə1
    I want drink \text{CLvolume/CLPL} soup
    ‘I would like to drink a bowl of soup/ some soup.’
    or: ‘I would like to drink the soup.’ (Cheng and Sybesma 2005: 266)

In other Sinitic languages, there is no requirements for definite noun phrases to not occur post-verbally. In Standard Mandarin, it is still quite common to prepose definite objects.

Standard Mandarin

119. 把 車子 買 了
    bǎ chēzi mǎi le
    OM car buy PRF
    ‘Bought the car.’
    ("把車子買了" on Google: 247,000 results; 12 Nov 2012)
    (The marked object is definite by default.)

120. 買 了 這 辆 車子
    mǎi le zhē liàng chēzi
    buy PFV this \text{CL} car
    ‘Bought this car.’
    ("買了這輛車子" on Google: 278,000 results; 12 Nov 2012)
With the Sinitic languages in the Central and Far-Southern zone, they allow definite noun phrases post-verbally. The Far-Southern Sinitic languages allow the preposing of objects much less readily than the Central Sinitic languages.

Shàowǔ Western Mín

121. 老板 買了 蜀 瑞 汽

\( lau^5\text{pan}^2 \text{ mie}^{33} \text{ cǐ} \text{ ka}^{35} \text{ tē\text{h}ia}^2 \)

boss buy-PFV one CL car

‘The boss bought a car.’

122. 老板 買了 ○ 蜀 瑞 汽

\( lau^5\text{pan}^2 \text{ mie}^{33} \text{ cǐ} \text{ kiōŋ}^3 \text{ ka}^{35} \text{ tē\text{h}ia}^2 \)

boss buy-PFV this one CL car

‘The boss bought this car.’

(Very Strange:

123. ?? 老板 拿 ○ 蜀 瑞 汽 買了

\( ?? lau^5\text{pan}^2 \text{ na}^{22} \text{ kiōŋ}^3 \text{ cǐ} \text{ ka}^{35} \text{ tē\text{h}ia}^{21} \text{ mie}^{33} \)

boss OM this one CL car bought-PFV

‘The boss bought this car.’)

Cantonese

124. 個 老板 買咗 瑞 汽

\( ka^{33} \text{ lou}^{13}\text{pan}^{25} \text{ maî}^{31}-\text{ts}^{25} \text{ ka}^{33} \text{ ts}^{13}\text{e}^{55} \)

CL boss buy-PFV CL car

‘The boss bought the/a car.’ (Li and Bisang 2012: 336)

("買咗瑞嘅" on Google: 43,900 results; 13 Nov 2012)

(Very strange:

125. ?? 個 老板 將 瑞 汽 買咗

\( ?? kô^{33} \text{ lou}^{13}\text{pan}^{26} \text{ tseŋ}^{55} \text{ ka}^{33} \text{ ts}^{13}\text{e}^{55} \text{ maî}^{31}-\text{ts}^{25} \)

CL boss OM CL car buy-PFV

‘The boss bought the car.’

("將瑞車買" on Google: 4 results; 13 Nov 2012))

\[37\] In both Cantonese and Western Mín, 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.
The Far-Southern Sinitic languages are like the other MSEA languages in not prohibiting postverbal definite objects. Below are examples of various preverbal and postverbal [classifier–noun] noun phrases all being definite in non-Sinitic MSEA languages.

Green Hmong

126. *khi tug dlev ces tug miv lug

\[\text{tie} \; \text{CL} \; \text{dog} \; \text{[and.then]} \; \text{CL} \; \text{cat} \; \text{come}\]

‘Tie up the dog and subsequently the cat will come!’ (Li 1989: 122)

Ong Be

127. *lar^33 vən^55 ho^33 sar^33tsu^33 biar^33 tuar^55 hu^55 uk^5 mia^55,

\[\text{exist} \; \text{day} \; \text{one} \; \text{rich:man} \; \text{release} \; \text{goat} \; \text{CL} \; \text{out} \; \text{come} \]

\[\text{ma}^13 \; \text{hu}^55 \; \text{[...]} \; \text{huk}^3 \; \text{tuar}^55 \; \text{hu}^55 \; \text{dai}^13 \; \text{voi}^33. \]

\[\text{dog} \; \text{CL} \; \text{make} \; \text{goat} \; \text{CL} \; \text{die} \; \text{FP}\]

‘[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)

4.2.3.3 Word order in clauses with three place predicates

MSEA languages in general have fewer instances of double object constructions. In the most extreme case, Enfield (e.g. 2007: 355–382) argues that there are no ‘real’ double object constructions in Lao. Some ways to avoid having two unmarked objects after the main verb in Lao are eliding an object, topicalising an object, put them in a serial verb construction (e.g. the ‘take’ serial verb construction), or make one of the objects an oblique object. The main point is that the prohibition is towards having two unmarked objects after the verb; it is not a prohibition towards having more than one constituent, as you could sometimes have both an unmarked object and a preposition-marked object after the verb.

Lao

Noun incorporation (not ‘real’ double object construction)

128. laaw2 thaa2 sii3 hùan2 lang3 nii4

3SG.FAM apply paint house CL DEM

‘She painted (i.e., ‘applied paint to’) this house.’ (Enfield 2007: 357)

129. *laaw2 thaa2 sii3 lùam5 hùan2 lang3 nii4

3SG.FAM apply paint shiny house CL DEM (Enfield 2007: 357)

Topicalisation

130. hùan2 lang3 nii4 laaw2 thaa2 sii3 lùam5
house CL DEM 3SG.FAM apply paint shiny
‘This house, she applied shiny paint (to).’ (Enfield 2007: 358)

131. sii3 lùam5 laaw2 thaa2 hùan2 lang3 nii4
paint shiny 3SG.FAM apply house CL DEM
‘Shiny paint, she applied (to) this house.’ (Enfield 2007: 358)

Serial verb construction
132. laaw2 qaw3 sii3 lùam5 thaa2 hùan2 lang3 nii4
3SG.FAM take paint shiny apply house CL DEM
‘She took shiny paint (and) applied (it to) this house.’ (Enfield 2007: 358)

Oblique strategy
133. laaw2 thaa2 hùan2 lang3 nii4 duaj4 sii3 lùam5
3SG.FAM apply house CL DEM with paint shiny
‘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 this being resolved by a ‘take’ serial verb construction.

Khmer
134. * kɔɔt ha:l khaoʔa:v thjay
he expose clothes sun

135. kɔɔt yɔɔk khaoʔa:v tɛu ha:l thjay
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 rather strong dispreference of having two constituents after the verb; one of the objects has to be placed in front of the verb somehow.

Hui’än Southern Mín

Theme topicalised
136. 伊 一 墨 筆 與 我
î tsi6>4 tsaiŋ>4 pet7 kɔɔ5 ua
3SG one CL 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. Notice that in Hui’ān Southern Min even indefinite noun phrases are often topicalised.)

Theme topicalised, plus object marking construction with a resumptive pronoun

137. 我 冊 共 伊 ○ 咋 桌 咋
   ua³ tsʰeʔ⁷ ka⁸ i¹ hio⁷>⁴ leʔ⁸>⁸ tɔʔ⁸ leʔ
1SG book OM 3SG put at table LOC
‘I put the book on the table.’
(The additional object marking construction is optional.)

Fùyáng Wú

138. * 伊 親得 我 一 口
   * fii ’wʰin-lə ɤɤ iə? kʰiu
3SG kiss-PFV 1SG one CL_mouth
(親 ’wʰin ‘kiss’ is a three-place predicate in Fùyáng Wú)

Passivised

139. 我 撥 伊 親得 一 口
   ɤɤ pəʔ? fii ’wʰin-lə iə? kʰiu
1SG PASS 3SG kiss-PFV one CL_mouth
‘I was kissed by him/her once.’

Object marking construction

140. 伊 ○ 我 親得 一 口
   fii kʰəʔ ɤɤ ’wʰin-lə iə? kʰiu
3SG OM 1SG kiss-PFV one CL_mouth
‘S/he kissed me once.’

In Wāxiāng, the most commonly used ditransitive construction involves a marked indirect object placed in front of the main verb.

Wāxiāng

141. 就 跟 它 放 到 ○ 裹
   tɕi’u⁵ʔ ka⁵ʔ tʰu⁵ʔ fyy³ tɔu³ pi³ la²⁵
then OM 3SG place to jar in
‘[...] then put it in the jar.’
On the other hand, the Far-Southern Sinitic languages are relatively free in having two constituents after the main verb, like the core of MSEA.

Cantonese

143. 佢 吊咗 我 一 吻
   \(k^b\)ey\(^{ij3}\) sek\(^2\)-tsɛ\(^{25}\) īj\(^{13}\) jet\(^6\) tam\(^{22}\)
   3SG kiss-PFV 1SG one CL
   ‘S/he kissed me once.’

144. 佢 界咗 呢 藥 我
   \(k^b\)ey\(^{ij3}\) pei\(^{25}\)-tsɛ\(^{25}\) tī\(^{55}\) jėk\(^3\) īj\(^{13}\)
   3SG give-PFV CL;MASS medicine 1SG
   ‘S/he gave me the/some medicine.’

145. 個 阿婆 收埋咗 呢 錢 嗨 檔桶 底
   ka\(^{33}\) a\(^{33}\)p\(^{h3l}\) suw\(^{55}\)mai\(^{ij3}\)-tsɛ\(^{25}\) tī\(^{55}\) tsʰin\(^{25}\) hvr\(^{25}\) kʰvr\(^{22}\)tuŋ\(^{25}\) tvr\(^{25}\)
   CL o.woman hide-PFV CL;MASS money at drawer under
   ‘The old woman hid the/some money under the table.’

Nánning Pínghuà

146. 佢 系了 的 藥 系 我
   ka\(^{i3}\) hvr\(^{25}\)-lu\(^{33}\) tık\(^5\) jėk\(^{23}\) hvr\(^{25}\) īj\(^{13}\)
   3SG give-PFV CL;MASS medicine DAT 1SG
   ‘S/he gave me the/some medicine.’

147. 你 放了 我 個 手機 住 那的?
   na\(^{i3}\) fuy\(^{55}\)-lu\(^{33}\) īj\(^{13}\) ko\(^{55}\) kəu\(^{33}\)koɾ\(^{53}\) tfəɾ\(^{22}\) na\(^{33}\)tık\(^5\)?
   2SG put-PFV 1SG MOD mobile.phone at where
   ‘Where did you put my mobile phone?’

Hainanese (Hǎinán Mín)

148. 我 分 蹤 〇 冊 (至) 伊
   guₐ\(^{2i}\) bʊn\(^{44}\) dziak\(^{13}\) bʊt\(^{4i}\) sɛʔ\(^{25}\) (tⁱ\(^{1}\)) t⁴
1SG  give  one  CL  book  to  3SG
‘I gave a book to him.’ (Lee 2011: 502-503)

The following are examples showing some non-Sinitic MSEA languages not having problems with having two constituents after the main verb.

Ong Be
149.  beu\textsuperscript{33} jua\textsuperscript{33} nə\textsuperscript{31} hiu\textsuperscript{55} (jou\textsuperscript{33}) hau\textsuperscript{55}
deliver  clothes  that  CL  to  1SG
‘Pass me that shirt/ Pass that shirt to me.’ (Liú 2009: 35)

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

151.  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)

The following table summarises the clause-level word order traits discussed in this §4.2.

|           | Thai | Khmer | Vietnamese | E Kayah Li | Burmese | Hmong Njua | Men | Cantonese | Nánging Pínghuá | Huí\textsuperscript{ān} S Mín | Fúqīng E Mín | Shāowù W Mín | Yèchăn Găn | Fúyáng Wú | Jīxī Huí | Gǔzhèng Wáxīng | Suīnīng C Mán\textsuperscript{ān} |
|-----------|------|-------|------------|------------|---------|------------|-----|-----------|----------------|-----------------|----------------|----------------|-----------|-----------|----------|-----------|-----------|-----------|
| 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 | +  | +  | +  | +  | -  | +  | +  | +  | +  | -  | +  | +  | +  | -  | ?  | +  | +  | ?  | -  | ?  |
In the table above, the absolute value of the total scores have little significant, 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 does show the relative difference in left-headedness on the clause level amongst the Sinitic languages, with the Far-Southern Sinitic languages being relatively left-headed, more similar to the core of MSEA, and the Southeastern Sinitic and Northern Sinitic languages being relatively right-headed, more similar to North Asia.

5. Conclusions and discussions

In this paper we have discussed some of the phonological and word order traits in the Sinitic languages. The Far-Southern Sinitic languages (roughly Yuè, Pínghuà, Hakka in Guǎngdōng, the Sinitic languages of Hǎinán and Léizhōu) are the most similar with the core of MSEA: highly tonal, conservative with codas, and relatively normal as SVO languages. In terms of word order, the least SVO-like languages are surprising not the Northern Sinitic languages, but the Southeastern Sinitic languages (roughly Mín, Wú and Huí). The strong prevalence of verb-final clauses in the Southeastern Sinitic languages is primarily an internal
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 5 above) sometimes show the Far-Southern Sinitic languages as having higher scores of ‘MSEA-ness’ than other Sino-Tibetan languages like Burmese and Southern Mín. They do not indicate that the Far-Southern Sinitic languages are more MSEA-like than these other Sino-Tibetan languages: they only indicate that the Far-Southern Sinitic languages have some traits that are more MSEA-like than Burmese and Southern Mín. There are many other typological traits, for instance grammaticalisation pathways, which would indicate the strong link from the core of MSEA to languages like Burmese and Southern Mín (see, e.g., Matisoff 1991, 2001a). What this paper is trying to argue is that, just as there are 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 ‘frindge’ as Northern Mandarin. Studies of the MSEA linguistic area would benefit immensely if data from the Southern Sinitic languages are always included in the MSEA linguistic area.

Abbreviations

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. Other people: John Phan, others. 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).

Bibliography


Djamouri, Redouane. forthcoming. Tàngwàng Language.


Zhèngzhāng, Shàngfāng 郑张尚芳. 1999. Báiyǔ shǐ Hán–Báí Yǔzú de yǐzhī dúlǐ 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, Fāng 石锋 and Wùyǔn Pān 潘悟云 (eds.). *Zhōngguó Yùyánxué de Xīn Tuōzhān 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.


