

Towards the copular status of clefts—Evidence from diachrony

Dawei Jin* Jun Chen[†]

Shanghai Jiao Tong University & University of Stuttgart

Abstract This paper draws attention to a robust grammatical change trajectory, here dubbed the copula-to-cleft pathway. We show that the emergence of the copular clause construction tends to lead to the emergence of the cleft construction, and the decline of one tends to be in tandem with the other. We further propose that the above pattern follows from a syntactic treatment that derives the cleft structure from a copular clause. Such copular syntax can be further combined with a semantic analysis where the focus interpretation of clefts is computed based on a copular structure.

Keywords: Cleft sentence, identificational focus, copula, syntactic change

1 Introduction

This paper presents evidence for a biclausal syntax of clefts from a diachronic perspective. We identify a recurring pattern of cleft development in the syntactic history of Chinese: A copula morpheme became employed in a cleft construction following its emergence in a copular clause construction, and subsequently its use in both constructions declined. This process is instantiated in three distinct copula morphemes with chronologically non-overlapping and independent developments, which can be taken to indicate that the copular clauses and the clefts in Chinese share a non-coincidental syntactic and semantic connection. We argue that this connection can be captured by a theory according to which clefts are a type of the copular construction. Both canonical copular clauses and clefts are thus headed by the same copula verb, hence the tendency for both copular structures to emerge and decline in a coordinated way. The assumption of homogeneity, which we will call the copular approach to clefts, contrasts with an alternative approach that

* dawejin@buffalo.edu, jun.chen@ling.uni-stuttgart.de

[†] This paper has a long gestation period. We take pride in ourselves for displaying determination in keeping pursuing this research program in face of adversity. We would like to express our deep gratitude to the three anonymous reviewers of *Acta Linguistica Academica* for their detailed and comprehensive comments. The paper improved significantly as a result. We are also indebted to Ryan Bennett and Jim Wood for their valuable feedback during the preparation of this manuscript. Further thanks go to Lihua Xu, Matt Barros, Matt Tyler, Jason Zentz, Zhenyu Chen, Emanuel ‘Manu’ Quadros, Bob Frank, Yaoying Lai, Maria Piñango and Daniel Hole for discussions. We also wish to thank audiences at the Linguistic Association of Great Britain annual meeting, ESSLLI student session, Penn Linguistics Colloquium and the weekly colloquium at the Institute of Linguistics of University of Stuttgart for their feedback. As usual, all the remaining errors are our own.

treats the copula morpheme in clefts as an operator participating in (exhaustive) focus marking.¹ Unlike the copular approach, we show that the focus-based approach faces difficulties in addressing the above historical pattern of coordinated emergence and decline. The upshot is that diachronic data can be a useful means in evaluating alternative theories of the cleft structure, complementing the rich body of existing syntactic analyses from the synchronic perspective.

The rest of this paper is structured as follows. Section 2 presents three copula-to-cleft grammatical processes. Section 3 shows how an analysis that assumes an underlying uniformity between copular clauses and clefts captures the similarity among the three processes. Section 4 concludes the paper.

2 Data

2.1 Some basics

(1) exemplifies a modern Mandarin Chinese copular clause headed by *shì*. Following Williams (1983), Partee (1987), den Dikken (2006) and Mikkelsen (2005, 2011), we define a copula as a verb that mediates between a (logical) subject and its complement. The copula has no semantic load, but it syntactically functions to establish a predication relation by which the complement is predicated of the subject. We assume in this paper that the copula's function is not order-sensitive: Thus, in predicational copular clauses like (1a), the pre-copula element is the entity-denoting subject, and the post-copula element is the property-denoting complement. In specificational copular clauses like (1b), the predicative complement precedes the copula, and the subject is post-copula.² We will return to this distinction later in Section 3, when the cleft structure is being dealt with in the context of specificational copular clauses.

- (1) a. Zhāngsān shì běijīng rén.
Zhangsan COP Beijing person
'Zhangsan is a Beijinger.'
- b. Dìyī míng shì zhāngsān.
first place COP Zhangsan
'The one that is placed the first (wins the first place) is Zhangsan.'

We define the cleft in terms of a syntactic pattern that introduces a linear separation between the discourse-prominent/focused constituent and the informationally backgrounded constituents. The cleft

¹ In this paper, we take the term 'copula' or 'copula verb' to correspond strictly to a verbal functional element that heads a predication projection in copular clauses (in the sense of den Dikken 2006). The term 'copula morpheme' is used as a theoretically neutral cover term. That is, it refers to both the copula verb proper, and the phonologically identical morpheme appearing in clefts, propositional assertions, etc., regardless of what analysis one adopts and which syntactic category we assign to this morpheme.

² The following abbreviations are used in the glossing: ACC: accusative; C: complementizer; CLF: classifier; COP: copula; DEM: demonstrative; FUT: future tense; NEG: negative, negation; NMLZ: nominalizer; NOM: nominative; PASS: passive; POSS: possessive; PRF: perfect; PRS: present; PRT: particle; PST: past; REL: relativizer; TOP: topic marker.

construction in modern Mandarin Chinese is exemplified in (2). Here the cleft sentence in (2B) provides an exhaustive answer to a prior congruent question in (2A). The linear partition of the cleft is realized in the sense that the focused phrase (henceforth the cleft phrase) is to the immediate right of the copula morpheme and the backgrounded portion (henceforth the cleft clause) follows the cleft phrase (as a tendency but not an absolute constraint). The copula morpheme is optionally preceded by a topic DP in the Chinese cleft.

(2) A: Who will come next week?

B: (Xiàe xīngqī) shì [Zhāngsān]_{focus} [yào lái]_{background}.
 Next week COP Zhangsan will come
 ‘(Speaking of next week), it is Zhangsan who will come.’

In relation to clefts, the same copulative structure can additionally introduce a full CP to the right of the copula as broad focus. We will follow Paul & Whitman (2008) by referring to such sentence focus construction in Chinese as the propositional assertion.³ Despite sharing the same structural correlate, the propositional assertion and the cleft address different questions under discussion. A propositional assertion is situated in a discourse structure that requires propositional answers. Most typically, this involves a prior explanatory/interpretive context that selects for a corresponding reason, cause, consequence or inference, etc. (Heggie 1988; Declerck 1992; Delahunty 2001; von Prince 2012). Thus, the utterance in (3B), differing from the cleft answer in (2B), is construed as an explanation in the immediate context of (3A), while ruling out other alternative explanations (of propositional type) (von Prince 2012).

(3) A: Why are you cleaning up your house?

3 Unlike the sentence-focus construction in the sense of Lambrecht (2000), which functions to introduce new events to the discourse and is compatible with discourse-initial contexts, propositional assertions resist non-restricted contexts. In (i), the propositional assertion cannot felicitously address an *out-of-the-blue* broad focus *wh*-context:

(i) A: What’s happening?

B: #Shì [wǒ zài dǎsǎo jiā]_{focus}.
 COP I PROG clean.up house
 ‘I am cleaning my house.’

Such sensitivity to ongoing/active questions under discussion places propositional assertions among the family of broad focus constructions like the so-called maximal focalization constructions in other languages (Huber 2006). (ii) illustrates a maximal focalizable sentence in French, which employs the same morpho-syntactic strategy used for encoding identificational focus in regular (term) clefts (Clech-Darbon et al. 1999). A similar pattern is observed for Hungarian, Scandinavian languages, Gaelic languages, Japanese, Akan, Swahili and the Gur languages (Kuno 1973; Bearth 1997; Kiss 1998; Huber 2006; Fiedler et al. 2009; von Prince 2012).

(ii) A: You look worried.

B: C’est la petit qui est tombé dans l’escalier.
 it.is the small.MASC who is fallen in the.stairs
 ‘The little one fell down the stairs.’

- B: Shì [Zhāngsān yào lái]_{focus}.
 COP Zhangsan will come
 ‘(It’s that) Zhangsan will come.’

It is safe to say that propositional assertions are not confined to explanatory and related contexts (these contexts are argued to form a delineated natural class characterizing the use of English *it’s that* S-construction, cf. Delahunty 2001). The literature does not yet have a definitive, structured taxonomy of the types of contexts where propositional assertions may felicitously embed. Of particular relevance is the classification by Sheil (2016) based on Scottish Gaelic, which is to date the most detailed taxonomy of the discourse structure for sentence focus constructions. Sheil identifies a number of environments that license the Scottish Gaelic propositional cleft (e.g. contrasting alternatives, exemplification, contrary to expectation, *only*-reading). We find many of them are also suitable contexts for Mandarin propositional assertions. For example, in the naturally occurring utterance below, the propositional answer (4B) addresses a prior *wh*-question that requests picking a valid proposal among contrasting alternatives.

- (4) A: How do we give the cat a shower? Do we use the bathtub? Put her into a shower cabinet?
 Use faucet water?
 B: Shì [xiǎoxīnde yòng shuǐchí de shuǐ yìdiǎndiǎn gěi tā chōng].
 COP carefully use faucet POSS water piece.by.piece to her spill
 ‘It’s that [we slowly turn on the faucet and let the water drip on her].’

Besides context, the meaning difference between the cleft and the propositional assertion can be diagnosed in other environments. For example, the *wh*-cleft structure (i.e. the post-copula cleft phrase is a *wh*-constituent) does not have an alternative broad focus reading. An illustration is given in (5). Here the inherently focal nature of the *wh*-phrase imposes a narrow focus interpretation for the entire sentence, preempting broad focus assignment.

- (5) Shì [shéi]_{focus} yào lái?
 COP who will come
 ‘Who is it that will come?’

Note that propositional assertions *per se* can be questioned. Examples (6) and (7) show that both clefts and propositional assertions can be interpreted within the scope of a polar question. Importantly, their difference in focus assignment leads to distinct inferences in question contexts: A negative answer to a cleft polar question excludes the referent of the cleft phrase from the set of individuals denoted by the cleft clause. The answer in (6B) entails that Zhangsan will not come. However, the entailment does not hold when a broad focus is being questioned. The negative answer in (7B) is compatible with a scenario where Zhangsan will come, although the fact that Zhangsan will come is *not the explanation for* the cleaning up.

- (6) A: Shéi yào lái? Shì [zhāngsān]_{focus} yào lái ma?
 who will come? COP zhangsan will come Q
 ‘Who will come? Is it Zhangsan who will come?’

B: Bù. Shì lǐsì yào lái.
no COP lisi will come
'No. It is Lisi who will come.'

(7) A: Wèishénme nǐ yào dǎsǎo? Shì [zhāngsān yào lái]_{focus} ma?
why you will clean.up? COP zhangsan will come Q
'Why will you clean up? Is it that Zhangsan will come?'

B: Bù. Fángjiān tài zāng le.
no room too dirty PRT
'No. The room is too dirty.'

The remainder of this section chronicles the development of the copular clause construction and the two copulative focus-marking constructions in the history of Chinese. We argue from diachronic evidence that these uses form a structured pattern, representing a recurring trend in the grammatical development of Chinese.

2.2 The copula-to-cleft pathway

The copula morphemes under investigation are *wéi*, *shì* and *xì*. We follow the chronological order in which relevant examples for each morpheme are first attested.⁴ The grammaticalization of each individual morpheme has already been studied previously. Our empirical novelty is to propose that these three morphemes together form a copula-to-cleft pathway: The emergence of the copular clause use invariably results in the employment of the copula morpheme in a propositional assertion strategy as well as a clefting strategy, and the process of decline is witnessed to target all the three uses alike. To our knowledge, this claim has not been made before.

In the following we first exemplify the grammaticalization process of the copula morpheme *wéi*, before proceeding to *shì* and *xì*. The data presented in the following (as well as their translations) are all taken from previous works. Examples (8a-8c) exemplify *wéi*'s functions as attested in the oracle bones inscriptions from the period of early Pre-Classical Chinese (from 14th century BC to 11th century BC).⁵

⁴ The Chinese writing system is logogram-based. Phonological forms are not transparent, so that linguists have relied on reconstruction to postulate the historical phonology of individual morphemes. The reconstructed forms are often not agreed upon, for example, the copula *wéi* has been reconstructed as **rəd* by Li (1971) and **ljuəj* by Schuessler (1987). The copula *shì* has been reconstructed as **djigx* (Li) and **dji* (Schuessler). In this paper, we follow accepted convention and transliterate historical Chinese morphemes using *pinyin*, the official transcription scheme for modern Mandarin.

⁵ Here we use the periodization of Chinese by Tai & Chan (1999), which differs from the mainstream, phonology-based periodization commonly used in the literature. Building upon previous works, Tai & Chan identify the transitional Spring and Autumn period, as well as the period from Late Han through The Six Dynasties (*Wèijìn Nánběicháo*) to Early Tang as pivotal milestones in terms of the syntactic and morphological development of Chinese. Defined by these milestones, we assume the following periods (dates are approximate):

(i) Pre-Classical Chinese: the Chinese language from the oracle bones inscriptions to the Spring and Autumn Period (14th century BC-3rd c. BC);
(ii) Classical Chinese: End of the Spring and Autumn Period to Early Han (3rd c. BC-1st c. AD);

Oracle bones are the undersides of turtle shells used for divination in early China. A fortune-teller carves several typically one-sentence inscriptions on the shells and burns them till the shells crack. The direction that the crack is pointing to is then examined, and the event described by the inscription appearing along the direction is believed to be the true happening. Under this background, the inscription in (8a) is understood as describing a potential future in which the nobin Bin will be minister (people believe this future will happen if the divination's crack pattern points that way). *Wéi* is interpreted as a copula modified by a future tense-marking auxiliary, which links a (proper name) subject argument with a predicative nominal complement (here the analysis is adopted from [Peyraube & Wiebusch 1994](#); [Djamouri 2001](#)). The inscription of (8b) is written to seek divine information about whether the king will take a future hostile action against another tribe. *Wéi* introduces a propositional assertion, with the entire post-*wéi* proposition under negotiation ([Takashima 1990](#); [Yang 2010](#)). (8c) instantiates *wéi*'s cleft use, with the immediately post-copula proper name phrase receiving an exhaustive focus. This is evidenced by the fact that similar inscriptions with a different post-*wéi* proper name are found alongside this inscription, indicating that these inscriptions are written to seek divine information about who placed a curse, given an alternative set of individuals ([Bisang 1998](#); [Liu et al. 2005](#)).

- (8) a. Yà bīn qí **wéi** chén.
 nobel Bin FUT WEI minister
 'The noble Bin will be minister.'
 (*Jiǎgǔwén Héjǐ* 'Collected Oracle Bones Inscriptions', 22301, 14th century BC-11th century BC, translation by [Djamouri 2001](#))
- b. **Wéi** [wáng fá _ fāng].
 WEI king fight missing.tribe.name tribe
 'It is the case that [the king will fight the (illegible tribe name) tribe].'
 (*Jiǎgǔwén Héjǐ*, 614, 14th century BC-11th century BC, translation by [Takashima 1990](#))
- c. **Wéi** [Fù Jiǎ] tā wáng.
 WEI Fu Jia place.a.curse.on king
 'It was [Fu Jia] who placed a curse on the king.'
 (*Jiǎgǔwén Héjǐ*, 1659, 14th century BC-11th century BC, translation by [Bisang 1998](#))

The chronological order of *wéi*'s three uses cannot be determined. Already in the oracle bones records (i.e. the earliest known writing system of Chinese), the three uses are attested with a certain productivity, evidenced by the observation that they occur with a wide range of lexical items and syntactic environments ([Wang 1940](#)). This suggests that all three uses had emerged by the period of pre-oracle bones Chinese, for which there are no written records. No quantitative study has been published in the literature on *wéi*'s corpus frequency, therefore we conducted our own survey. Based on the 67 complete full-sentence *wéi*-inscriptions that [Liu et al. \(2005\)](#) extracted from *Jiǎgǔwén Héjǐ* with translations provided, we found

- (iii) Early Medieval Chinese: Late Han through the Six Dynasties to Early Tang (1st c. AD-10th c. AD);
- (iv) Late Medieval Chinese: Late Tang to the Ming dynasty (11th c. AD -16th c. AD);
- (v) Pre-Modern Chinese (from the Ming dynasty onwards)

that altogether 65 sentences (or 97%) belong to one of the three uses, registering a ratio of 19 (copular clause): 30 (propositional assertion): 16 (cleft).

Unlike the status with *wéi*, the available corpus evidence allows us to look into the grammaticalization process in which *shì* developed into a copula verb. According to a mainstream analysis, prior to its copula use, *shì* starts out as an anaphoric demonstrative pronoun that occurs in a topic-comment structure (Wang 1940; Yue-Hashimoto 1969; Li & Thompson 1977; Feng 1993; Chang 2006). As (9) schematizes, the copula use emerges when the topic-comment structure becomes reanalyzed as a subject-predicate structure. That is, the demonstrative pronoun, which is coreferential with the topic in the comment of a topic-comment construction, is reanalyzed as a copula that intervenes between a subject and a predicate in a subject-predicate construction.

(9) Topic₁ + [_{comment} Pronoun₂ + Predicate₃] → [Subject₁ + Copula₂ + Predicate₃]

(10a) lends itself to a topic-comment interpretation, with the predicate in the comment part connected by the intervening demonstrative pronoun *shì*. In (10b), *shì* is most plausibly analyzed as a copula verb, as it is preceded by a separate demonstrative pronoun *cǐ* ‘this’.

- (10) a. Qióng yǔ jiàn, **shì** rén zhī suǒ wù yě.
Poverty and debasement, SHI people POSS NOM dislike DECL.PRT
‘Poverty and debasement, **that** (is) what people dislike.’
(*Lúnyǔ* ‘The Analects’, vol. 23, 5th century BC, translation by Li & Thompson 1977)
- b. Cǐ **shì** ānshì suì jīn.
this SHI Anshi bit gold
‘These **are** Anshi’s fragments of gold.’
(*Shìshuō Xīnyǔ* ‘A New Account of the Tales of the World’, ch. 4, section 87, 403-444 AD, translation by Li & Thompson 1977)

Shì’s copular clause use is believed to have emerged by the 1st century AD (e.g. Wang 1940; Li & Thompson 1977; Guo 1997), although it is not productively attested in the sources until later (after 6th century AD at the earliest and up to 1000 AD). Attested early uses of the proposition assertion and the cleft function are found between the 5th and the 10th century AD. The two uses are exemplified in (11a-11b), respectively. The proposition introduced in (11a) stands in an inferential relation with the prior context, whereas (11b) has a narrow focus reading (Wang 1940; Shi & Xu 2001; Dong 2004).

- (11) a. Wángníng yì móu, yún **shì** [qīng wèi qí jì].
wangning conspire, say SHI you for him scheme
‘Wang Ning is conspiring for a coup. People said [you are masterminding for him].’
(*Shìshuō Xīnyǔ*, ch. 2, section 100, 403-444 AD, translation by Mather 2002)
- b. Yíqiè zhòngshēng píngděng, jiē yǒu zhēnrúfǎshēn, fēi **shì** [dù] shǐnǎi yǒu yě.
all lives equal, DIST have dharma, not SHI deliverance then have PRT
‘All lives are equal, and all have dharma. It is not [via deliverance] that we have dharma.’
(*Dūnhuáng Biànwén_Wéimójiéjīng Jiǎngjīngwén* ‘Dunhuang Transformation Texts_Lectures on the Vimalakirti Sutra’, vol.2.717, ca. 900 AD, translation by Chen 2019)

Mather (2002) has conducted a quantitative survey of *Shìshuō Xīnyǔ*, a colloquial style document that is among the oldest sources with a significant number of *shì*-copula sentences. Mather finds that 48 (or 73%) out of a total of 66 *shì*-sentences have a structure of [NP/pronoun *shì* NP], thereby potentially instantiating a copular clause use. 9 (or 13%) out of the 66 *shì*-sentences are interpreted as potential propositional assertions that also allows for ambiguous interpretations. No *unambiguous* cleft use is attested by this point. We have conducted another quantitative survey of the 10th century colloquial document *Dūnhuáng Biànwén*, which features some of the earliest attestations of *shì*-clefts. We classify 7 out of 67 (or 10.4%) *shì*-sentences as potentially instantiating a cleft use (a cleft use is identified if a focus-background linear partition is established based on context), in comparison with 11 out of 67 (or 16.4%) potential propositional assertion uses, indicating that the cleft use has reached a level of productivity comparable to the propositional assertion by that time.⁶

The copula morpheme *xì* has generally been agreed to have a lexical origin. It is argued that *xì* started out as a verb meaning ‘to tie/to attach’ (Wang 1940; Gao 1948; Tang 2009), illustrated by (12):

- (12) Jīn zhǐ mǎ zhī bǎi tí ér bù dé mǎ, ér mǎ xì yú qián zhě,
 today point.to horse NOM hundred body.parts yet NEG get horse, yet horse tie to front NOM,
 lì qí bǎi tí, ér wèi zhī mǎ yě.
 stand its hundred body.part then call it horse PRT
 “Now if we point individually to all the body parts of a horse, yet we do not get an idea of what a horse is like. However, if we just need to tie an actual horse to the front and let its body parts stand up, then we can call it a horse.”
 (Commentaries to Zhuangzi, vol.8.322, 52-312 AD)

What happens next is a bleaching process, so that aside from denoting physical attachment, *xì* came to denote a ‘belonging to’ relation, that is, the referent of *xì*’s subject is a part/subcategory of the referent of its second argument (Tang 2009). Note that the part-of relation lends itself to further ambiguity. For instance, in (13a), the mountain in question can be understood as belonging to the territory of the kingdom. An alternative interpretation, available from the same context, treats the construction as a predication copular clause, such that the mountain possesses the property of being the territory of the kingdom. This systematic ambiguity is argued to underlie *xì*’s further development into a copula verb (Tang 2009). (13b) exemplifies an early case of *xì*’s unambiguous copular clause use. Here the referent of the post-*xì*

⁶ Zhan & Traugott (2015:475-476) also investigated the cleft uses in *Dūnhuáng Biànwén*. However, their survey is restricted to the so-called [*shì*...*de*] cleft sentence, with the assumption that a sentence-final particle *de* contributes to the exhaustive focus reading in *shì*-clefts (Paul & Whitman 2008; Hole 2011; Zhan & Sun 2013), so that only tokens where *de* is attested should be analyzed as *bona fide* cases of the cleft construction. This paper does not stand by Zhan & Traugott’s assumption, in view of recent experimental results showing that in both acceptability and processing tasks of Mandarin Chinese, participants draw an exhaustive focus inference in the cleft construction without the *de*-particle (Liu & Yang 2017; Chen 2019; Liu & Xu 2019). These formal results are also compatible with fieldwork findings that in non-Mandarin Sinitic languages such as Xiang, clefts featuring a counterpart of the *de*-particle is dispreferred and rarely used in natural oral discourse. We leave it to future research for understanding the pattern of correlation between the historical development of these two cleft constructions.

constituent does not stand in a part-of relationship with that of the pre-*xì* constituent, ruling out a ‘belong to’ interpretation (cf. Sun 2018).

- (13) a. Hónghé’ěr dàshān zìlái xī běi cháo dìtǔ.
 Honghe’er mountain since.early XI north dynasty territory
 ‘The Honghe’er Mountains have long since belonged to the territory of the Northern Kingdom.’
 (*Yǐmǎo Rùguó Zòuqǐng*, ca. 1031-1095 AD, translation from Chinese based on Sun (2018))
- b. Zǐ páo rén xī yān rén wúxiàomín. Bái páo rén xī jīn rén.
 purple robe person XI Yan person wuxiaomin. White robe person XI Jin person
 ‘The man in purple robe is Wu Xiaomin, a native of Yan. The man in white robe is a native of Jin.’
 (*Sāncháo Běiméng Huìbiān* ‘Collection of documents concerning alliances with the north during the three reigns’, vol.28, ca.1137 AD, translation from Chinese based on Fan 2000)

(14a) presents one of the earliest attested uses of *xì* as a copular clause, around the turn of the 10th and the 11th century AD (Wang 1940; Tang 2009; Jin 2017; Sun 2018). *Xì*’s propositional assertion use and cleft use emerged in quick succession, exemplified in (14b) and (14c), respectively.

- (14) a. Yǐshàng bā rén, yì yuán xī běn zhōu wǔwèi dìjiǔ zhǐhuì, jiàngyúhóu,
 above 8 people, also originally XI local province garrison 9th zhihui, jiangyuhou,
 chéngjú.
 chengju.
 ‘The above mentioned eight people were previously also the ninth *zhihui*, the *jiangyuhou* and the *chengju* of the local garrison, respectively (*zhihui*, *jiangyuhou* and *chengju* are military titles within the military rank of the Song Dynasty-era China).’
 (*Ōuyáng Wénzhōnggōng Jí* ‘Collections of Ouyang Xiu’, vol.118.2.1.15, ca.1072 AD, translation by Sun 2018)
- b. Zìlái guóshū zhǐ shì sī fēn rén xiūxiě, jūyú
 since.long diplomatic.letters only corresponding ministry assign person transcribe, follow
 tǐlì, zì wú rěbǐ. Jīn xī [zhǔshàng qīn yù hàn mò].
 protocols, naturally no illegible.writings this.time XI [His.Majesty self take.hold pen ink]
 ‘(The Commander Zhan Han scolded the diplomat that the diplomatic letter presented to him had illegible writings, showing that the appointment of people in charge of writing was not careful enough. The diplomat answered:)
 It’s long been the case that diplomatic letters are in the sole charge of corresponding ministries, who assign scribes and follow all the protocols, so that naturally there are no illegible writings. This time, it is the case that [His Majesty took up his pen and ink Himself to write the letter] (that’s why there are illegible writings).’
 (*Sāncháo Běiméng Huìbiān* ‘Collection of documents concerning alliances with the north during the three reigns’, vol.15.13, ca.1137 AD, translation from Chinese based on Fan 2000)

- c. Zì cǐ měi zhǎn, bìng xī hàn'ér zǎixiàng jí zuǒyòu qīnjìn lángjūn
 since this each CLF, all XI Han.ethnic prime.minister and surrounding close entourage
 guì jìn.
 kneel.down offer.wine
 ‘(A man was standing with a purple coat and wearing a silk belt, the man was recognized as the Prime Minister of the Han nation, Zuo. He knelt down in front of the emperor and offered wine to pay homage....The emperor finished the cup of wine, and ordered all those who were seated to kneel down and offer wine again...)
 Since that round, in each round, it was always [the Han prime minister and his entourage] who offered wine while kneeling down.’
 (*Sāncháo Běiméng Huìbiān* ‘Collection of documents concerning alliances with the north during the three reigns’, 516716.15.5.9, ca.1137 AD, our own translation)

In a study of transcriptions of oral conversations and correspondences in the early 16th century source *Míngshìzōng Shílù*, Jin (2017) reports a ratio of propositional assertions versus clefts at 17:11, indicating that both uses have reached a comparable level of productivity by that time.

All the three copula morphemes underwent independent decline processes. It has been claimed that *wéi* had disappeared from the three uses we have shown in (8), and had changed into a distinct exclusive adverb by the 6-4th century BC (Wang 1940; Pulleyblank 1995; Meisterernst 2010; Yang 2010). As an anonymous reviewer points out, however, the timeline of this path of change could last significantly longer, complicated by periods of transition. Suffice it to say that, from available evidence, by the 4th century BC, new structures had emerged on the basis of Pre-Classical *wéi*-clefts. A potential environment of structural change involves object fronting, illustrated in (15). As Meisterernst (2010) discusses in detail, here the preposed object focus is followed by the pronoun *shì* (e.g. 15a) or the particle *zhī* (e.g. 15b).⁷

- (15) a. Wéi lì shì shì.
 WEI strength SHI look.at
 ‘It was strength (success) that I had in mind.’
 (*Zuǒ, Xī*, 24, 375-360 BC, translation by Meisterernst 2010)
- b. Wú wéi zǐ zhī yuàn.
 I WEI you GEN resent
 ‘I will resent only you.’
 (*Zuǒ, Wén*, 7, 375-360 BC, translation by Aldridge 2019)

The presence of both *wéi* and *shì/zhī* reflects a mixture of old and new structures that indicates a period of transition (Wang 1940). *Wéi* can still be analyzed as retaining its previous function (see Aldridge (2019) for a copular analysis of the structures instantiated by (15a-15b)). Interestingly, though, the insertion of new elements triggered further structural changes. Meisterernst observes that (unlike in Pre-Classical

⁷ I assume with Aldridge (2019) and previous works that *shì* here functions as an anaphoric pronoun that instantiates a doubling strategy. The particle *zhī* is analyzed as projecting a genitive phrase that selects for the VP. We appreciate an anonymous reviewer for drawing my attention to Aldridge (2019)’s work.

corpora) by Classical texts, there are very few attested cases where preposed objects are solely introduced by *wéi* without a following *shì*.⁸ The attested cases either involve the co-occurrence of both morphemes as in (15), or alternatively *wéi* is dropped as in (16).

- (16) Jiāng guó **shì** miè, hé ài yú Yú?
 FUT Guo SHI destroy, what feel at Yu
 ‘And if it really destroys Guo, what kind of feelings will it have about Yu?’
 (Zuǒ, Xǐ, 5, 375-360 BC, translation by Meisterernst 2010)

The emergence of the latter construction indicates a tendency for the gradual loss of the cleft interpretation, and for the gradual emergence of a new focalizing strategy in which the copula role of (the now optional) *wéi* is open to reanalysis. It is plausible to hypothesize that over time, the reanalysis of the overall construction also led to a change in the meaning assigned to *wéi* (Meisterernst 2010:89): As the new focalizing structure [(*wéi*) NP_{obj} *shì* V] still encodes exhaustive focus, *wéi* might be frequently interpreted as ‘only’, leading to its reanalysis as an exclusive adverb.

The full-fledged development of *wéi* into an exclusive adverb is reflected in further changes in its distribution. For instance, with *shì* later developed into a copula, *wéi* is witnessed to modify the copula *shì* (see 17a). Second, when *wéi* is in the scope of negation, the corresponding *wéi*-focus construction may be followed up by an additive continuation that introduces other alternatives (see 17b). Note that a negated cleft sentence is not compatible with an additive continuation.⁹

⁸ The transitional nature of the mixed structure within this period can be further seen from an asymmetry between object and subject focus. In the following example (i), which involves a subject focus, the focused constituent is solely introduced by *wéi*, whereas *shì/zhī* does not appear. In such case, the construction is indistinguishable from the earlier, Pre-Classical cleft construction. The evidence thus suggests that object focus (fronting) most likely serves as the catalyst for *wéi*’s change, and it is very possible that its copular use has yet to vanish by that period. We are indebted to an anonymous reviewer for drawing my attention to such subject-object asymmetry.

- (i) Zhèng rén jiē xǐ, **wéi** Zíchǎn bù shùn.
 Zheng person all glad, WEI Zichan not agree
 ‘The people of Zheng were all glad. [Zichan]_{focus} disagreed.’
 (Zuǒ, Xiāng, 8, Aldridge (2019:20))

⁹ A standard way to diagnose the difference between the cleft construction and the exclusive focus construction involves their different behaviors when the exhaustive component is under negation, an observation that goes back to at least Horn (1981). The distinction is also seen in modern Mandarin Chinese as follows.

- (i) a. *Bù **shì** zhāngsān yào lái. Lǐsì yě yào lái.
 not COP Zhangsan will come. Lisi also will come
 ‘It is not Zhangsan that will come. Lisi also will come.’
 b. Bù **zhǐ** zhāngsān yào lái. Lǐsì yě yào lái.
 not only Zhangsan will come. Lisi also will come
 ‘Not only Zhangsan will come. Lisi also will come.’

The contrast in (ia) and (ib) would be accounted for, if we assume that the exhaustive reading *shì*-clefts generate is not part of what is asserted (i.e. the at-issue dimension of meaning, in the sense of Potts 2004). In other words, cleft exhaustivity is derived from the level of presupposition, or the level pertaining to conventional implicature or conversational implicature

- (17) a. Yóu **wéi** shì gū chén, jiāng hé bào guó?
 You(First Name) **only** COP unattached official, will how repay country
 “You is only a government official unattached to any factions, how will You be able to repay the country?”
 (*Fànzhōngxuān Jí* ‘Collections of Fan Zhongxuan’, vol.7, 1351 AD, translation by [Pulleyblank 1995](#))
- b. Gǔ rén jìsì wú bù yòng shī. Fēi **wéi** jìsì jiā
 ancient person offer.sacrifice never not use surrogate. not.be WEI offer.sacrifice family
 xiān yòng shī. Jì wài shén yì yòng shī.
 ancestor use surrogate. offer.sacrifice external gods also use surrogate
 ‘The ancient people never failed to use surrogates (figurines/statuettes symbolizing the deceased ancestors) in sacrifice ceremonies to honor the dead. It is not only in honoring the family ancestors that (they) used surrogates. When honoring outer gods, they also used surrogates.’
 (*Zhūzǐ Yǔlèi* ‘Dialogues of Zhu Xi’, vol.90.2.65, ca. 1263 AD, my own translation)

The other two copula morphemes *shì* and *xì* have also undergone decline processes. It has long been observed that Sinitic languages (descendants of historical Chinese) are divided into the *shì*-type and the *xì*-type: As summarized in Table 1, *shì* is in productive use today in modern Mandarin Chinese and a subset of other Sinitic languages, yet has ceased to be used in other Sinitic languages such as Cantonese and Hakka. In those languages where *shì* has disappeared, *xì* is found to be in productive use ([Wang 1940](#); [Gao 1948](#); [Yue-Hashimoto 1969](#); [Hashimoto 1973](#); [Tang 2009](#); [Jin 2017, 2020](#)). This tendency for only one copula to be productively used within an individual language thus points to a competition effect, in which the redundancy of copulas with identical functions triggered the decline of one copula and the retaining of the other.

Type of copula	Sinitic languages
<i>shì</i> -type	Gàn, Mandarin, Mǐn, Wú, Xiāng
<i>xì</i> -type	Cantonese, Hakka, Huī, Píng

Table 1: Grouping of Sinitic languages by copula type

([Hartmann & Veenstra 2013](#)). Following this idea, in (ia), the proposition within the scope of negation from the first sentence asserts that Zhangsan is not a member of the contextually restricted set of individuals denoted by the sentence predicate. The negated proposition is incompatible with the follow-up additive continuation, assuming that the additive operator presupposes the existence of a member in the predicate’s alternative set. Conversely, in (ib) the exhaustive meaning is asserted in the *only*-sentence, while the truth of *only*’s prejacent is presupposed/implicated. This way, (ib) creates no contradiction: The first sentence asserts that Zhangsan is not the *sole* member of the predicate’s alternatives, compatible with the follow-up additive statement. Assigning cleft exhaustivity to the non-at-issue dimension also explains the widespread feeling that the cleft has a weaker exhaustive inference compared against *only*. This is expected, under the assumption that a non-asserted exhaustive reading is more violable than an asserted one.

Crucially, the mutually complementary pattern of *shì* vis-à-vis *xì* is observed to apply to all the three environments of copular clauses, propositional assertions and clefts: As previous surveys and our own fieldwork have confirmed, the generalization is that for a given Sinitic variety, if a reflex of the historical Chinese *shì*-copula is used in one of these environments, then it is also used for the other two environments. The same applies to reflexes of the *xì*-copula. We do not find cases where (for instance) a Sinitic language employs a *shì*-copula for copular clauses but employs a *xì*-copula for marking clefts.

copula	Mandarin corpus (CallFriend 1996-1997)	Cantonese corpus (HKCanCor 1997.3-1998.8)
<i>shì</i> -type	3513 (copula: prop-assertion: cleft = (approx.) 58:23:17)	2
<i>xì</i> -type	24 (copula: prop-assertion: cleft = 10:7:7)	5387 (copula: prop-assertion: cleft = (approx.) 3:1:1)

Table 2: Tokens of copula morphemes in modern Mandarin and Cantonese

The pattern of decline can be quantitatively corroborated in modern Mandarin and Cantonese, the two Sinitic languages where oral corpus data are available. We have calculated the token frequency of *xì* against *shì* in two oral corpora.¹⁰ Our search results in Table 2 show that Mandarin predominantly uses the *shì*-copula, whereas the *xì*-copula has all but disappeared from colloquial discourse. Meanwhile, colloquial Cantonese almost exclusively uses the *xì*-type copula.

To summarize, we have brought attention to three copula morphemes that emerged in distinct periods of historical Chinese. We unveil a common grammaticalization pattern, namely, when a copula morpheme began to be used productively in copular clauses, it also expanded to the focus-marking constructions (propositional assertions and clefts). Moreover, when the focus-marking copula morpheme underwent change/decline, it also declined from the copular clause construction. Figure 1 summarizes the appearance and the decline of the individual uses.

¹⁰ The spoken Mandarin [CallFriend corpus](#) (Canavan et al. 2018) consists of annotated transcriptions of 24 hours of unscripted Mandarin telephone conversations between family members and friends (with part-of-speech tagging) in the year 1996-1997. The spoken Cantonese [HKCanCor corpus](#) (Luke & Wong 2015) collects 93 radio talk shows and telephone transcripts (230,000 words, also annotated with part-of-speech tagging) by Hong Kong Cantonese speakers from March 1997 to August 1998.

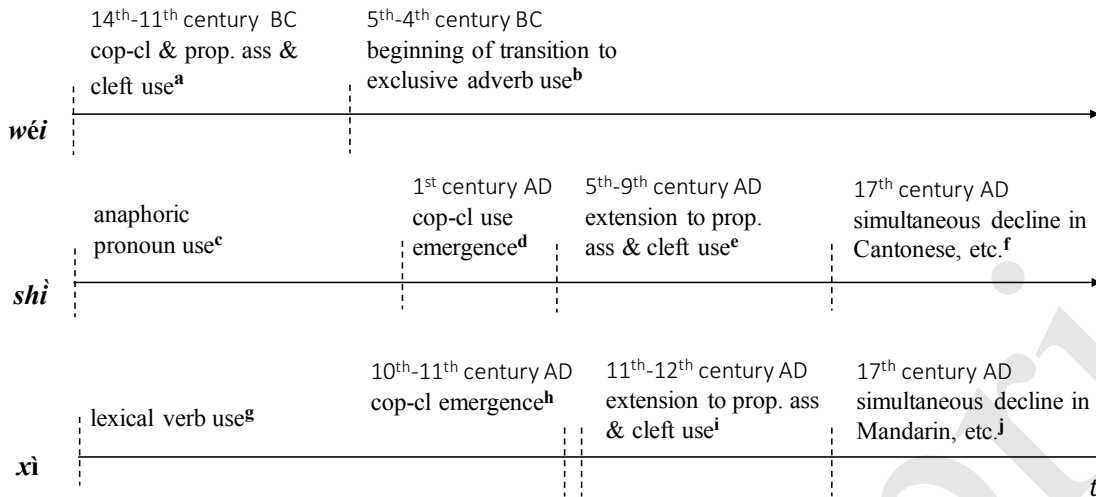


Figure 1: A timeline of copula development. Source texts in superscript (see Appendix, Table 4)

- a. *Jiǎgǔwén Héjǐ Shìwén* b. *Zuǒ* c. *Lúnyǔ* d. *Shìshuō Xīnyǔ*
e. *Dūnhuáng Biànwén* f. *HKCanCor* g. *Yǐmǎo Rùguó Zòuqǐng* h. *Ōuyáng Wénzhōnggōng Jí*
i. *Sāncháo Běiméng Huìbiān* j. *CallFriend*

3 In favor of a copular approach to cleft constructions

In this section, we show that the copula-to-cleft pathway supports treating Chinese clefts as underlyingly a copular clause. We start in section 3.1 by showing how the cleft structure is derived from a copular structure, and then in section 3.2 we show the alternative characterization of the cleft as a clause with a focus-sensitive modifier. The syntactic analyses we present below are not our own. We draw upon well-established previous proposals to make possible a precise illustration of how the cleft sentence can be analyzed under each of the two competing approaches. We then present in section 3.3 our argument that the copula-to-cleft pathway discussed in Section 2 is compatible with the copular approach, while presenting challenges for the focus-based approach.

3.1 Clefts as a copular sentence

According to an idea dating back to Jespersen (1928), the copula morpheme in the cleft is a copula verb, which projects a special kind of copular clause. Some earlier accounts of Chinese clefts within the copular approach have specifically subscribed to the view that clefts are derived from pseudoclefts (Chao 1968; Yue-Hashimoto 1969; Tang 1983; Li & Thompson 1989; Shi 1994; Zhan & Sun 2013). The standard pseudocleft analysis is based on Percus (1997)'s account of the English *it*-cleft (see also Akmajian 1970; Bolinger 1972; Hedberg 2000; Han & Hedberg 2008). The account treats the *it*-cleft structure as derived from a specificational pseudocleft. This can be seen clearly by comparing (18a) with the specificational pseudocleft in (18b).

- (18) a. [It]_{complement} is [Zhangsan]_{specificational value} [who will come]_{headless relative}.
 b. [(The one) who will come]_{complement} is [Zhangsan]_{specificational value}.

In (18a), the copula heads a clause with a logical subject and a pronominal predicative complement that co-refers with a right-dislocated headless relative (treated as adjunction to a right-peripheral clause position via A'-movement, cf. Buring & Hartmann 1997).¹¹ Similar to (18b), the complement provides a variable, and the subject specifies the value of that variable.¹²

While there is an established tradition analyzing the cleft structure as pseudoclefts, such analysis has few, if any, *Chinese-internal* motivations. As one anonymous reviewer points out, the core claim that the cleft structure is derivable from the pseudocleft is challenged by their non-identical distribution, e.g. not all types of focus constituents in a Mandarin specificational pseudocleft are focussable in a corresponding cleft (see Cheng (2008) for more counterarguments). There is an additional lack of support for the claim that the cleft clause in Chinese undergoes CP extraposition. By postulating extraposition, the pseudocleft analysis assumes that the cleft structure involves one additional step of movement compared to the propositional assertion. Such movement has yet to be diagnosed with syntactic or prosodic evidence.¹³

The body of evidence thus undermines characterizing the Chinese cleft with a Percus syntax. In what follows we adopt a different copular-based analysis by Cheng (2008) (see also Cheng & Vicente 2013). Cheng (2008) treats clefts in Mandarin as distinct from (and not derived from) pseudoclefts. The account has the additional merit of having fully fleshed out the role of the copula (underspecified in most pseudocleft-based analyses). Based on the framework developed in Stowell (1981), Cheng (2008) proposes that the copula projects a small clause establishing a predication relation between the (logical) subject and the predicate. To show how the derivation works, we present a syntactic analysis based on the *shì*-cleft in (19).¹⁴

- (19) a. Shì Zhāngsān yào lái.
 SHI Zhangsan will come
 'It is Zhangsan that will come.'

11 We have adopted the standard assumption that in a specificational copular clause, the pre-copula element is a predicative complement, and the post-copula element is an entity-denoting subject. We refer the reader to the discussions above example (1).

12 In a number of languages, the agreement morphology on the copula is subject to the person features of the nominal arguments on par with pseudoclefts, providing explicit evidence that the cleft construction is derived from the latter (Percus 1997; Gribanova 2013; Hartmann & Veenstra 2013).

13 Support for the cleft clause occupying an extraposed position typically comes from OV languages such as Dutch, where the relative clause must appear radically final rather than appear pre-verbally (den Dikken 2013). Phonology-based evidence has also been resorted to in the literature. For instance, the intonational phrase pattern in Zulu shows that the cleft pivot resides at the CP edge, presenting evidence that the cleft clause is extraposed to a CP-external position (Cheng & Downing 2013).

14 Here we focus on the so-called bare *shì*-cleft, which is the object of investigation in our historical survey. According to Cheng (2008), bare *shì* and *shì...de* clefts share the copular-based structure, with the copula projecting a small clause in both constructions. Cheng does not treat the two structures as homogeneous: Narrow focus constituents in bare *shì* clefts are proposed to move to [Spec, CP], whereas such movement is constrained in *shì...de* clefts, probably due to the presence of the *de*-particle.

- b. *shi* [SC [subject [CP *Zhangsan yaolai*]] [predicate *pro*]]
 c. *pro*_i *shi* [SC [subject [CP *Zhangsan yaolai*]] [predicate *t*_i]]

As (19b) illustrates, *shi* selects for a small clause, with a CP residing in the small clause subject position, and a *pro* element residing in the predicate position. The *pro* predicate raises further to the left of the copula, creating an inverse predication structure as in (19c), similar to what has been proposed for the garden-variety specificational copular clause. The assumption of a non-pronounced *pro* predicate associated with the pre-copula slot is compatible with a plethora of agreement and anaphor-based evidence coming from typologically distinct languages (Moro 1997; Adger & Ramchand 2003; den Dikken 2006; Cheng & Downing 2013; Sheil 2016).¹⁵

In addition, the pre-copula slot may be occupied by (frame-setting) topical constituents,¹⁶ derived via either movement or base generation depending on framework (Xu & Langendoen 1985; Xu 1986, 2006;

¹⁵ To our knowledge, there is no explicit evidence for the inversion of *pro* in Mandarin. However, we would like to note that the Mandarin behaviors are in conformity with a standard test of inverse predication in specification (Moro 1997; den Dikken 2006, 2013): In (ia), the specificational copular clause embedded under the propositional attitude verb *consider* must contain a copula (a predication clause thus embedded freely allows for copula dropping). (ib) shows that an underlying specificational structure with a silent *pro*-predicate parallels (ia) in terms of the obligatoriness of the copula, compatible with a raising analysis of the *pro*-predicate.

- (i) a. Nü zhujue shi xiaohua. Zhen dang nü zhujue *(shi) xiaolian la?
 female protagonist COP xiaohua really consider female protagonist (COP) xiaolian PRT
 ‘The heroine is Xiao Hua. Did (you guys) really consider the heroine to be Xiao Lian?’
 b. Nü zhujue shi xiaohua. Zhen dang *(shi) xiaolian la?
 female protagonist COP xiaohua really consider (COP) xiaolian PRT
 ‘The heroine is Xiao Hua. Did (you guys) really consider (it) to be Xiao Lian?’

In den Dikken (2006)’s analysis, the copula initially heads the small clause, before subsequently undergoing movement to a higher functional head, motivated by its role as a linker that circumvents the locality constraint otherwise incurred during the long-distance raising of the *pro*-predicate. In this sense, den Dikken shows that the obligatoriness of the copula testifies to the inversion of the predicate. We will leave more evidence of inverse predication to future work.

¹⁶ It has been widely observed that topicalization in Chinese is not sensitive to the subject island constraint (Xu & Langendoen 1985; Xu 1986; Li 1990; Huang 1992; Lin 2005; Xu 2006). Examples like (i) sound very natural to most speakers.

- (i) Zhei-ge ren de mingzi, [ni mei ting-guo] hen zhengchang.
 DEM-CLF person POSS name you have.not hear-EXP MOD normal
 ‘That you haven’t heard of this guy’s name is quite normal.’

Similar claims about the absence of subject islands have been made in other East Asian languages like Japanese (Ross 1967; Kuno 1973; Takahashi 1994). In general, there appears to be wider cross-linguistic variation in terms of subject island effects compared with other strong island constraints (See Stepanov 2007 and Chaves & Dery 2019 for a crosslinguistic review, and see Sprouse et al. 2016 for an experimental study showing Italian lacks subject island effects). The above empirical evidence thus offers us a context as to why many works of Chinese clefts consider it tenable to relate the sentence-initial topic to a position inside the subject of the small clause.

Li 1990; Takahashi 1994).¹⁷ The topic is assumed to be sentence-initial, as *pro* resides in the specifier of the phrase which the copula heads.

- (20) a. Zhāngsān shì míngtiān yào lái.
 Zhangsan COP tomorrow will come
 'As for Zhangsan, it is tomorrow that (he) will come.'
 b. *Zhangsan*_k *pro*_i *shi* [_{SC} [_{CP} *t*_k *mingtian yaolai*] [*pro*_i]]

The propositional assertion receives the same underlying copular structure as the cleft under Cheng (2008). The focus-marking behavior of the two constructions follows from the assumption that specificational copular sentences have a fixed, post-copula focus position (Higgins 1979; Mikkelsen 2005). This is evidenced by the contrast between (21) and (22), indicating that focus in a specificational copular clause has to be associated with the post-copula position within the small clause, rather than the inverse predicate position.¹⁸

- (21) Post-copula focus is fine.
 Q: Who is the mayor?
 A: The mayor is [John]_{focus}.
 (22) Pre-copula focus is infelicitous.
 Q: Who/What is John?
 A: #[The mayor]_{focus} is John.

That focus is associated with the post-copula position in specificational structures hence straightforwardly allows for focus assignment to the CP subject in the small clause of *shì*, capturing the focal behavior of the propositional assertion. Cheng further proposes that in the case of the cleft, narrow focus assignment is derived via movement of the focused constituent to [Spec, CP] (either through prosody-conditioned covert movement or overt movement).¹⁹

¹⁷ See Li (1990), Takahashi (1994), Stepanov (2007) and Jin (2016) for analyses that movement from subjects follow a different mechanism from those in other strong island contexts. See Xu & Langendoen (1985), Xu (1986), Huang (1992), Cheng (1973), *inter alia*, for proposals that topicalization across subjects involves base generation, with both the topic and an empty pronoun generated in their surface positions and linked via a semantic binding mechanism that is insensitive to locality constraints.

¹⁸ Predicational copular clauses are different and have a flexible focus structure. The following observations indicate the flexible focus placement in the predicational structure:

- (i) Focus on complement or subject
 Q: Is Sam the mayor?
 A1: No, Sam is the [fire chief]_{focus}.
 A2: No, [John]_{focus} is the mayor.

¹⁹ Compatible with Cheng (2008)'s view, Sheil (2016) presents a more detailed account of agreement-triggered movement to derive the narrow focus reading. She assumes that the C head contains an uninterpretable focus feature, attracting a constituent

Semantically, the pre-copula *pro* in specificational copular structures can be construed as anaphoric to a prior discourse object (Mikkelsen 2005, 2011). Specifically, we follow Declerck (1992) and von Prince (2012) in assuming that *pro* in the cleft evokes the immediate Question under Discussion (iQuD), which is addressed by the post-copula (logical) subject through specifying a value. The process of specification is constrained by the question-answer congruence (Roberts 1996), requiring that the alternatives triggered by the focus within the answer are those alternatives that the iQuD ranges over (see also Büring 2003; Constant 2014). This means that in a felicitous cleft answer, the iQuD takes the form congruent to a bipartite structure, partitioned into a *wh*-restriction corresponding to the cleft focus and a nuclear scope corresponding to the backgrounded property (Barros 2014).

We follow Sheil (2016)'s proposal that *pro* in the propositional assertion is construed in a similar way. *Pro* also evokes iQuD, which is constrained by the question-answer congruence and corresponds to a bipartite structure. Here Sheil assumes that the bipartite involves an elided backgrounded part (what she terms as the covert Context Set partition). Consider a typical explanatory context, the elided part is pragmatically recovered as a presupposed consequent *q*. Given an iQuD of the form *why q?*, a congruent bipartite structure assumes the form of *pro shi* [*p*]_{focus} [(*q*)]_{background}, i.e. partitioned into the broad focus *p* and the elided consequent *q* with *p* specifying the explanation of *q*.²⁰

Finally, Cheng (2008) assumes that the focus interpretation in clefts and propositional assertions can be derived based on the mechanism of 'exclusion by identification' (Kenesei 2006), which applies to identificational structures in general. The gist is that by selecting one element of the set of individuals within the universe of discourse, all the other members are simultaneously excluded. Note that the exhaustive reading associated with identificational focus can be computed in this way, according to proposals dating back to at least Szabolcsi (1981). Kenesei builds upon Szabolcsi (1981) and proposes to capture the identification process by resorting to the maximality (iota) operator. Thus, given a specificational structure with a property *K* denoted by the predicate and a specificational value *a* introduced by the logical subject, Kenesei argues for the following paraphrase in (23).

- (23) a. $\iota x (K(x)) = a, x \in R$ (where *R* is the relevant set in the universe of discourse)
 b. the *x*, such that $x \in R$, for which *K*(*x*) is the case, is identical with *a*

In the following we briefly illustrate how the exhaustive focus mechanism works with an example. Here we opt to use a modified mechanism by Büring & Križ (2013), rather than the more established Szabolcsi/Kenesei proposal making use of a maximality operator (see also Heim 1999, 2015), for the

to its specifier. On the other hand, she allows that the broad focus alternatively stays *in situ* (within IP), in which case she assumes a sufficiently local relation with the C head (necessary for agreement) is still obtained, drawing crucially on the mechanism of Anti-Locality in Abels (2003). We refer the reader to Sheil (2016:132-133) for details.

²⁰ See Delahunty (2001:520) for corpus examples of the English sentence focus construction, where a backgrounded cleft clause can be overtly realized.

- (i) I wonder if it was that they hadn't room enough for them up in the house [that they put them out here in the woods].

reason that maximality does not suffice to derive the exhaustivity of identificational focus in negative contexts. We refer to Buring & Križ (2013:6) for detailed discussions.²¹

The crucial insight from Buring & Križ (2013) is that the exhaustive reading in identificational focus arises from a conditional presupposition.²² The presupposition is such that the individuals denoted by the logical subject are either not included within the predicate, or when they are, they exhaust the individuals denoted by the predicate. Buring & Križ assume that the presupposition is encoded in a definiteness operator, represented as (24). Note that the types in the formula can be left underspecified, e.g. P may as well range over sets of propositions (aside from ranging over sets of individuals), and thus be made compatible with the exhaustive reading in propositional assertions.

$$(24) \quad \llbracket \text{DEF} \rrbracket = \lambda z. \lambda P : \forall x \in \max(P) [z \not\sqsubset x]. P(z)$$

Here DEF is treated as a generalized quantifier determiner that takes in a predicate and an entity as argument. It asserts nothing more than that the predicate P is predicated of the entity z . Meanwhile, the formula between the colon and the period defines the domain of the partial function, which can be thought of as a presupposition (see e.g. Heim & Kratzer 1998). Hence, the definiteness operator presupposes that its entity argument is *not* a proper part of (i.e. $\not\sqsubset$) any maximal element of the predicate's extension, where the part-of relation \sqsubset is mereologically defined. We now return to the structure in (19a), while assuming that the copula is semantically vacuous. Based on the denotation of DEF, (19a), under a cleft reading, is mapped to the following meaning contribution:

(25) (19a) asserts: $\llbracket \text{will come} \rrbracket (\llbracket \text{Zhangsan} \rrbracket)$

(19a) presupposes: $\llbracket \text{Zhangsan} \rrbracket$ is not a proper part of the maximal element(s) in $\llbracket \text{will come} \rrbracket$.

Let's use the notation K to refer to the property associated with the individuals who will come, and use a to refer to the individual Zhangsan. Given a situation where the extension of K contains a and only a , the presupposition in (19a) is satisfied (the maximal element within K 's extension is a , and a is not a proper

²¹ To see this informally, the negated Mandarin *shì*-cleft in (i) is intuitively not true in a situation where Zhangsan and Lisi will come. Nevertheless, this intuition fails to be captured under a maximality presupposition account: If the iota operator applying to the predicate *yaolai* extracts the maximal element of the individuals that will come, it then follows that the individual Zhangsan does not possess the property of being this maximal individual, hence the sentence is expected to be true under such situation, contrary to our intuition. Note that the interpretation problem arises for identificational structures in general, e.g. *The winners of the game are not Zhangsan and Lisi*.

(i) Bú shì zhāngsān yào lái.
not COP Zhangsan will come
'It's not Zhangsan that will come.'

By contrast, Buring & Križ (2013)'s parthood-based presupposition account gets us the correct prediction. Under this account, (i) presupposes that Zhangsan is not a part of the individual that will come. In the situation given here, however, the individual that will come is the maximal individual Zhangsan+Lisi, of which Zhangsan is a part. The presupposition is therefore violated, hence the sentence is correctly predicted to be anomalous due to presuppositional infelicity.

²² The treatment here is compatible with the view that the exhaustive reading *shì*-clefts generate pertains to the non-at-issue dimension of meaning. See footnote 9 for a detailed discussion.

part of itself). Simultaneously, the asserted content in (19a) is satisfied, hence Buring & Križ's framework predicts that (19a) is true under this situation. Under another situation where K 's extension contains not only a but at least another atomic individual e.g. b , then the maximal element within K 's extension is $a+b$. Since a is a proper part of $a+b$, the presupposition in (19a) cannot be satisfied. Therefore, the utterance of (19a) in this situation will be infelicitous. Finally, under a situation where the extension of K does not contain a , then the presupposition in (19a) is satisfied (a is not identical to any of the atomic individuals within K 's extension, hence a is not a proper part of K 's maximal element). However, this situation would cause the asserted content to be false, therefore the utterance of (19a) is again infelicitous. In other words, the only context where the statement is true is when Zhangsan will come but no other individuals will come. Consequently, based on Buring & Križ's parthood-based semantics, we correctly derive the exhaustive inference of specificational sentences in both positive and negative contexts.

3.2 A focus marker analysis of clefts

In the following we turn to the alternative approach dating back to [Jespersen \(1937\)](#), according to which clefts are not a type of the copular clause, but rather a single CP structure, with the copula morpheme functioning as a focus-sensitive modifier directly encoding the exhaustive semantics associated with clefts, in which cases the cleft morpheme is homophonous with the copula verb. Similar to the copular-based approach, the focus-based approach has also been taken up by a plethora of accounts of Chinese clefts (e.g. [Teng 1979](#); [Huang 1982](#); [Zhu 1996](#); [Shi & Xu 2001](#); [Dong 2004](#); [Yang 2011](#); [Erlewine 2020](#)). In many languages, e.g. Japanese, Uzbek and Wolof, it has been argued that clefts are distinct from copular clauses and instantiate a monoclausal structure (cf. [Hiraiwa & Ishihara 2002](#); [Gribanova 2013](#); [Klecha & Martinović 2015](#)). In Japanese, for instance, it is argued that the cleft phrase bears the same case as that of the focused constituent in a corresponding *in situ* focus construction. Accordingly, [Hiraiwa & Ishihara \(2002, 2012\)](#) argue that the cleft phrase is derived from the *in situ* focus phrase (where it is licensed for case assignment), suggesting an analysis of Japanese clefts as involving a focus marker.

We will present an idea of how this approach yields the cleft structure and captures its exhaustive interpretation. Specifically, there are two distinct proposals to implement the focus-based approach. According to one analysis, *shì* is a focus-sensitive sentential modifier adjoined at the CP level that must c-command its associate (focus) ([Rooth 1985](#)), and is additionally subject to a low attachment constraint such that while c-commanding its focus associate, it occurs adjacent to the focus whenever possible ([Yang 2011](#); [Erlewine 2020](#)). The preference for low attachment has been independently proposed for German and Dutch exclusive focus operators ([Jacobs 1983](#); [Buring & Hartmann 2001](#)).

Based on this analysis, when *shì* is in a pre-subject position, it either associates with the subject or with the entire clausal prejacent due to low attachment, enabling a unified underlying structure for clefts and propositional assertions, as in (26a) and (26b).

- (26) a. [*shì* [[*Zhangsan*]_{focus} *yaolai*]] (cleft)
 b. [*shì* [*Zhangsan yaolai*]_{focus}] (propositional assertion)

The linear adjacency between the copula and the post-subject focus constituents in Chinese clefts can be similarly captured by the low attachment constraint. For instance, the default position of an adjunct in Chinese syntax is between the subject and the VP. When *shì* is adjoined at the C-level and associates with an adjunct focus within its c-command domain, the low attachment constraint is predicted to come into effect, causing the subject to scramble/topicalize to a higher C-level position, so that the adjunct is positioned next to the copula morpheme with no ‘intervening’ element in between, i.e. [subject_i [*shì* t_i adjunct_{focus} VP]].²³

In the second, constituent-based analysis, *shì* can be seen as a dedicated functional head attracting an exhaustively identified focus constituent via relevant syntactic feature-checking (Horvath 1995; Brody 1995; Horvath 2010). This solution where the copula morpheme triggers focus movement has been independently proposed for a wide range of languages (Kiss 1998; Hiraiwa & Ishihara 2002, 2012; Gribanova 2013). In Germanic languages, the tense marking of the copula depends on that of the embedded V, indicating that the copula originates within the cleft clause and that focus movement takes place (Meinunger 1998).²⁴ (27) illustrates how the cleft and the propositional assertion are derived based on the functional head approach in Hiraiwa & Ishihara (2002, 2012): The copula morpheme initially merges at a focus projection head, posited to be FocP, and triggers the movement of the cleft phrase (see 27a) or the entire clausal focus (see 27b) to [Spec, FocP], again unifying clefts and propositional assertions. The copula morpheme subsequently undergoes head movement to a projection structurally higher than FocP.²⁵

²³ The low attachment analysis further predicts that with a VP under focus, all pre-VP elements (subjects, adjuncts) will be scrambled across *shì*, so that *shì* will attach directly next to the VP focus. A VP cleft is indeed attested in Mandarin (Hole 2011) and also in historical Chinese (Dong 2004).

²⁴ Hiraiwa & Ishihara (2002) has also drawn upon island sensitivity as a diagnostic for probing the existence of focus movement: If one assumes that the copula morpheme attracts an exhaustively identified cleft phrase to undergo syntactic movement, then it is predicted that a locality effect is induced when the cleft phrase’s base position is contained within an island domain. This is shown to be the case for Japanese *da* ‘be’-clefts. Complex NP island violations hence provide evidence for the existence of focus movement. In Mandarin Chinese, nevertheless, judgments from island data are not clear-cut. In (i), the constructed complex NP island sentence does not invite a strongly degraded judgment, and the judgments vary significantly across consultants (based on twelve Mandarin speakers from North China).

(i) Hǎoxiàng shì [nèi-dào tí]_i xiǎoliú zhǎo bú dào [néng gěi tā jiǎng míngbai e_i] de lǎoshī.
 seem COP [that-CLF problem]_i xiaoliu find not out [can give him explain clearly e_i] REL teacher
 ‘Seems like it is [that problem]_i that Xiao Liu couldn’t find a teacher who [could explain e_i to him clearly].’

The absence of a clear island effect in (i) might not be evidence against focus movement, as it could be the result of an empty pronominal being merged locally in the complex NP domain in Chinese, instead of a trace left behind by A’-movement (e.g. Lin 2005; Li 2007). If we further assume that the sentence-initial DP constituent is also base-generated and co-indexes with the empty pronominal, then no movement-created operator-variable binding relation is involved. This non-movement syntactic representation might be available to those speakers who can accept (i).

²⁵ In some other languages, the movement site of the copula morpheme is posited to be a topic projection (Meinunger 1998; Hiraiwa & Ishihara 2002; Sleeman 2012), but the motivation for it is not very clear, other than the need to get the word order right (Hartmann & Veenstra 2013).

- (27) a. $[shi_k [_{\text{FocP}} Zhangsan_i [_{\text{Foc}'} \text{Foc}^0 t_k [_{\text{FinP}} t_i yaolai]]]]$ (cleft)
 b. $[shi_k [_{\text{FocP}} [Zhangsan yaolai]_i [_{\text{Foc}'} \text{Foc}^0 t_k [_{\text{FinP}} t_i]]]]$ (propositional assertion)

Semantically, there are mechanisms compatible with the above focus-marker syntax that capture the exhaustive reading of clefts. For example, Erlewine (2020) proposes that *shì* is a sentence modifier that lexically realizes Velleman et al. (2012)'s CLEFT operator (note that the CLEFT operator may be covert, so that it is not necessarily adopted for a focus-based approach). According to Velleman et al., CLEFT as well as the exclusive operator *only* projects a two-layered meaning that targets both the at-issue and the presuppositional dimension. The denotations of CLEFT and *only* à la Velleman et al. (2012) are presented in (28) (Here we adopt a modified version in Erlewine 2020, the subscript S indexes the current context):

- (28) CLEFT: = $\lambda w. \lambda p$: No true answer is strictly stronger than p at w.

There is a true answer at least as strong as p at w.

only: = $\lambda w. \lambda p$: There is a true answer at least as strong as p at w.

No true answer is strictly stronger than p at w.

Formally,

CLEFT: = $\lambda w. \lambda p$: $\forall q [(q >_S p) \rightarrow \neg q(w)]. \exists q [q(w) \wedge q \geq_S p]$

only: = $\lambda w. \lambda p$: $\exists q [q(w) \wedge q \geq_S p]. \forall q [(q >_S p) \rightarrow \neg q(w)]$

The idea in (28) is meant to capture the intuition that both CLEFT and *only* serve an inquiry-terminating function. Given a question under discussion *who will come?*, an answer in the form of *Only Zhangsan will come* and *It is Zhangsan that will come* both address the question completely so that this line of inquiry is terminated. According to Velleman et al, the meaning components of CLEFT and *only* are identical, the difference being that the mapping of *only*'s two layers of meaning to the two dimensions is the mirror image of CLEFT. The 'switch side' treatment is meant to capture Horn (1981)'s observation that the exhaustive (inquiry-terminating) inference of *only* is derived from the at-issue dimension, whereas that of CLEFT is from the non-at-issue dimension (see fn#9 for a discussion). In the following we only focus on how the cleft interpretation is derived. Once again we illustrate with the Mandarin cleft from (19a) *Shi Zhangsan yao lai*. 'It is Zhangsan that will come', where we assume with Erlewine (2020) that *shì* is the CLEFT operator, the prejacent of which is $p = \wedge \text{Zhangsan will come}$. Suppose the relevant context S contains a QUD *who will come?* and a domain consisting of individuals {Zhangsan, Lisi, Wangwu}, we arrive at a set of propositions based on atomic individuals: $\{\wedge \text{Zhangsan will come}, \wedge \text{Lisi will come}, \wedge \text{Wangwu will come}\}$. These propositions form a Boolean lattice based on the entailment relation, from which we have all the possible Hamblin answers. Using *a*, *b*, *c* to stand for Zhangsan, Lisi and Wangwu, respectively, and using *K* to refer to the extension of *will come*, the set of Hamblin answers are then represented as $\{K(a), K(b), K(c), K(a+b), K(a+c), K(b+c), K(a+b+c)\}$.

We rank the strength of propositions along the entailment scale, meaning that a proposition *q* is at least as strong as a proposition *p* iff *q* entails *p*, and *q* is strictly stronger than *p* iff *q* asymmetrically entails *p*. According to the denotation of CLEFT in (28), the formula between the colon and the period is treated

as a presupposition. Thus, given $p = \wedge \text{Zhangsan will come}$, $\text{CLEFT}(p)(w)$ presupposes that no true answer is strictly stronger than $K(a)$ at w . This means that to satisfy the presupposition, $K(a+b)$, $K(a+c)$, $K(a+b+c)$ cannot be true at w , as they asymmetrically entail $K(a)$ and hence are strictly stronger than $K(a)$. Additionally, $\text{CLEFT}(p)(w)$ asserts that there is one true answer that is at least as strong as $K(a)$ at w . This means that one of the following Hamblin answers $K(a)$, $K(a+b)$, $K(a+c)$, $K(a+b+c)$ must be true. Combining the presuppositional and the at-issue content, we can conclude that $\text{CLEFT}(p)(w)$ is felicitous when $K(a)$ is true at w , while the stronger alternatives $K(a+b)$, $K(a+c)$, $K(a+b+c)$ are all false at w . Therefore, we arrive at an exhaustive interpretation, according to which Zhangsan is the sole individual who will come.

Aside from Velleman et al. (2012)’s semantics that is couched in the sentential modifier analysis of *shì*, an exhaustive semantics couched in the focus head analysis of *shì* can also be formulated, by analogy to the semantics of constituent *only* (see e.g. Rooth 1992; Wagner 2006). We will leave out the details here.

3.3 Capturing the copula-to-cleft pathway under the copular approach

The previous discussions have shown that both the copula-based approach and the focus-based approach are able to account for the basic aspects of the cleft construction. Both derive the tendency for the copula morpheme to be immediately followed by the focus. Both derive the exhaustive focus interpretation of clefts, when interfacing with an appropriate semantic mechanism.

We now point out that the diachronic copula-to-cleft pathway identified above (summarized briefly in Table 3, see Figure 1 above for more details) can be drawn upon to support the copular approach to Chinese clefts.

Copula	Grammatical change process
<i>wéi</i>	Attested uses of cop-cl, prop-assertion & cleft (14th to 11th century BC) → Decline in use of cop-cl, prop-assertion & cleft, reanalysis of <i>wéi</i> as an exclusive adverb (5th to 4th century BC)
<i>shì</i>	Attested uses of cop-cl, prop-assertion & cleft (1st to 9th century AD) → (cop-cl, prop-assertion & cleft) Simultaneous decline in a subset of modern Sinitic languages (Cantonese, Hakka, etc.) (17th century AD)
<i>xì</i>	Attested uses of cop-cl, prop-assertion & cleft (10th-12th century AD) → (cop-cl, prop-assertion & cleft) Simultaneous decline in another subset of modern Sinitic languages (Mandarin, etc.) (17th century AD)

Table 3: A timeline of the grammatical development of the three copula morphemes

To start with, let’s assume in accordance with previous works (Harbsmeier 1981; Pulleyblank 1995; Hale 1998, 2007; Herforth 2003; LaPolla 2015) that word order, clause structure and construction type have remained largely stable for the previous stages of Chinese syntax. That is, Chinese has consistently been a predominantly SVO language, and similarly in a Chinese copular structure, the copula has

consistently intervened between the subject and the complement. We further assume that adopting a copular approach of clefts amounts to saying that a cleft structure has consistently been an underlyingly copular structure that is headed by a copula.

We now start with the grammatical development of the *wéi*-copula. The three copula-related functions in question are repeated from (8) for illustration.

- (29) a. Yà bīn qí **wéi** chén.
nobel Bin FUT WEI minister
'The noble Bin will be minister.'
- b. **Wéi** [wáng fá _ fāng].
WEI king fight missing.tribe.name tribe
'It is the case that [the king will fight the (illegible tribe name) tribe].'
- c. **Wéi** [Fù Jiǎ tā wáng].
WEI Fu Jia place.a.curse.on king
'It was [Fu Jia] who placed a curse on the king.'

Under the copular approach, the fact that *wéi* appeared in all three constructions follows from the structural homogeneity among these structures. Specifically, we follow the copular analysis of Cheng (2008) laid out above and assume that the copular structures in (29a)-(29c) are analyzed as in (30a)-(30c), respectively.

- (30) a. *Yabin_i wei* [SC [subject *t_i*] [predicate *Chen*]]
- b. *pro_i wei* [SC [subject [CP *Wang fa _ fang*]] [predicate *t_i*]]
- c. *pro_i wei* [SC [subject [CP *Fu Jia ta wang*]] [predicate *t_i*]]

We thus account for *wéi*'s pattern of development by assuming that as the propositional assertion and the cleft structure emerged on the basis of simple *wéi*-copular clauses, the lexical insertion of the copula *wéi* to these structures was triggered.

The copular approach similarly predicts the paths of change following the categorial reanalysis of *shì* and *xì* as a new copula item. Under the copular approach, the emergence of a new copula morpheme in a historical stage means that learners acquire a new lexical item for the category of copula verb within her/his lexicon. As a new copula member in the lexicon, the lexical insertion of this morpheme to the verbal head of a copular clause structure is triggered. Given that in this case, the propositional assertion and the cleft structure are already present in the learner's grammar, we expect a straightforward extension process, accounting for both the case when the *shì*-copular structure extends to *shì*-propositional assertions and *shì*-clefts, as well as a similar case for the *xì*-copula. Importantly, that the above process of expansion will be replicated each time a new copula use emerges is expected, since the process of copula development can be reduced to a reanalysis-and-extension process (Harris & Campbell 1995): The reanalysis of a morpheme as instantiating a copula verb category results in the extension of this morpheme to structures that host the copula verb category.

The copular approach also explains the pattern of decline witnessed in the three copula morphemes. For *wéi*, the initial use in copular clauses, clefts and propositional assertions gave way to a distinct exclusive operator use with distinct distributional behaviors. According to the copular approach, this change came about via a reanalysis: The copula verb is required in a cleft structure, and when external change caused the structure to be no longer analyzed as a cleft by learners, the exhaustive focus interpretation associated with the structure leads to a reinterpretation of the copula verb as an exclusive adverb. Note here that the posited lexical change of *wéi* from a copula verb into an exclusive adverb is not to be conflated with the analysis of *wéi* under the focus-based approach to clefts: In the former case, it is assumed that *wéi* is a copula verb in clefts. The cleft structure later changed into a different structure, and this change then triggered the reanalysis of *wéi* into a different category, which is also accompanied by a change of *wéi*'s distributional behaviors (See section 2.2 for discussions). As *wéi* no longer retains the copula verb meaning, its disappearance from the copular clause is accounted for.²⁶

We have further seen that the use of *shì* has declined in a number of modern Sinitic languages, and the use of *xì* has declined in other Sinitic languages. Crucially, in each language the decline process is coordinated in the three environments: We do not find a case where a copula morpheme is unproductive in copular clauses, but remains productive in clefts/propositional assertions, or the other way round. According to the copular approach, such coordinated decline pattern follows from the assumption that the copular clause and the two focus-marking constructions are underlied by the same copula verb. It follows that the loss of productivity of a given copula verb will lead to its decline in use across the three environments.

We now show that compared against a copula-based analysis, a focus-based analysis does not fare equally well in capturing the full array of diachronic data. To account for the patterns where a newly emerged copula expanded to propositional assertions and to clefts, the focus-based approach may posit that in three separate, individual cases, the copula verb underwent change into a focus marker, and afterwards the focus-marking copula morpheme started to appear in clefts and propositional assertions. A severe difficulty encountered by the focus-based approach is the fact that a change taking place in the focus-marking uses of the copula morpheme always affects the copular clause use: When *wéi* in clefts changed into an *only*-like adverb, *wéi* also disappeared from copular clauses. Furthermore, the loss of

26 In our paper, we consider it possible that the copula verb in a cleft sentence (under the assumption that the cleft sentence is a copular structure) may undergo further development into a distinct focus marker. Here we assume that there are *only*-like adverbs marking exclusivity in Chinese, which head an exclusive focus construction, and the copula in the cleft may be reanalyzed into such an exclusive adverb. If/when this reanalysis takes place, the resulting exclusive focus construction is not a cleft structure: The exhaustive focus being expressed pertains to the at-issue dimension, and the two constructions have a number of distributional differences. Note that the focus-based approach to *shì*-clefts, in treating *shì* as a focus-sensitive operator, also maintains that *shì* is different from an *only*-adverb. *Shì* encodes exhaustive focus from the non-at-issue level, and exhibits other distribution behavior differences from the *only*-adverb.

We assume what is disallowed in Chinese grammar is a focus-based cleft structure, which would be nondistinguishable from the copular-based cleft structure in terms of distributional behaviors. Our paper rules out the possibility that a copular-based cleft may be reconfigured into a focus-based cleft while remaining unaltered in terms of syntactic distribution.

productivity of *shì* and *xì* in clefts/assertions correlated with their loss of productivity in copular clauses. A focus-based analysis of Chinese clefts would commit to positing two homophonous lexical entries for the copula morpheme in the copular clause and the copula morpheme in the propositional assertion/cleft, respectively. A direct consequence is the lack of convincing reason to account for why the loss of both lexical entries should be closely correlated.

In sum, the above discussions provide evidence that the copular approach to clefts better captures the diachronic development in Chinese. Note that our discussions lend support to the general idea that clefts are copular clauses, and do not by itself support any specific analysis within the copular approach. We will need more investigations to understand whether the historical data can lend further support to a particular copula-based analysis.

4 Conclusion

We point out a substantive link between three independent and chronologically separate grammatical developments in historical Chinese. Three copula morphemes emerged in the copular clauses and the focus-marking cleft construction, and subsequently both uses in all three copula morpheme have declined.

The individual copular and cleft uses of the three morphemes have already been observed in the literature. Much of our empirical novelty is given over to the observation that these uses form a structured pattern. That is, they represent a recurring trend in the grammatical development of Chinese.

Taken from here, this account goes on to argue for a copular characterization of Chinese clefts. We show how the recurring trend of copula verbs occurring in clefts presents a convincing case for the syntactic and semantic homogeneity of the copular clause and the cleft. In all, the current analysis provides an example that the diachronic development of clefts, in addition to being worthy of a study in its own right, provides a useful probe into the cleft structure in general.

Sources

Historical Chinese corpora

The historical Chinese source texts cited in this paper are listed in the following table. The authors' names are included when they are known.

Periods	Approximate years	Texts
Pre-Classical Chinese	14th to 11th century BC	a. <i>Jiǎgǔwén Héjī Shìwén</i> 'An Annotated Edition of the Collected Oracle Bones Inscriptions' Edited and annotated by Hu Houxuan et al. Beijing 1999: China Social Sciences Press.
	circa 375-360 BC	b. <i>Zuǒ</i> , Annotated and translated by Stephen Durrant, Wai-yee Li and David Schaberg. Seattle 2017: University of Washington Press.
Early Medieval Chinese	551-479 BC	c. <i>Lúnyǔ</i> 'The Analects', Translated by D.C. Lau. London 1998: Penguin.
	circa 403-444 AD	d. <i>Shìshuō Xīnyǔ</i> 'A New Account of the Tales of the World' Written by Liu Yiqing. Edited and translated by Richard Mather. Ann Arbor 2002: University of Michigan Press.
	circa 900-950 AD	e. <i>Dūnhuáng Biànwén</i> 'Dunhuang Transformation Texts' Annotated and translated by Victor Mair. Cambridge 2008: Cambridge University Press.
Late Medieval Chinese	1031-1095 AD	g. <i>Yīmǎo Rùguó Zòuqǐng</i> 'Plea to enter the Country in the Year Yimao', Written by Kuo Shen. Edited and Annotated by Tao Li. Available at https://ctext.org/wiki.pl?if=gb&chapter=402594&remap=gb
	circa 1072 AD	h. <i>Ōuyáng Wénzhōnggōng Jí</i> 'Collections of Ouyang Xiu' Written by Ouyang Xiu. Available at https://ctext.org/wiki.pl?if=gb&res=230618&remap=gb
	circa 1137 AD	i. <i>Sāncháo Běiméng Huìbiān</i> 'Collected Accounts of the Treaties with the North under Three Reigns' Written by Xu Mengxin. Available at https://ctext.org/wiki.pl?if=gb&res=363390&remap=gb

Table 4: Historical Chinese source texts

Modern Chinese corpora

Further details about the modern Mandarin and Cantonese corpora used in this paper can be found in the following web links:

f. HKCanCor: <http://compling.hss.ntu.edu.sg/hkcancor/>

j. CallFriend: <https://catalog.ldc.upenn.edu/LDC2018S09>

References

- Abels, Klaus. 2003. *Successive cyclicity, Anti-locality, and adposition stranding*: University of Connecticut dissertation.
- Adger, David & Gillian Ramchand. 2003. Predication and equation. *Linguistic Inquiry* 34(3). 325–359.
- Akmajian, Adrian. 1970. On deriving cleft sentences from pseudo-cleft sentences. *Linguistics Inquiry* 1. 149–168.
- Aldridge, Edith. 2019. Subject/non-subject movement asymmetries in Late Archaic Chinese. *Glossa* 4(1). 1–38.
- Barros, Matt. 2014. *Sluicing and identity in ellipsis*: Rutgers University dissertation.
- Bearth, Thomas. 1997. Inferential and counter-inferential. *Swahili Forum* 51(4). 203–230.
- Bisang, Walter. 1998. Grammaticalization and language contact, constructions and positions. In Anna Giacalone-Ramat & Hopper Paul (eds.), *Limits of grammaticalization*, 13–58. Philadelphia: John Benjamins Publishing Company.
- Bolinger, Dwight. 1972. A look at equations and cleft sentences. In Firchow Evelyn (ed.), *Studies for Einar Haugen*, Hague: Mouton de Gruyter.
- Brody, Michael. 1995. Focus and checking theory. In István Kenesei (ed.), *Approaches to Hungarian* 5, 29–44. Szeged: JATE.
- Büring, Daniel. 2003. On D-Trees, Beans, and B-Accents. *Linguistics and Philosophy* 26. 511–545.
- Büring, Daniel & Katharina Hartmann. 1997. Doing the right thing. *The Linguistic Review* 14. 1–42.
- Büring, Daniel & Katharina Hartmann. 2001. The syntax and semantics of focus-sensitive particles in German. *Natural Language and Linguistic Theory* 19. 229–281.
- Büring, Daniel & Manuel Križ. 2013. It's that, and that's it: Exhaustivity and homogeneity presuppositions in clefts (and definites). *Semantics and Pragmatics* 6. 1–29.
- Canavan, Alexandra, George Zipperlen & John Bartlett. 2018. *CALLFRIEND Mandarin Chinese-Mainland Dialect, second edition*. Philadelphia: Linguistic Data Consortium.
- Chang, Jung-hsing. 2006. The Chinese copula *shi* and its origin: A cognitive-based approach. *Taiwan Journal of Linguistics* 4(1). 131–156.
- Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Chaves, Rui & Jeruen Dery. 2019. Frequency Effects in Subject Islands. *Journal of Linguistics* 55(3). 475–521.
- Chen, Jun. 2019. Locus of the exhaustive reading in Chinese cleft sentences. In *Exhaustivity in questions and answers workshop*, vol. 1, 508–526.
- Cheng, Lisa. 1973. *On the typology of wh-questions*. New York: Garland.
- Cheng, Lisa Lai-Shen. 2008. Deconstructing the *shi...de* construction. *The Linguistic Review* 25. 235–266.
- Cheng, Lisa Lai-Shen & Laura Downing. 2013. Clefts in Durban Zulu. In Hartmann Katharina & Veenstra Tonjes (eds.), *Cleft structures*, 141–164. Amsterdam: John Benjamins Publishing Company.

- Cheng, Lisa Lai-Shen & Luis Vicente. 2013. Verb doubling in Mandarin Chinese. *Journal of East Asian Linguistics* 22. 1–37.
- Clech-Darbon, Anne, Georges Rebuschi & Annie Rialland. 1999. Are there cleft sentences in French? In Georges Rebuschi & Laurice Tuller (eds.), *The grammar of focus*, 141–164. Amsterdam: John Benjamins Publishing Company.
- Constant, Noah. 2014. *Contrastive topic: Meanings and realization*: University of Massachusetts Amherst dissertation.
- Declerck, Renaat. 1992. The inferential *it is that* construction and its congeners. *Lingua* 87. 203–230.
- Delahunty, Gerald. 2001. Discourse functions of inferential sentences. *Linguistics* 39(3). 517–545.
- den Dikken, Marcel. 2006. *Relators and linkers: The syntax of predication, predicate inversion, and copulas*. Cambridge: MIT Press.
- den Dikken, Marcel. 2013. Predication and specification in the syntax of cleft sentences. In Hartmann Katharina & Veenstra Tonjes (eds.), *Cleft structures*, 35–70. Amsterdam: John Benjamins Publishing Company.
- Djamouri, Redouane. 2001. Markers of predication in Shang bone inscriptions. In Chappell Hilary (ed.), *Sinitic grammar*, 143–171. Oxford: Oxford University Press.
- Dong, Xiufang. 2004. Shi de jinyibu yufahua: you xuci dao cinei chengfen [Further grammaticalization of *shi*: From grammar words to internal elements]. *Dangdai Yuyanxue* 6(1). 35–44.
- Erlewine, Michael. 2020. Mandarin *shì* clefts and the syntax of discourse congruence. In Gergel Remus, Augustin Speyer & Ingo Reich (eds.), *Particles in German, English, and beyond*, 178–189. Amsterdam: John Benjamins.
- Fan, Chaokang. 2000. *A study of the lexicon in sanchao beimeng huibian*: Guizhou University dissertation.
- Feng, Li. 1993. The copula in Classical Chinese declarative sentences. *Journal of Chinese Linguistics* 21(2). 277–311.
- Fiedler, Ines, Katharina Hartmann, Brigitte Reineke, Anne Schwarz & Malte Zimmermann. 2009. Subject focus in West African languages. In Malte Zimmermann & Caroline Féry (eds.), *Information structure: Theoretical, typological, and experimental perspectives*, 141–164. Oxford: Oxford University Press.
- Gao, Mingkai. 1948. *Hanyu yufalun [thesis on Chinese grammar]*. Beijing: The Commercial Press.
- Gribanova, Vera. 2013. Copular clauses, clefts, and putative sluicing in Uzbek. *Language* 89(4). 830–882.
- Guo, Xiliang. 1997. *Hanyushi lunwenji [collections of papers on the history of Chinese]*. Beijing: The Commercial Press.
- Hale, Mark. 1998. Diachronic syntax. *Syntax* 1(1). 1–18.
- Hale, Mark. 2007. *Historical linguistics: Theory and method*. Oxford: Blackwell.
- Han, Chung Hye & Nancy Hedberg. 2008. Syntax and semantics of it-clefts: a tree adjoining grammar analysis. *Journal of Semantics* 25(4). 345–380.
- Harbsmeier, Christoph. 1981. *Aspects of Classical Chinese syntax*. London and Malmö: Curzon.
- Harris, Alice & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge, UK: Cambridge University Press.

- Hartmann, Katharina & Tonjes Veenstra. 2013. Introduction. In Katharina Hartmann & Tonjes Veenstra (eds.), *Cleft structures*, 1–32. Amsterdam: John Benjamins Publishing Company.
- Hashimoto, Mantaro. 1973. *The Hakka dialect*. Cambridge, UK: Cambridge University Press.
- Hedberg, Nancy. 2000. The referential status of clefts. *Language* 76(4). 891–920.
- Heggie, Lorie. 1988. *The syntax of copular structures*: University of Southern California dissertation.
- Heim, Irene. 1999. Notes on superlatives. Manuscript, Massachusetts Institute of Technology.
- Heim, Irene. 2015. Definiteness and indefiniteness. In Klaus von Heusinger, Claudia Maienborn & Paul Portner (eds.), *Handbook of semantics*, 996–1025. Berlin & Boston: De Gruyter.
- Heim, Irene & Angelica Kratzer. 1998. *Semantics in generative grammar*. Oxford: Blackwell Publishers.
- Herforth, Derek. 2003. A sketch of Late Zhou Chinese grammar. In Graham Thurgood & Randy LaPolla (eds.), *The Sino-Tibetan languages*, vol. 3, 59–71. London: Routledge.
- Higgins, Francis Roger. 1979. *The pseudo-cleft construction in english*. New York: Garland.
- Hiraiwa, Ken & Shinichiro Ishihara. 2002. Missing links: Cleft, sluicing, and “no da” construction in Japanese. *MIT working papers in linguistics* 43. 35–54.
- Hiraiwa, Ken & Shinichiro Ishihara. 2012. Syntactic metamorphosis: Clefts, sluicing, and in-situ focus in Japanese. *Syntax* 15(2). 142–180.
- Hole, Daniel. 2011. The deconstruction of Chinese *shì*... *de* clefts revisited. *Lingua* 121(11). 1707–1733.
- Horn, Laurence. 1981. Exhaustiveness and the semantics of clefts. *Proceedings of NELS* 11. 125–142.
- Horvath, Julia. 1995. Structural focus, structural case, and the notion of feature-assignment. In Katalin É Kiss (ed.), *Discourse configurational languages*, 28–64. New York: Oxford University Press.
- Horvath, Julia. 2010. Discourse features, syntactic displacement and the status of contrast. *Lingua* 120. 1346–1369.
- Huang, Chu-Ren. 1992. Certainty and functional uncertainty. *Journal of Chinese Linguistics* 20. 247–288.
- Huang, James. 1982. *Logical relations in Chinese and the theory of grammar*: MIT dissertation.
- Huber, Stefan. 2006. The complex function of *it*-clefts. In Valéria Molnár & Susanne Winkler (eds.), *The architecture of focus*, 549–578. Berlin: Mouton de Gruyter.
- Jacobs, Joachim. 1983. *Fokus und Skalen: Zur Syntax und Semantik der Gradpartikeln im Deutschen*. Tübingen: Niemeyer.
- Jespersen, Otto. 1928. *A modern English grammar on historical principles*. Heidelberg: Carl Winter.
- Jespersen, Otto. 1937. *Analytic syntax*. London: Allen und Unwin.
- Jin, Dawei. 2016. *The semantic-pragmatics interface and island constraints in chinese*: State University of New York at Buffalo dissertation.
- Jin, Dawei. 2017. Copulas and the class of copular constructions in a cross-Sinitic perspective. Talk presented at the North American Conference in Chinese Linguistics. Rutgers University.
- Jin, Dawei. 2020. Copula functions in a cross-Sinitic perspective. *Folia Linguistica* 54(1). 89–132.
- Kenesei, István. 2006. Focus as identification. In Valéria Molnár & Susanne Winkler (eds.), *The architecture of focus*, 137–168. Berlin: Mouton de Gruyter.
- Kiss, Katalin É. 1998. Identificational focus versus information focus. *Language* 74(2). 245–273.

- Klecha, Peter & Martina Martinović. 2015. Exhaustivity, predication and the semantics of movement. In *Proceedings of the 41st annual meeting of the berkeley linguistics society*, vol. 41, Berkeley Linguistics Society.
- Kuno, Susumu. 1973. *The structure of the japanese language*. Cambridge, MA: The MIT Press.
- Lambrecht, Knut. 2000. When subjects behave like objects: an analysis of the merging of S and O in sentence-focus constructions across languages. *Studies in Language* 3. 611–682.
- LaPolla, Randy. 2015. Sino-Tibetan syntax. In William Wang & Chaofen Sun (eds.), *The Oxford handbook of Chinese linguistics*, 45–57. Oxford: Oxford University Press.
- Li, Audrey. 1990. *Order and constituency in mandarin chinese*. Dordrecht: Kluwer.
- Li, Audrey. 2007. Identifying empty categories. Unpublished manuscript, University of Southern California.
- Li, Charles & Sandra Thompson. 1977. Subject and topic: A new typology. In Charles Li (ed.), *Subject and topic*, 419–444. New York: Academic Press.
- Li, Charles & Sandra Thompson. 1989. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Li, Fang-Kuei. 1971. Studies in Old Chinese phonology. *Tsinghua Journal of Chinese Studies* 9(1-2). 1–61.
- Lin, Jonah. 2005. Does *wh-in-situ* license parasitic gaps? *Linguistic Inquiry* 36(2). 298–302.
- Liu, Ying & Hongzhi Xu. 2019. Clefts in Mandarin: How exhaustive are they? A large-scale corpus and experimental study of *shi... (de)*. Talk presented at Theoretical East Asian Linguistics (TEAL) 12. University of Macao.
- Liu, Ying & Yu'an Yang. 2017. To exhaust, or not to exhaust: An experimental study on mandarin *shi*-clefts. In *GLOW in Asia XI*, vol. 2, 103–117.
- Liu, Zhiji, Deshao Zhang, Ken-ichi Takashima & Kehe Zang. 2005. *Modern translations and catalogues of the oracle bones inscriptions*. Nanning: Guangxi Education Press.
- Luke, Kang Kwong & May Wong. 2015. The Hong Kong Cantonese Corpus: Design and uses. *Journal of Chinese Linguistics* 25. 309–330.
- Mather, Richard. 2002. *A new account of tales of the world*. Ann Arbor: University of Michigan Press.
- Meinunger, André. 1998. A monoclausal structure for (pseudo-)cleft sentences. In Tamanji Pius & Kusumoto Kiyomi (eds.), *Proceedings of NELS*, vol. 28, 283–298. University of Massachusetts.
- Meisterernst, Barbara. 2010. Object preposing in Classical and Pre-Medieval Chinese. *Journal of East Asian Linguistics* 19(1). 75–102.
- Mikkelsen, Line. 2005. *Copular clauses: Specification, predication and equation*. Amsterdam: John Benjamins Publishing Company.
- Mikkelsen, Line. 2011. Copular clauses. In Claudia Maienborn, Klaus von Heusinger & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, 1805–1829. Berlin: Mouton de Gruyter.

- Moro, Andrea. 1997. *The raising of predicates: Predicate noun phrases and the theory of clause structure*. Cambridge, UK: Cambridge University Press.
- Partee, Barbara. 1987. Noun phrase interpretation and type-shifting principles. In Jeroen Groenendijk, Dick de Jongh & Martin Stokhof (eds.), *Studies in discourse representation: Theory and the theory of generalized quantifiers*, 115–143. Dordrecht: Foris Publication.
- Paul, Waltraud & John Whitman. 2008. *Shi...de* focus clefts in Mandarin Chinese. *The Linguistic Review* 25(3-4). 413–451.
- Percus, Orin. 1997. Prying open the cleft. In *Proceedings of NELS*, vol. 27, 337–352. University of Massachusetts.
- Peyraube, Alain & Thekla Wiebusch. 1994. Problems relating to the history of different copulas in Ancient Chinese. In Matthew Y. Chen & Ovid J. L. Tseng (eds.), *Honor of William S.Y. Wang: Interdisciplinary studies on language and language change*, 383–404. Taipei: Pyramid Press.
- Potts, Christopher. 2004. *A logic for conventional implicatures*. Oxford: Oxford University Press.
- von Prince, Kilu. 2012. Predication and information structure in Mandarin Chinese. *Journal of East Asian Linguistics* 21(4). 329–366.
- Pulleyblank, Edwin. 1995. *Outline of Classical Chinese grammar*. Vancouver: UBC Press.
- Roberts, Craige. 1996. Information structure in discourse: Towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics* 49.
- Rooth, Mats. 1985. *Association with focus*: University of Massachusetts at Amherst dissertation.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116.
- Ross, John. 1967. *Constraints on variables in syntax*: MIT dissertation.
- Schuessler, Axel. 1987. *A dictionary of Early Zhou Chinese*. Honolulu: University of Hawaii Press.
- Sheil, Christine. 2016. *Scottish Gaelic clefts: Syntax, semantics and pragmatics*. University of California, Berkeley dissertation.
- Shi, Dingxu. 1994. The nature of Chinese emphatic sentences. *Journal of East Asian Linguistics* 3. 81–100.
- Shi, Yuzhi & Jie Xu. 2001. Hanyushi shang yiwen xingshi de leixingxue zhuanbian jiqi jizhi [The process and mechanism of the typological change of the question form in the history of Chinese]. *Zhongguo Yuwen* [*Linguistics in China*] 284. 454–479.
- Sleeman, Petra. 2012. Quantifier preposing in french and italian as a root phenomenon: a syntactic or a pragmatic approach? *Bucharest Working Papers in Linguistics* 14(1). 5–22.
- Sprouse, Jon, Ivano Caponigro, Ciro Greco & Carlo Cecchetto. 2016. Experimental syntax and the variation of island effects in English and Italian. *Natural Language and Linguistic Theory* 34. 307–344.
- Stepanov, Artur. 2007. The end of CED? Minimalism and extraction domains. *Syntax* 10(1). 80–126.
- Stowell, Tim. 1981. *Origins of phrase structure*: MIT dissertation.
- Sun, Pinjian. 2018. *Mingqing gongwen yuti de yufa tezheng yanjiu* [a study of the grammar characteristics of administrative language styles in ming and qing china]: Beijing Language University dissertation.
- Szabolcsi, Anna. 1981. Compositionality in focus. *Folia Linguistica* 15. 141–161.

- Tai, James & Majorie Chan. 1999. Some reflections on the periodization of the Chinese language. In Alain Peyraube & Chaofen Sun (eds.), *Studies on Chinese historical syntax and morphology*, 109–154. Paris: CRLAO.
- Takahashi, Daiko. 1994. *Minimality of movement*: University of Connecticut dissertation.
- Takashima, Ken-ichi. 1990. A study of the copulas in Shang Chinese. *The Memoirs of the Institute of Oriental Culture* 42. 1–92.
- Tang, Ting-chi. 1983. Guoyu de jiaodian jiegou: ‘fenlieju’, ‘fenliebianju’, yu ‘zhunfenlieju’ [Focusing constructions in Chinese: cleft sentences and pseudo-cleft sentences]. In Tang Ting-chi, Cheng Robert & Li Ying-che (eds.), *Studies in Chinese syntax and semantics*, 127–226. Taipei: Student Book Co.
- 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.
- Teng, Shou-Hsin. 1979. Remarks on cleft sentences in Chinese. *Journal of Chinese Linguistics* 7. 101–112.
- Velleman, Leah, David Beaver, Emilie Destruel, Dylan Bumford, Edgar Onea & Liz Coppock. 2012. It-clefts are IT (inquiry terminating) constructions. In Anca Chereches (ed.), *Proceedings of SALT 22*, 441–460. Linguistic Society of America.
- Wagner, Michael. 2006. Association by movement: evidence from NPI-licensing. *Natural Language Semantics* 14. 297–324.
- Wang, Li. 1940. *Papers on the history of the Chinese language*. Beijing: Beijing Press.
- Williams, Edwin. 1983. Semantic vs. syntactic categories. *Linguistics and Philosophy* 6. 423–446.
- Xu, Liejiong. 1986. Free empty category. *Linguistic Inquiry* 17. 75–93.
- Xu, Liejiong. 2006. Topicalization in Asian languages. In Henk van Riemsdijk & Martin Everaert (eds.), *The blackwell companion to syntax*, 137–174. London: Wiley Online Library.
- Xu, Liejiong & Terence Langendoen. 1985. Topic Structures in Chinese. *Language* 61. 1–27.
- Yang, Barry Chung-Yu. 2011. Intervention effects and wh-construals. *Journal of East Asian Linguistics* 21(1). 43–87.
- Yang, Jianqiao. 2010. *Lecture notes of Classical Chinese grammar*. Shanghai: Fudan University Press.
- Yue-Hashimoto, Anne. 1969. The verb ‘to be’ in Modern Chinese. In John Verhaar (ed.), *The verb ‘be’ and its synonyms*, 72–111. Berlin and Heidelberg: Springer.
- Zhan, Fangqiong & Chaofen Sun. 2013. A copular analysis of *shi* in the Chinese cleft construction. *Language and Linguistics* 14(4). 755–789.
- Zhan, Fangqiong & Elizabeth Closs Traugott. 2015. The constructionalization of the Chinese cleft construction. *Studies in Language* 39(2). 459–491.
- Zhu, Yao. 1996. *The focus-marking function of shi in Mandarin Chinese*: University of Minnesota dissertation.