

Copulas and the Class of Copular Constructions in a Cross-Sinitic Perspective

Dawei Jin

Shanghai Jiao Tong University

This paper motivates the claim that the copula verb *xi* should be treated on a par with the empirically better described copula verb *shi* in Sinitic syntax. Based on fieldwork, the constructions in which the copula appears are investigated for a set of five Sinitic languages. The comparison demonstrates that the copular clause construction and the cleft construction inter-correlate in a robust way. This lends preliminary support to a theory positing an underlyingly identical syntactic structure for both constructions. Furthermore, the availability of particles in cleft constructions warrants a reexamination of treatments on cleft exhaustivity in Sinitic semantics.

1. Xi in Sinitic languages: the reinstatement of a marginalized copula

Yue-Hashimoto (1993), drawing upon Wang (1940)'s pioneering observation, articulates that Sinitic languages fall into two groups based on the copulas being employed: Cantonese and Hakka use a *xi*-type copula, that is, a cognate of the Classical Chinese copula verb *xi* 係. In contrast, the other Sinitic languages employ a cognate of *shi*. This purported restricted distribution of the *xi*-copula led many authors to assume that *xi* did not develop fully in Classical Chinese, and its current productivity in Cantonese/Hakka is a case of language-specific innovation.

Some evidence has been proposed with regard to this observation. For instance, the Gan language, which has close affinity with Hakka, is argued to be a *shi*-type language (e.g. Li & Zhang 1992; Tang 2009). Tang (2009) draws upon Li & Zhang's (1992) fieldwork survey and claims that an opposition obtains between Hakka and Gan in negative copular clause constructions: *m-he* 'NEG-XI' constructions occurs exclusively in varieties of Hakka. By contrast, negative copular clauses with negative morpheme *bat* or *mao* are found in Gan varieties. The conclusion Tang draws is that the Hakka *xi*-copula is most likely a recent innovation that arises due to close contact with Cantonese. Combined with the fact that *xi* is close to defunct in modern colloquial Mandarin, Tang concludes that the Middle Chinese copula verb *xi* clearly lost out in a competition with the other copula verb *shi* in all descendant languages (i.e. Sinitic) but the Cantonese/Hakka group.

Zhang & Tang (2011) further argue that the copula *xi* is grammatically more impoverished. They claim that the syntactic distributional environment of Cantonese *hai* is more restricted than that of Mandarin *shi*. For one thing, *hai* may only be flanked by

two NP arguments, instead of other categories. Furthermore, the two flanking NP arguments tend to denote concrete (non-abstract) entities.

Partly due to these claims, the *xi*-copula has garnered very little attention, compared with the *shi*-copula. However, a truly watertight conclusion about the lack of *xi*-copula verb in Sinitic languages should await empirical investigations of Sinitic varieties with a large enough sample, or at least an exhaustive chronicling of all the major Sinitic languages, which to this date has not been done. In many field studies of Sinitic varieties, the copula verb is either ignored, or to the extent that it is reported at all, simply presumed to be of the *shi*-type. Thus, in Hui Chinese, the copula verb, phonemically transcribed as *ɛi*, is mentioned as the analog of Mandarin *shi*, even though it can be seen elsewhere that the Hui equivalent of *shi* has a distinct pronunciation (*si*).¹

2. Toward a geography of copula types in Sinitic languages

This paper sets about testing previous claims about the *xi*-copula. As a pilot study, the task is established such that one representative dialectal spot is designated for each of ten Sinitic languages.² A minimum of three native speakers is consulted for each spot (local residence, no immigration history). The copula in question is further cross-checked with the phonetic realization of the etymologically related morpheme that forms part of the compound meaning ‘connection, relation’ (i.e. the morpheme *xi* in *lianxi* or *guanxi* in Mandarin). The consultants’ self-reports are compared with recordings and online chat records to the extent available.³

The elicitation results constitute a first approximation towards a typology of Sinitic copula types, illustrated in Table 1. The place name given in the parenthesis stands for the representative dialectal spot of the relevant Sinitic language elicited. A practical transcription scheme is adopted in Table 1 and throughout this paper. The

¹ Yue-Hashimoto feels necessary to qualify her claim by noting that the Jiahe dialect of the Xiang language uses both *shi* and *xi* for the copula. She believes that language contact may play a role, as this dialect might be influenced by the neighboring Hakka or Cantonese dialects. However, she still believes that there are differences of stratum, and the *shi*-copula forms the substrate for the Jiahe dialect. My general survey, on the other hand, suggests a much broader distribution of the *xi*-copula.

² Sinitic languages have been argued to number between ten to thirteen or fourteen (Norman 1988; Tang & van Heuven 2007; Handel 2015). My classification is based on the conservative view adopted in *The Chinese dialect atlas* (1987) and Ethnologue’s *Languages in China* (18th edition).

³ My data combine oral corpora of spontaneous speech (several hours of conversation and storytelling recordings during my fieldwork) with elicited native speaker reports. Such self-built corpus is then manually phonemically transcribed and parsed. In each corpus, I manually exclude non-syntactic tokens of the copula morpheme (in most cases, this means occurrences of the copula morpheme as a component of compound words, e.g., *dan-shi* ‘however’, *yu-shi* ‘then’). The remaining occurrences are further categorized into several construction types analyzed below.

Mandarin data are transcribed using the standard pinyin system. The Cantonese data are transcribed using the Eitel Cantonese Romanization scheme, one of the official Romanization schemes used in Cantonese-speaking regions. Broad phonemic transcriptions are used for other Sinitic languages, given the lack of unanimous Romanization programs to follow.

Type of copula	Language (dialectal spot): copula form
shi-type	Jin (Pingyao): <i>shi</i>
	Mandarin (Beijing): <i>shi</i>
	Min (Southern Min: Quanzhou): <i>si</i>
	Wu (Shanghai): <i>si</i>
	Xiang (Xiangtan): <i>si</i>
xi-type	Cantonese (Hong Kong): <i>hai</i>
	Gan (Fuzhou): <i>ɛi</i>
	Hakka (Wuhua): <i>hɛ</i>
	Hui (Wuyuan): <i>ɛi</i> (<i>xi-type</i>)/ <i>si</i> (<i>shi-type</i>)
	Ping (Binyang): <i>hai</i>

Table 1 *Grouping of Sinitic languages by copula type*

Apart from Cantonese and Hakka, Fuzhou Gan, Wuyuan Hui and Binyang Ping also use the *xi*-copula consistently.⁴ Furthermore, the observation by Tang (2009) that languages such as Gan have no negative copular constructions is also not supported by my fieldwork findings, where all language varieties that bear a *xi*-copular clause allow modification of the *xi*-copula by a negation verbal modifier (e.g. Ping: *mou-xai*; Gan: *baʔ-ɛi*; Hui: *pu-ɛi*).

Figure 1 illustrates the geographical distribution of *shi*-type languages versus *xi*-type languages (gray: *shi*-languages, red: *xi*-languages).

⁴ For example, in Fuzhou Gan, only 2 tokens using *si* are identified in recorded speech, against an aggregate of 24 *xi*-tokens. Both tokens take the form of the phrase *zun si*, where the copula *si* is modified by a quantificational adverbial *zun*. Importantly, *zun* sounds archaic to the native Gan speakers I consulted. The more colloquial alternative adverbs are *toi* ‘all’.

JIN: Xi in Sinitic languages

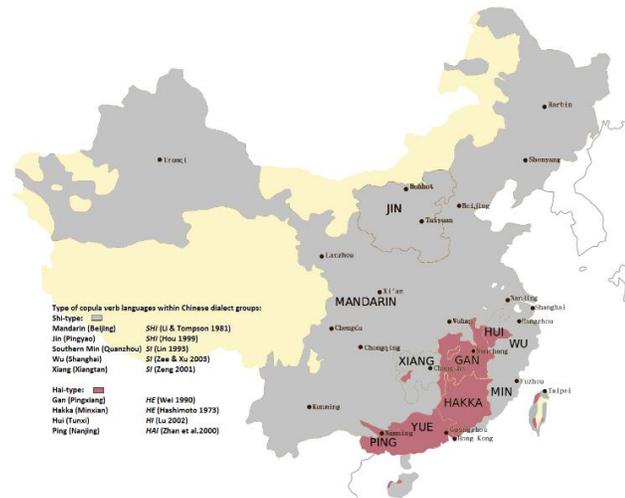


Figure 1 *Geography of copulas*

The current pilot study thus suggests that while Sinitic languages indeed exhibit an opposition between a *shi*-type subgroup and a *xi*-type subgroup, the distribution of the *xi*-subgroup is not limited to Cantonese and Hakka, but is broader than previously assumed. The geographical distribution identified in this study, where all *xi*-languages are mutually adjoining, is indicative of an areal feature that results from language contact. While the identification of a linguistic area characterized by the use of the copula *xi* does not invalidate the postulation of *shi*-languages' predominance, it does lead us to modify the assumption that the Middle Chinese copula *xi* has all but disappeared from most Sinitic languages except for a residual use in Cantonese, or that the copula property of *xi* is a Cantonese-specific innovation. It seems more plausible to assume that the *xi*-copula was fully productive in Middle Chinese up to its split into its daughter languages. This productivity was retained in the South Sinitic languages that maintained close contact with one another, yet was gradually lost in the North Sinitic languages.⁵ In short, it seems that modern daughter languages of Classical Chinese opt for one single copula morpheme, while the other copula morpheme inherited from the ancestor language tends to be demoted. At present, this finding highlights the need to put the *xi*-copula on an equal footing with the *shi*-copula in typologically-oriented studies on comparative dialectal grammar. Thus, the investigation below serves the empirical purpose of chronicling for the first time the syntactic distribution of the constructions where copula *xi* occurs from a comparative and typological perspective. Apart from empirical reasons to investigate the

⁵ Alternatively, it is also possible to think of the *xi*-copula as of indigenous origin in South China. That is, the *xi*-copula was initiated in South Sinitic languages and was subsequently spread to North China. The northbound *xi*-copula lost out in competition to the more prevalent *shi*-copula, but remained productive in regions of its origin. At present, I am aware of no good way to determine between these two scenarios. A concerted effort, combining textual/archival research and fieldwork at a micro-level, is needed to shed light upon this distribution pattern in the future.

xi-type languages, valuable insights can be drawn from a comparative perspective on the distributional environments of constructions bearing the copular morpheme *xi* vis-à-vis *shi*. As both copulas evolved independently, comparing their distribution in modern daughter languages enables us to tease apart etymological and historical coincidence, and explore to what extent variation exhibits itself within the two language groups. Consequently, it also enables us to formulate certain hypotheses regarding the structural relation between copula-like constructions.

3. Toward a first approach to the distribution of the copula-class constructions

In this section, I investigate the syntactic distribution environment of Mandarin and four *xi*-type Sinitic languages (Hakka is left out due to lack of access to data). The first construction type involves the copular clause construction exemplified by copulative sentences such as the following.⁶

(1) a. *Cantonese predicational copular clause*

kui hai kingkek ke yattoi tsungsi.
 he COP Peking.Opera REL a.generation giant

‘She said: He is a giant in his generation of Peking Opera.’

b. *Cantonese specificational copular clause*

Nei yeungyeung tau ho: patkwo ngo tsungyi ke hai Tongtong.
 You everything PRT good yet I love REL COP Tongtong

‘Everything about you’s good, but the girl I am in love with is Tongtong.’

A more accurate characterization would be to treat copular clauses as a family of mutually related constructions. Thus, (1a) illustrates a predicational copular clause, in which the pre-copula subject is entity-denoting, and the post-copula predicational complement denotes a property that is applied to the subject (Higgins 1979; Mikkelsen 2005). In (1b), the pre-copula subject is property-denoting, predicated of the post-copula referential complement.

The homogeneity exhibited between the *shi*-type and the *xi*-type Sinitic languages is unsurprising, given that the family of copular clause constructions is generally analyzed to be, in Construction Grammar terms, underlied by a single overarching copular clause construction type, and, in generative terms, a family of transformationally derived structures.

Aside from copular clauses, the second syntactic distributional environment of the *xi*-copula involves the (term) cleft construction. In keeping with much previous syntactic

⁶ Glossing in this paper follows the Leipzig Glossing Rules, including the following abbreviations:

ADV: adverbial, COP: copula, FOC: focalizing particle, CLF: classifier, DECL. PRT: declarative particle, DIST: distributive operator, EMP: emphatic morpheme, NEG: negation morpheme, NOM: nominalizer, PASS: passive morpheme, POSS: possessive, PRT: particle, REL: relativizer.

literature on clefting in Chinese syntax, a cleft construction in Sinitic languages is identified as a structure that includes a linear sequence of two post-copula parts (Shyu 1995; Simpson & Wu 2002; Cheng 2008; Paul & Whitman 2008; Hole 2011). The immediately post-copula clefted phrasal constituent (an argument or adjunct) encoding information-new sentential focus is followed by an open sentence (predicate) that encodes information-old backgrounded content. Cantonese example is provided below.⁷

- (2) a. *Cantonese argument-cleft*
 Zhe-jian shiqing [shi Zhangsan] fuze.
 This-CLF affair COP Zhangsan in.charge,
 ‘It is Zhangsan who is in charge of this matter.’
- b. *Cantonese adjunct-cleft*
 Zhangsan [shi zuotian] lai de.
 Zhangsan COP yesterday come DE
 ‘It is yesterday that Zhangsan came.’

A further characteristic of the cleft construction is the optional presence of a pre-copula constituent that functions as frame-setters or discourse topics (Hole 2011). The topical status can be demonstrated via the attachment of a discourse-level suffixal particle to the pre-copula constituent, which is independently shown to be unacceptable when the host constituent is non-topical (Xu 2000; Xu & Liu 2007; Constant 2014). A final diagnostic of clefts has to do with the exhaustive nature of focus semantics, illustrated among other things by the incompatibility with an additive reading. In (3), a continuation involving an additive adverb *jitdou* ‘also’ leads to unacceptability, which can be accounted for if the cleft sentence requires an exhaustiveness interpretation.

- (3) *Cantonese*
 #Keoidei hai camjat tai dinjing, gamjat keoidei jitdou tai dinjing.
 They COP yesterday watch movie, today they also watch movie
 #‘It was yesterday that they watched a movie. They also watched a movie today.’

In Mandarin, clefting strategy may be achieved with what is termed a bare *shi*-cleft, schematized as [(*topic*)+*shi*+*clefted constituent*+*open sentence*] without any overt particle attached to the open sentence. Alternatively, the clefting strategy in Mandarin

⁷ Unlike English or other Sino-Tibetan languages such as Burmese, cleft constructions in Chinese do not involve overt cleaving, in that the copula morpheme does not overtly partition the focused constituents (e.g., *Zhangsan* in 2a) from the backgrounded materials (e.g., *fuze* ‘to take charge’ in 2a) (cf. Erlewine 2016). However, I follow Hole (2011) and Hole & Zimmermann (2013) in assuming that languages vary in whether the partitioning between focused and backgrounded materials is achieved in overt or covert syntax, with Sinitic languages falling into the latter category.

may be achieved by a periphrastic [*shi* V O *de*] construction. Furthermore, in Northern varieties of Mandarin, a periphrastic construction is often used in parallel with the above-mentioned [*shi* V O *de*] construction (that is, the particle *de* is placed in between the verb and its object NP).

- (4) a. Mandarin [*shi* V O *de*] construction
 Zhangsan shi zuotian lai **de**.
 Zhangsan COP yesterday come DE
 ‘It is yesterday that Zhangsan came.’
 b. Mandarin [*shi* V de O] construction
 Wo shi xie-**de** shi.
 I COP write-DE poems
 ‘It is poems that I wrote.’

In Mandarin clefts, the presence of the *de*-particle has been argued to be the locus of exhaustivity. A characteristic of Sinitic languages of the *xi*-type is that analogs of the Mandarin *de*-particle is omissible where *de* is obligatory in Mandarin. First, an intervening of the *de*-like particle in between a predicate and its object is disallowed in these Sinitic languages. [*hai* V *ge* O] is not possible, where *ge* functions similarly with Mandarin *de*-particle (Tang 1998; Cheung 2007; Tang 2005; Wakefield 2010; Matthews and Yip 2013). Compare the Cantonese data in (5a) with the Mandarin data in (5b):

- (5) a. Cantonese
 *Keoidei hai cammann tai **ge** bo.
 They COP last.night watch GE ball.game
 ‘It was last night that they watched the game.’
 b. Mandarin
 Tamen shi zuowan kan **de** qiu.
 They COP last.night watch DE ball.game
 ‘It was last night that they watched the game.’

Second, sentence-final *ge* or *gaa* (a fused morpheme of *ge* and *aa*) is witnessed, but unlike Mandarin their appearance is optional, and omission of particles still maintains acceptability of the cleft sentence, illustrated by the following contrasts.

- (6) a. Cantonese
 Keoidei hai camjat tai dinjing {*ge/gaa*}.
 They COP last.night watch movie {GE/GAA}
 ‘It was last night that they watched movies.’
 b. Mandarin
 Tamen shi zuowan kan dianying *(de).

- (7) a. *Cantonese*
 They COP last.night watch movie *DE
 ‘It was last night that they watched movies.’
 Toifung fungkau hai jau tinmantoï fatfung {ge/gaa}.
 Typhoon wind.ball COP by Observatory release {GE/GAA}
 ‘It was by the Observatory that the typhoon wind ball was released.’
- b. *Mandarin*
 Taifeng fengqiu shi you tianwentai fafang *(de).
 Typhoon wind.ball COP by Observatory release *DE
 ‘It was by the Observatory that the typhoon wind ball was released.’

Examples in (8) involve a different type of construction containing the copula morpheme *xi*. Instead of a partition between informationally prominent (focused) and backgrounded components, an entire clausal argument, which follows the copula morpheme, is conveyed as new information.

- (8) a. *Mandarin*
 Ta mei gen wo dazhaohu, wo juede shi ta mei renchu wo lai.
 He NEG-PRF with me greet, I think COP he NEG-PRF recognize me out
 ‘He didn’t greet me, do you think it’s that he failed to recognize me?’
- b. *Ping*
 na mou-jou hat-løk hak tankun, kø nø tsaktak ei na mou touŋø kua
 He NEG-PRF eat-finish box egg.roll, then I think COP he NEG hungry MOD-PRT
 ‘He didn’t finish that box of egg roll, and so I think it might be that he wasn’t hungry.’

Statements such as (8) function as a propositional assertion (Hole 2011), in which the copula marks the entire proposition denoted by the clausal argument as focused material. Several diagnostics can be drawn upon to motivate the case that in (8), the entire post-copula clause represents a new propositional assertion. For instance, (8a) is felicitously uttered as an answer to the QUD: *why didn’t he greet you?* The post-copula clause is construed as a propositional answer to the *why*-question (i.e. construed as an implicit *because*-clause as explanation).⁸ Next, if A’s answer is negated by another speaker, as in B’s answer. The negation has to be construed as negating the asserted content, which provides another way to diagnose what is asserted. As we can see, B’s continuation is felicitous if the negation is followed by an alternative explanation of the QUD, demonstrating that what is asserted in A’s answer is the entire propositional

⁸ I assume Belnap’s (1969) view that the answer of a *why*-question relates a proposition to another proposition that serves as the former’s explanation. That is, it encodes a binary causal relation between propositions (see also Scheffler 2005).

content as an explanatory because-clause. If B's negation is followed by replacing him with another individual (9), the continuation becomes infelicitous.

(9) *Mandarin*

QUD: Weishenme ta mei gen ni dazhaohu?

A: Xiao Zhang mei gen wo dazhaohu, wo juede shi ta mei renchu
 Xiao Zhang NEG-PRF with me greet, I think COP he NEG-PRF recognize
 wo lai.
 me out

'Xiao Zhang didn't greet me, I think it's that he failed to recognize me.'

B: Ni cuo-le. Ta tai congmang-le.

You be.wrong-PRF. He too be.hurry-PRF

'You are wrong. (It's that) he has been too much in a hurry.'

B: #Ni cuo-le. Xiao Wang mei renchu ni lai.

You be.wrong-PRF. Xiao Wang NEG recognize you out

'You are wrong. (It's that) Xiao Wang didn't recognize you.'

In addition, a copula may mark a predicate denoted by an open sentence as newly asserted material. In this construction, a topic-suffixed constituent precedes the copula, and unlike proposition assertion this constituent serves a frame-setting function and is informationally separated from the rest of the proposition (the open sentence) by encoding information already familiar to the interlocutors.

(10) *Gan*

ŋo [ei t^hiŋ p^hjɛʔŋinka wasi], ŋo tsikan puʔ ɛjɛwtɛʔ laŋ pan.

I COP follow others decide, I self NEG know how act

'What I do is listen to others giving orders. I have no idea how to do this.'

The frame-setting function of the pre-copula constituent is made clear in a typical context where its referent is already salient in immediately prior discourse. Similar to propositional assertion, we can diagnose that the post-copula predicative part represents what is asserted, as it is the content that negation targets.

(11) *Mandarin*

A: Benlai-ne Xiao Zhang shi dasuan liu zaijia.

Initially-TOP Xiao Zhang COP plan stay at.home

'Initially, Xiao Zhang was planning to stay at home.'

B: Ni cuo-le. Ta dasuan qu KTV wan.

You be.wrong-PRF. He plan go KTV play

'You are wrong. He was planning to go to a KTV.'

B: #Ni cuo-le. Xiao Wang dasuan liu zaijia.

You be.wrong-PRF. Xiao Wang plan stay at.home

‘You are wrong. Xiao Wang was planning to stay at home.’

Another environment a copula may appear in involves what is termed by the literature as the A-not-A question: the copula occurs in a reduplicative form, with a negative morpheme inserted in between two identical copies of the copula. In resemblance to a term cleft construction, an A-not-A question may similarly mark a term focus that appears immediately after the reduplicative copula-negation-copula form, shown in (12).

- (12) a. *Cantonese*
 Toifung fungkau hai-m-hai jau tinmantoï fatfong?
 Typhoon wind.ball COP-NEG-COP by Observatory release?
 ‘Is it the case that the typhoon wind ball is released by the Observatory?’
- b. *Mandarin*
 Zuowan neixie dianhua xi-mao-xi ni da gei wo de?
 Last.night those phone.calls COP-NEG-COP you call to me DE?
 ‘Those phone calls last night, is it the case that you made them to me?’

What is more, both predicate-level assertion and proposition-level assertion find corresponding reduplicative A-not-A questions, demonstrated as follows.⁹

(13) **Proposition assertion**

⁹ The question of whether the A-not-A question where the copula is reduplicated is syntactically derived from a declarative assertion sentence is subject to controversy, as A-not-A questions do not always converge with assertion sentences. For example, positive assertion is not felicitous where a reduplicative question is felicitous.

- (i) a. ?Shi ta genben jiu bu xiang bang ni.
 COP he definitely PRT NEG want.to help you
 ‘It’s that he definitely doesn’t want to help you.’
- b. Shi-bu-shi ta genben jiu bu xiang bang ni?
 COP-NEG-COP he definitely PRT NEG want.to help you
 ‘Is it that he definitely doesn’t want to help you?’
- Secondly, reduplicative question infelicitous when a sentence-final particle is attached:
- (ii) a. Ta shi xiang bang ni de.
 He COP want.to help you DE
 ‘It is the case that he wants to help you.’
- b. ?Ta shi-bu-shi xiang bang ni de?
 He COP-NEG-COP want.to help you DE
 ‘Is it the case that he wants to help you or not?’
- Moreover, reduplicative question allows for multiple occurrence of *shi*:
- (iii) Shi-bu-shi ta shi zhunbei liu xialai de?
 COP-NEG-COP he COP plan.to stay down DE
 ‘Is it the case that he planned to stay?’

Cantonese

Keoi mou tung ngo daziufu, nei lam hai-m-hai keoi mei jingceut ngo lai?
 He NEG with me greet, you think COP-NEG-COP he NEG-PRF recognize me out
 ‘He didn’t greet me, do you think it’s that he didn’t recognize me?’

(14) **Predicate assertion**

Gan

Nei ei-pu?-ei moŋ paŋ ŋo pan tan koi-te^hjen si ko?
 You COP-NEG-COP hope.to help me achieve DEM-CLF matter PRT?
 ‘Is it the case that you are willing to help me accomplish this matter?’

The rest of this section deals with copula-class constructions that exhibit variation among Sinitic languages. In Mandarin, Hui and Gan, the copula allows for what is termed by Li & Thompson (1989: 151-154) as the emphatic construction. In this use, the copula precedes a predicate and the predicate must encode a piece of familiar information that appears in prior discourse. Thus, sentence (15) must be uttered in a context that affirms what has been said earlier or what has been suspected or inferred by the interlocutors (example provided in Li & Thompson 1989).

(15) *Mandarin*

A: Wo xiang ta hen qiong, suoyi bu ken shang guanzi.
 I think he intensifier be.poor, so not willing.to go.to restaurants.
 B: Ta shi mei qian, keshi you zhiqi.
 He EMP not.have money, but have principles
 ‘A: I thought he got no money, so he wouldn’t dine out.
 B: It’s true that he got no money, but he got his pride.’

The surface distribution of the emphatic construction resembles a *shi*-introduced predicate assertion, however several behavioral characteristics serve to distinguish between these two distinct uses. First of all, emphatic *shi*-construction carries a special prosodic pattern, with a primary stress associated with the morpheme *shi*. In contrast, *shi* in predicate assertion is never stressed, whereas prosodic prominence is spread across the post-copula predicate (open sentence) (Wang 2011).

Secondly, prior familiarity is explicit or presumed. As (16) indicates, when the prior content under discussion cannot be ascertained, a *shi*-construction only receives a predicate assertion reading, and cannot be construed with an emphatic reading.

(16) *Mandarin*

A: Bu zhidao xingqitian tushuguan kaimen bu.
 NEG know Sunday library open NEG.
 B: Yinggai shi kaimen. Ni keyi wangzhan shang cha.
 Should COP open. You can website LOC look.up
 ‘A: Not sure if the library opens on Sunday or not.’

B: Should open. You can look it up on the library website.’

Thirdly, emphatic *shi* fits in with the behaviors associated with supplemental materials (Potts 2005), by exhibiting what is termed ‘scopelessness’, such that it fails to be interpreted within the scope of another scope-taking operator. For instance, it is not possible for an emphatic *shi* to be embedded under negation.

(17) *Mandarin*

#Bu shi hen taoyan.
NEG COP intensifier be.annoying

Intended: ‘It is not the case that indeed/truly she is annoying. /It is not indeed the case that she is annoying (It’s complicated./We cannot be positive yet).’

Note that the above sentence would become felicitous when interpreted as a simple assertion, in which it is denied that the annoying property is plainly predicated of said referent. In other words, while an assertive *shi* clearly allows itself to be embedded under negation, something special about an emphatic *shi* is blocking this scoping-under reading. Similarly, emphatic *shi* cannot be embedded under a modal operator. The following sentence only receives a predicate assertion reading, not an emphatic reading.

(18) *Mandarin*

Yinggai shi hen taoyan.

Possible reading: ‘It is possible that (she) is annoying.’

Impossible: ‘It is possible that it is indeed the case (truly) that (she) is annoying.’

In Hui, which employs both a *shi*-copula and a *xi*-copula. Importantly, *shi* occurs for an emphatic construction, whereas it is unacceptable to replace *shi* with *xi*, exemplified in the following contrast:

(19) *Hui*

A: i-phu tɛiθtɔtɕhie, ŋ pu ɕiɔtɔ θ iəm tsi tɕikɔ kɔtsĩ mɔ ke.
DEM-CLF bicycle, you NEG know I use PRF how.many price buy PRT

B: ɕi ɕiəpan tsi itɔŋ.
COP unworthy PRF a.bit

‘A: This bike, you wouldn’t guess how much I paid to get it!

B: True the price is a bit not worth it.’

Another construction, witnessed in Cantonese and Ping, expresses universal quantification. The copula morpheme, when followed by the maximality operator *dou*, yields a universal, ‘no matter what’ reading. As (20) demonstrates, the combination of the

copula morpheme and the *dou*-operator signals that the act or state denoted by the following predicate holds regardless of circumstances.

- (20) a. *Ping*
 mat na ei t^hoŋ mou sək tsou?
 how.come you COP PRT NEG know fault
 ‘How come he wouldn’t admit it is his fault no matter what!’
- b. *Cantonese*
 Nei hai dou jiu gik haa ngo sin hoisam!
 You COP PRT will irritate a.bit me then be.delighted
 ‘You would make me angry no matter what, and take delight in that!’

Finally, in Gan and Hui, *shi* appears as a conditional marker that is suffixed to an antecedent clause and connects it with the following consequent clause, as illustrated in the following Gan example.

- (21) *Gan*
 kɛ pu? t^hoŋji ei, koi-tɛ^hjɛn si tɛ^hju pan pu? tan tɛ.
 He NEG agree COP, DEM-CLF affair then achieve NEG RES PRT
 ‘If he does not agree (to that), we won’t be able to achieve that.’

Table 2 summarizes the distributional environments in Sinitic copula-class constructions.

	Mandarin	Cantonese	Gan	Hui	Ping
Copular clause	+	+	+	+	+
Term cleft	+	+	+	+	+
Propositional cleft	+	+	+	+	+
Predicate cleft	+	+	+	+	+
A-not-A question	+	+	+	+	+
Topic marker	-	-	+	+	-
Emphatic construction	+	-	+	+	-
Universal reading	-	+	-	-	+
Conditional marker	-	-	+	+	-

Table 2 A taxonomy of copula distribution in Sinitic languages

In sum, all Sinitic languages under survey converge in the pattern of copula distribution, in that they simultaneously occur in a family of (canonical) copular clause constructions, as well as focus-marking constructions such as term clefts (where the immediately post-copula term is focused) and broad assertions (where the entire post-copula clause or open clause is focused). Variation within these construction types lies in how liberal *de* and its analogs are licensed. A *shi*-type language (i.e. Mandarin) requires the presence of *de*-like particles for the exhaustivity reading to be available, whereas *xi*-type languages allow the absence of *de*-like particles while still retaining the exhaustivity reading. On the other hand, Sinitic languages diverge on a variety of constructions, i.e. constructions where the copula serves an emphatic, universal, conditional and topic-marking function.

4. Making sense

The different functions that the copula assumes in Sinitic languages have been independently observed to be crosslinguistically robust. Copulas tend to grammaticalize into dedicated topic markers, and vice versa (what is termed the ‘Copula Cycle’ by Lohndal 2009). Similarly, a robust bi-directional pathway obtains between a copula verb and a conditional clause marker, and speech act-level operators such as assertion particles tend to change into markers of affirmation and emphasizing (Heine & Kuteva 2002). It is thus plausible to assume that the different uses observed in this study are all to some extent related, and the multifunctionality of the copula is possibly developed from historical processes (e.g. reanalysis).

Importantly, though, it would be less than desirable to posit that the copular clause use and the emphatic use, which exhibits variation across the Sinitic languages, fall under a single, overarching lexical entry. Given that Mandarin and some *xi*-type Sinitic languages that do not feature an emphatic use of the *xi*-copula descend from the same common ancestor language, positing a single entry would commit us to theorizing that an innovation that takes place in Mandarin after the split has changed its parameter setting in a way that structurally unifies copular clause construction and emphatic construction. Since we have seen that in addition the topic marker construction, the conditional construction and the universal construction are witnessed in a subset of Sinitic languages, respectively, we would have to further posit that each Sinitic language may undergo its individual innovation so as to accommodate the variation. However, it would be unclear what independent evidence exists to justify these innovations. A more plausible solution would involve treating all these uses as distinct lexical entries. Given the semantic relatedness, the change from copular clauses to emphatic clauses represent a tendency that target certain Sinitic languages following the split, but not all the Sinitic varieties.

In contrast to this, things would be different given that the distribution of the following constructions exhibit no inter-language variation (regardless of *xi*-type and *shi*-type):

- (22) [copular clauses < --- > clefts < --- > broad < --- > A-not-A] (here < --- > is taken to mean that where the left side of the arrow is witnessed, the right side is also witnessed and vice versa)

The robustness of this pattern lends preliminary support for some structural homogeneity among said constructions. That is, clefts, broad clefts and A-not-A questions are underlyingly a copular structure. An alternative scenario, in which some neat reanalysis process applies indiscriminately to all these languages, is, albeit clearly possible, quite marked given the variation shown elsewhere.

Let us explore with some detail what a homogeneity analysis amounts to. In the Mandarin literature, propositional and predicate assertion have been treated as closely related to the informationally partitioned term cleft (referred to as ‘broad clefts’) (Cheng 2008; Paul & Whitman 2008; Hole 2011). Information-structurally, the bracketed CP brings discourse salience to a newly asserted proposition. Derivationally, Moro (1997) argues that *that John left* in (23a) resides in the same structural position within the copula *be*’s argument as a clefted phrase such as *John* in the *it*-cleft (23b).

- (23) a. (Speaker A realizes that Speaker B is upset and asks what is bothering him)
 B: It’s not that John and I argued, it’s [that John left].
 b. It’s [John] that left.

In this sense, the example in (23a) may be viewed as having a clefted CP. Huber (2006) further observes that Swedish and French allow more liberal uses of maximally focused clefts than English, expressing propositions in contexts not limited to discourse starters/framers. It thereby raises the possibility that Chinese proposition assertion sentences (broad clefts) form part of a continuum of maximally focalizable clefts.

Despite the lack of partition, previous authors have proposed that the copula in both constructions perform a uniform discourse strategy of asserting the immediately post-copular element as discourse-new focus information, with free focus assignment of variable scope. When an entire proposition falls within the scope of focus assignment, the relevant *shi*-clause expresses a propositional (or predicate) assertion, stated as an update against a background of non-stated prior knowledge. This differs from the strategy of term clefts, where only part of the proposition carries focus information, leaving the rest of the proposition overtly expressed as backgrounded. The uniform approach to both constructions is further assumed under an overarching structural treatment, according to which the propositional argument in broad clefts occupies the same structural correlate as the focused constituent of a smaller unit (an NP argument or an adjunct) in term clefts.¹⁰

¹⁰ For instance, Cheng (2008) proposes that a clefted CP patterns together with a regular cleft phrase (NP/adjunct), both merging at the subject position of the small clause argument of the copula. See also Paul & Whitman (2008) for arguments against a uniform analysis of both types of clefts.

Another piece of evidence comes from diachronic pattern. Based on diachronic studies on Early Mandarin corpora (16th century to 19th century), it has been shown that the use of the copula morpheme *xi* in copular clauses, clefts and broad clefts have vanished side by side during the transition from Early Mandarin to Modern Mandarin (Jin 2016; Chen 2017). The coordinated decline pattern would be unsurprising if these construction types share the same copular structure. For a focus-based analysis of clefts, nevertheless, the pattern would be mysterious, as the copula morpheme is treated as homophonous lexical entries distinct from one another, and a direct consequence is the lack of convincing reason to account for why the loss of distinct lexical entries should be closely correlated.

Let us briefly spell out what an alternative syntactic theory works like. The focus movement approach adopts a monoclausal analysis, in which a Chinese cleft is not headed by a copula verb. Rather, the copula moves with the focused phrase to the left periphery to check the [exhaustive] feature (Teng 1979; Zhu 1996; Erlewine 2016). Assuming a Rizzi-style articulated CP, it is argued that the focused phrase undergoes focus movement to [Spec, FocP] from its base position at FinP. One characterization of the copula morpheme's role during focus movement is that *shi* is syntactically an adverb analogous to the English adverb *only*. An adverb-like focus marker resides in the left periphery but simultaneously stays as closely to the focus it associates with as possible. Another possibility is that the copula morpheme initially merges at the head of the focus projection (Rizzi's FocP) and subsequently undergoes remnant movement to a projection structurally higher than FocP (e.g. TopP, cf. Frascarelli and Ramaglia 2009).

Apart from the empirical issue with the mutually entailing nature of copular clauses, term clefts and propositional assertions (as well as the coordinated decline pattern in Mandarin), an adverb-based approach shown above also faces the additional burden of accounting for the reduplicative A-not-A questions. To posit a bifurcation between a *bona fide* copula verb category in the case of copular clauses, and a focus-marking adverbial category in the case of clefts, such an account would seem to be committed to positing two types of reduplicative processes when both copular clauses and clefts appear in A-not-A questions.¹¹ Such assumption, however, does not seem to find any independent empirically-grounded motivations. A further issue is the plausibility of positing a reduplicative process that targets an adverbial element, as elsewhere only predicative elements (verbal and adjectival) are known to allow for reduplication. Note that this does not pose a problem if clefts feature a copula verb in Chinese.

Finally, the observation that Sinitic languages exhibit variation in terms of the obligatoriness of the *de*-like particle in clefts factors into the locus of exhaustivity reading

¹¹ As far as I know, focus-based approaches have not explicitly addressed how to incorporate propositional/predicate assertion into the division between copular clauses and clefts. One would assume that either these assertions are subsumed by a copular clause, or by a term cleft, or they project their own construction types. This problem (and whether an attempt to resolve it raises new theoretical challenges) will not be touched upon here.

in clefts. That is, it disfavors proposals (Hole 2011; Hole & Zimmermann 2013) in which exhaustivity is derived from the *de*-particle, which lexically encodes a meaning component of maximality. It seems more compatible with theories such as Cheng (2008), in which the particle modulates speech act (e.g. assertion) and expresses sentence mood. Exhaustivity is derived elsewhere, for example, by means of a maximal presupposition triggered from within the open sentence.

5. Conclusion

Studies on copulas and copular constructions from a cross-Sinitic perspective are still in an inception stage. As a result, the type of copula verb that each individual Sinitic language employs, as well as the extent to which Sinitic languages vary in the syntactic constructions a copula might appear in, remains a desideratum.

This paper thus fills in a noticeable lacuna, by mapping Sinitic languages to their respective copula types, and establishing that a group of four South Sinitic languages employ reflexes of *xi* in their copular constructions. This investigation is then followed by a pilot survey of the syntactic frames each *xi*-type language licenses, in comparison with the case of the *shi*-copula in Mandarin. The comparative syntactic work reaffirms the need to posit multifunctionality of the copula morpheme in the Sinitic language area. That is, it is plausible to assume that the copula morpheme represents homophonous (and historically/pragmatically interrelated) yet distinct lexical items. What's more, the comparative survey lends first support toward unifying several copular constructions as a family of assertion sentences with structural homogeneity. I have shown that the empirical picture is harder to accommodate given an approach that draws a distinction between a *bona fide* copula verb category and an adverbial category that occurs in clefts and cleft-like sentences. Finally, the comparison disfavors certain proposals that derive the focus exhaustivity reading of clefts from overt materials (i.e. sentence-final particles), by demonstrating that the presence of particles are fluid across languages.

While the present study looks into a variety of *xi*-type languages, it equates *shi*-type languages with Mandarin, thus a more comprehensive typological survey is needed in future research. It is hoped that this paper opens up a new line of research, and the syntactic variation unveiled enables more empirically grounded and more testable discussions over theoretical issues of the nature of copular-like constructions in Chinese syntax.

REFERENCES

- CHENG, LISA LAI-SHEN, 2008. Deconstructing the *shi... de* construction. *The Linguistic Review* 25 (3-4): 235-266.
- CHEUNG, HUNG-NIN. 2007. *Cantonese Grammar*. Hong Kong: The Commercial Press.

- DEN DIKKEN, MARCEL. 2006. *Relators and linkers: The syntax of predication, predicate inversion, and copulas*. Cambridge, Mass.: MIT Press.
- ERLEWINE, MICHAEL. 2016. The semantics of the Mandarin focus marker shi. Talk presented at the European Association for Chinese Linguistics 9, Stuttgart.
- FRASCARELLI, MARA, AND FRANCESCA RAMAGLIA. 2014. (Pseudo)clefts at the syntax prosody-discourse interfaces. *Cleft structures*, edited by Hartmann Katharina, and Veenstra Tonjes, 97–138. Amsterdam: John Benjamins Publishing Company.
- HANDEL, ZEV. 2015. The Classification of Chinese: Sinitic (The Chinese Language Family). In Wang, William & Sun, Chaofen (eds.). *The Oxford Handbook of Chinese Linguistics*, 34–44. Oxford: Oxford University Press.
- HIGGINS, RICHARD. 1979. *The pseudo-cleft construction in English*. New York: Garland.
- HIRAIWA, KEN, and SHINICHIRO ISHIHARA. 2002. Missing links: cleft, sluicing and ‘no da’ construction in Japanese. *MIT Working Papers in Linguistics* 43: 35-54.
- HOLE, DANIEL. 2011. The deconstruction of Chinese shi... de clefts revisited. *Lingua* 121.11: 1707-1733.
- HOLE, DANIEL, and MALTE ZIMMERMANN. 2013. Cleft partitionings in Japanese, Burmese and Chinese. *Cleft structures*, edited by Katharina Hartmann, and Tonjes Veenstra, 285-317. Amsterdam: John Benjamins.
- JIN, DAWEI. 2017. Toward a copular syntax of clefts: evidence from Chinese. Talk presented at the 2017 LAGB (Linguistic Association of Great Britain) annual Meeting.
- LI, CHARLES, and SANDRA THOMPSON. 1989. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- LI, RULONG, and SHUANGQING ZHANG. 1992. *Kegan fangyan diaocha baogao [A fieldwork report of the Hakka and the Gan dialects]*. Amoy: Xiamen University Press.
- LOHNDAL, TERJE. 2009. The Copula Cycle. *Cyclical Change*.
- MATTHEWS, STEPHEN, and VIRGINIA YIP. 2013. *Cantonese*. London: Routledge.
- MIKKELSEN, LINE. 2005. *Copular clauses*. Amsterdam: John Benjamins.
- MIKKELSEN, LINE. 2011. Copular clauses. In Maienborn, Claudia, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning, 1805-1829*. Berlin: Mouton de Gruyter.
- NORMAN, JERRY. 1988. *Chinese*. Cambridge, UK: Cambridge University Press.
- PAUL, WALTRAUD, and JOHN WHITMAN. 2008. Shi... de focus clefts in Mandarin Chinese. *The Linguistic Review* 25(3-4): 413-451.
- SHYU, SHU-ING. 1995. The syntax of topic and focus in Mandarin Chinese. Los Angeles, California: University of Southern California Dissertation.
- SIMMONS, RICHARD VANNESS. 1999. *Chinese Dialect Classification*. Amsterdam: John Benjamins.
- TANG, SZE-WING. 1998. Parametrization of features in syntax. University of California at Irvine dissertation.

- TANG, SZE-WING. 2005. *Cantonese Grammar*. Hong Kong: The Commercial press.
- TANG, YUMING. 2009. Jindai Hanyu de Panduan Dongci Xi jiqi Liubian [The copula verb xi in Early Modern Chinese and its development]. *Zhongshan Daxue Xuebao* [*Bulletins of Sun Yat-sen University*] 6.3: 55-59.
- TANG, CHAOJU, and VAN HEUVEN, VINCENT. 2007. Predicting mutual intelligibility in Chinese dialects from subjective and objective linguistic similarity. *Interlingüística* 17: 1019–1028.
- TENG, SHOU-HSIN. 1979. Remarks on cleft sentences in Chinese. *Journal of Chinese Linguistics* 7: 101-112.
- WAKEFIELD, JOHN. 2010. The English Equivalents of Cantonese Sentence-final Particles. Hong Kong Polytechnic University Dissertation.
- WANG, LI. 1940. *Hanyushi Lunwenji* [*Papers on the History of the Chinese Language*]. Beijing: Beijing Press.
- XU, LIEJONG, and LIU, DANQING. 2007. *Huati de jiegou yu gongneng* [*The structure and function of topic*]. Shanghai: Shanghai Education Publishing House.
- YUE-HASHIMOTO, ANNE. 1969. The verb ‘to be’ in Modern Chinese. The verb ‘be’ and its synonyms, edited by John Verhaar, 72-111. Dordrecht: Reidel.
- ZHAN, FANGQIONG, and CHAOFEN SUN. 2013. A copular analysis of shi in the Chinese cleft construction. *Language and Linguistics* 14.4: 755-789.
- ZHANG, HEYOU, and TANG, SZE-WING. 2011. On the syntax and semantics of the idiosyncratic copular sentences and empty categories. *Contemporary Linguistics* 12.1: 14-23.
- ZHANG, ZHENXING. 1987. *Language Atlas of China: Volume of Chinese Dialects*. Beijing: The Commercial Press.
- ZHU, YAO. 1996. The focus-marking function of shi in Mandarin Chinese. PhD dissertation, University of Minnesota.