

The development of the Chinese copula *shì* construction

A diachronic constructional perspective

Fangqiong Zhan and Elizabeth Closs Traugott
East China Normal University | Stanford University

This paper investigates the development of the copula *shì* construction in Chinese from the perspective of diachronic construction grammar (Traugott & Trousdale 2013). In prior work the development has been conceptualized in a grammaticalization framework, with focus on the individual expression *shì*, rather than on its development in the context of both the immediate syntactic frame and also of other constructions such as the demonstrative *shì* construction and the copula *wéi* construction. We show that the copula *shì* construction went through various types of expansion and reduction. The change was not unidirectional in the way predicted by a grammaticalization model such as is proposed in e.g. Lehmann (1995, 2004) and Haspelmath (2004).

1. Introduction

In recent years there has been growing interest in construction grammar, a branch of cognitive linguistics. It is a model of the speaker's knowledge of language, and its architecture is based in form-meaning pairings or signs in the extended sense of words, phrases, clauses, and complex sentences (see e.g. Croft 2001; Fillmore *et al.* 1988; Fillmore & Kay 1997; Goldberg 1995, 2006; Sag 2012). It has begun to be adopted in analyses of Chinese (see e.g. Chen 2009; Lu 2004; especially Zhan & Sun 2013 on the copula *shì* in cleft constructions). Although most work on construction grammar to date has been synchronic, there is a growing body of research from a diachronic perspective. The study of the development of constructions has been undertaken mainly in connection with European and European-derived languages (see e.g. Barðdal 2008; Barðdal *et al.* 2015; Bergs & Diewald 2008; Hilpert 2013; Petr 2014; Traugott & Trousdale 2013). Constructional historical studies of Chinese include Bisang (2010), Peng (2013) and Zhan & Traugott (2015) on how the Chinese cleft construction developed over time. The present

study focuses on the development of the copula *shì* construction prior to the development of the cleft. It is a contribution to constructional studies in Chinese, and explores ways in which the perspective of work on constructionalization differs from the perspective of work on grammaticalization.

The paper is structured as follows. Section 2 outlines some basic points about the perspectives on grammar and change adopted. Section 3 provides a brief overview of the data sources and methodology. Section 4 provides a detailed analysis of the development of the copula *shì* construction. Section 5 presents a quantitative analysis of this development. Section 6 is the conclusion.

2. The perspectives on grammar and change adopted

In this section we briefly outline some of the main points relevant to this paper about the architecture of construction grammar and the model of grammar adopted (Section 2.1). We then go on to review some key points about our perspective on change. We start with an overview of constructionalization, a framework for researching the coming into being of new constructions (Section 2.2) and compare this approach with one that might be adopted in a grammaticalization framework (Section 2.3).

2.1 Construction grammar

In construction grammar, the form-meaning pairs or signs that are the building blocks of grammar are called “constructions”, and range on a continuum from fully contentful (lexical) to procedural (syntactic) (see Hoffmann & Trousdale 2013a:1–2). Of the several varieties of construction grammar (see Hoffmann & Trousdale 2013b) we adopt the usage-based model developed by Goldberg (1995, 2006), which embraces all aspects of a speaker’s knowledge of language, including expressions that in the past have often been considered peripheral to grammar, such as idioms and pragmatic markers. Examples of constructions in contemporary Chinese that will be relevant to the present paper include:

- (1) a. Word construction:
 老師
 lǎoshī
 ‘teacher’

b. Nominal predicate copular construction:

他是老師

tā shì lǎoshī

3SG COP teacher

'He is a teacher.'

c. Verbal predicate copular construction:

我的愛好是看電視

wǒ de àihào shì kàn diànshì

SG1 ASSOC hobby COP watch TV

'My hobby is to watch TV.'

Key to Goldberg's view of construction grammar is that constructions are organized in a hierarchic network. Attested instances, such as occur in conversation or in written texts, are "constructs". Types abstracted away from constructs that have "substantive", phonological representation, such as those in (1), are called "micro-constructions". "Schemas" are abstractions over sets of micro-constructions. They may have subschemas in which sets of micro-constructions with similar properties are organized. Figure 1 illustrates a partial copular network (COP is short for copula, Cxn for construction). In English, when VPs are used as subjects or objects, the form is marked as nominal, be it gerund, infinitive, or complement clause. However, in Chinese, when VPs are used as subjects or objects, there is no change in form. Although the VPs denote nominal fact, the form is still verbal and can take verbal inflections, e.g. the perfective marker *le* in the construct of Micro-Cxn 3 in Figure 1. Moreover, diachronically, as we will see in the discussion below, VPs occurred in both subject and post-copula position at the same time as NPs, which means that VPs and NPs co-existed from the time the copular construction emerged.

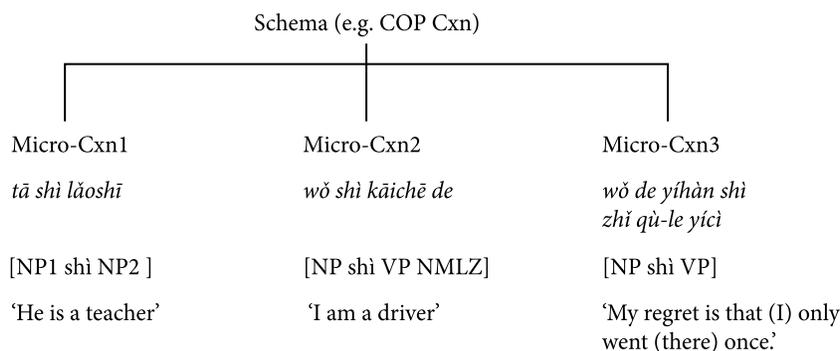


Figure 1. Partial network of copular constructions in Chinese

Schemas always involve structural slots, some to a lesser extent than others. Those with a mix of abstract slots and substantive items are said to be “partially filled”, e.g. [NP1 *shi* NP2], where NP1, NP2 are syntactic slots of a particular kind, and *shi* is substantive. Those that consist of slots only are “unfilled”, e.g. [NP1 COP NP2]. We adopt Booij’s (2010) schematization of an abstract construction as [FORM] ↔ [SEM]. For example, (1b) is an instantiation of Micro-Cx_{n1} such as:

- (2) Predicational copular construction (Modern Chinese)
 [NP_i COP NP_j] ↔ [SEM_i linking;predicational SEM_j]

This is to be read as follows: the formal string [NP copula NP] correlates with the semantics of the first NP linked predicationally to the semantics of the second NP.

For the semantics of the copular construction, we adopt the following terms: “predicational” for the descriptive and property denoting copula (Blom & Daalder 1977; Declerck 1986), and “specificational” for copulas that encode a referential member and non-referential but restricted set relationship (Patten 2012). Given that there is no formal distinction in Chinese between identifying, equational copulas that denote two equative referential entities and more general specificational copulas, we consider equational copulas to be a subclass of specificational ones (see also Patten 2012; Zhan & Sun 2013).¹

2.2 Constructionalization

Over time constructions may come into being, obsolesce, or be reorganized. Here we adopt Traugott & Trousdale’s (2013:22) definition of constructionalization: the development of new conventionalized signs/constructions, i.e. of form_{new}-meaning_{new} pairs at either the substantive or the schematic level, whether contentful or procedural. In this model, constructionalization is the outcome of a series of small-step changes in either form or meaning that are called “constructional changes”; it is typically also followed by constructional changes, as will be illustrated below.²

1. More explanations on the difference between ascriptive (what we call predicational) and specifying (what we call specificational) copular constructions may be found in Huddleston (2002: 266). For the difference between specifying copular constructional and equation/identity statements, see Langacker (1991: 64–71), Petré (2014: 53–54) and Stassen (1997). Many thanks to an anonymous reviewer for the references.

2. The distinction between constructionalization and constructional changes is not without its problems since it is not always easy to tell from the data whether a new construction/change has occurred (for a critique, see Börjars *et al.* 2015). These are problems endemic to attempts to distinguish between change and contexts that enable change (see e.g. Diewald 2006 on the

Three factors are often discussed in work on constructionalization: productivity, schematicity, and compositionality. Over time productivity and schematicity tend to increase and compositionality tends to decrease. Increase in productivity is related to frequency and to generalization of use and meaning. Like Baayen (2001), Barðdal (2008), Bybee (2003, 2006) and Hilpert (2013), among others, we distinguish token from type frequency. Token frequency refers to the frequency of instantiations of a particular construction, or the number of times a particular unit occurs in a text. Type frequency refers to the number of different expressions that are licensed by a particular pattern. Increase in token frequency may enable change in the first place and it may also be the outcome of change. When new constructions are formed, they typically ‘spread by gradually increasing their frequency of use over time’ (Bybee & McClelland 2005: 387). Increase in type frequency (such as increase in the collocates or “host-classes”³ of the copula) is evidence for entrenchment or storage. Such increase in type frequency contributes to the freezing, fixing, and conventionalization in a community of a new construction. Increase in schematicity is related to the extent to which types are generalized over, for example, the extent to which a constructional pattern may sanction new members. Finally decrease in compositionality concerns the transparency of the match between meaning and form. In constructionalization, the transparency of the link between form and meaning tends to become misaligned and therefore reduced.

As we show in Sections 4 and 5, the development of the copula *shi* from Archaic Chinese around 500 BCE to the present involves all the factors mentioned above: constructional changes, constructionalization, increase in productivity and schematicity, and reduction in compositionality at the (sub)schema level of organization. It also involves reanalysis and analogy, two mechanisms that have been much discussed in historical linguistics, particularly in work on grammaticalization (see Section 2.3).

problem in grammaticalization), given the paucity of evidence in the textual record, and also given the fact that language use is dynamic and shifting.

Two other approaches to the distinction between constructionalization and constructional changes may be mentioned here. Smirnova (2015: 82) regards “the succession of the growth and decline of contextual restrictions” as constructionalization, and a wider set of changes that are not necessarily in succession or context-sensitive as “constructional changes”. Hilpert (2013) presents a constructionist approach to language change without the distinction between constructionalization and constructional changes. To him, any changes are constructional changes if they selectively seize a symbolic unit, and alter it in terms of its form, function, frequency and distribution in the community. However, he does not consider the emergence of distinctly new $\text{form}_{\text{new}}\text{-meaning}_{\text{new}}$ constructions that we call constructionalization.

3. This term appears to originate with Himmelmann (2004).

Reanalysis refers to the replacement of old structures by new ones. Analogy, by contrast, refers to the attraction of extant forms to already existing constructions (Hopper & Traugott 2003:64). Langacker (1977:58) defined reanalysis as “change in the structure of an expression or class of expressions that does not involve any immediate or intrinsic modification of its surface manifestation”. Following Andersen (2001:231fn3) we will use the term “neoanalysis” rather than reanalysis, because it does not presuppose that a language acquirer’s knowledge is restructured, only that the analysis is new relative to the earlier one (Traugott & Trousdale 2013:21). Neoanalysis is motivated by parsing, the online analysis of a string of symbols, and is primarily syntagmatic. By contrast, analogy concerns pattern match with other members of a category and the focus is on similarity. The term “analogy” has been used for both motivation and mechanism. Traugott & Trousdale (2013) suggest distinguishing these two as follows. The motivation, which may or may not lead to change, is “analogical thinking”, the cognitive ability to match aspects of meaning and form that we bring to the tasks of comprehension and production. The mechanism leading to change is “analogization”. It brings about a new fit to an extant pattern. When analogization happens, it is simultaneously neoanalysis (Traugott & Trousdale 2013:58) because each case of analogization involves a slight restructuring of what the speaker or hearer knows about a particular expression. While much work on analogy has concerned paradigmatic relations, recently it has been studied in the syntagmatic dimension as well (e.g. De Smet 2012; Fischer 2007). De Smet (2012: 607) points out that “language change often advances most easily where it is least obtrusive, apparently thriving on structural ambiguities and (possibly superficial) resemblances to existing patterns”. After constructionalization, analogization is manifested by gradual, potentially overlapping and sneaky expansion (De Smet 2012:608–609).

2.3 Comparison of constructionalization with grammaticalization

Grammaticalization understood as the study of the ways in which grammatical expressions arise has been of central importance in historical linguistics in recent years. Questions have sometimes been asked whether constructionalization is equivalent to grammaticalization or incorporates it. This is no doubt partly because several of the researchers who work on historical construction grammar were doing research on grammaticalization before construction grammar became a focus of attention, and have tended to revisit examples of change already discussed in terms of grammaticalization. In one of the first papers outlining a historical approach to diachronic construction grammar, Noël (2007) focused on the relation between constructionalization and grammaticalization. As Noël (2007:195) interprets construction grammar, constructions “are by defi-

nition grammatical, so that the historical emergence of constructions amounts to becoming part of the grammar". This might seem to qualify as "grammaticalization", which Lehmann (2004:183) defines broadly as "the creation of grammatical categories", but "becoming part of grammar" in Noël's terms means becoming part of the speaker's knowledge of language. This knowledge ranges over contentful as well as procedural expressions and is therefore far more extensive than the domain Lehmann had in mind. Researchers on grammaticalization, following Meillet (1978[1912]), have been mainly concerned with the development of lexical into grammatical material (e.g. Chinese *liǎo* 'finish' > *le* 'perfective'), and syntacticization of topic-oriented systems into subject-oriented ones (e.g. Old to Middle English).

It follows that constructionalization and grammaticalization cannot refer to the same phenomena, although the former can incorporate aspects of the latter (see e.g. Noël 2007; Hilpert 2013; Traugott & Trousdale 2013; Trousdale 2010). Even procedural constructionalization, the development of procedural constructions (referred to as "grammatical constructionalization" in Traugott & Trousdale 2013) cannot be equivalent to grammaticalization. This is because grammaticalization is not usually conceptualized in terms of schematic templates. A rare exception is Heine (1997), in which schemas for possession, comparison and other domains are proposed. For the most part the focus of work on grammaticalization has been on the development of individual morphemes, which are typically simple or 'atomic' in structure, although their contexts may be complex, e.g. the development of the copula *shi* has been interpreted in terms of the individual, atomic word *shi* (cf. Peyraube & Wiebusch 1994; Wang 1937; Yen 1986), rather than of the larger copular schema (see the introduction to Section 4).

Researchers have tended to approach grammaticalization from two perspectives (Kiparsky 2012:18). One is "How does form change?" (e.g. Lehmann 1985, 2004; Haspelmath 2004). The other is "How does function change?" Here the question is how markers of categories such as aspect, modality, quantification, and complementation arise (e.g. Bybee *et al.* 1994; Heine *et al.* 1991). These two perspectives have contributed to two models of grammaticalization. They are:

- i. grammaticalization as reduction of form (e.g. Lehmann 1995) and as increased dependency (e.g. Haspelmath 2004); this model underlies the hypothesis that grammaticalization is unidirectional,
- ii. grammaticalization as expansion (here focus is on context and function, e.g. Himmelmann 2004).⁴

4. The distinctions are not absolute. For example, despite emphasizing context and function, Heine and his colleagues adhere to the model of grammaticalization as reduction.

Early work on grammaticalization focused on shifts from lexical to grammatical and from pragmatic to syntacticized structures, two types of reduction identified in Meillet (1958[1912]). For example, Lehmann (1985:1) defined diachronic grammaticalization as “a process which turns lexemes into grammatical formatives and makes grammatical formatives still more grammatical” (cf. Kuryłowicz 1965:52). Givón (1979:208) suggested that discourse structure is constantly eroded and “‘pragmatic’ discourse structures develop – over time – into tight ‘grammaticalized’ syntactic structures”. In the late 1990s, the focus of interest shifted from reduction to expansion. Himmelmann (2004) defines grammaticalization as context expansion, and classifies three types of expansion: host class, e.g. of *auxiliary BE going to* from animate subjects only to animate and inanimate subjects; syntactic context, e.g. of articles from core argument position to core and peripheral argument positions; and semantic-pragmatic, e.g. of articles to associative anaphoric uses (*a wedding – the bride, a house – the front door*) (such associative uses are unavailable for demonstratives).

The two models might appear to be orthogonal but in fact they are complementary, because reduction and expansion are intertwined. If a lexical item is reduced to a clitic or an affix, it is used more frequently in a wider range of contexts. If a variety of “pragmatic” discourse expressions is syntacticized, the streamlining of syntactic options results in expanded use of the new structure. Traugott & Trousdale (2013) suggest that a construction grammar approach in which the difference between substantive and schematic constructions is privileged allows us to account for the intertwining of reduction and expansion. While increase in productivity and schematicity, and reduction in compositionality are directional, UNIdirectionality is not expected. This is well supported by evidence from the development of the copula *shi*, as will become apparent in Section 4.

One of the issues that has been widely discussed in work on grammaticalization is the relevance of reanalysis and analogy (for a summary see Traugott 2011). In much of the literature on grammaticalization, reanalysis has been closely associated with grammaticalization but has been considered to be independent because it is not unidirectional, does not imply loss of autonomy or information, and is not gradual (e.g. Haspelmath 1998; Lehmann 2004). However, in the Minimalist tradition, Roberts (1993) suggested that grammaticalization is a subtype of reanalysis. By contrast, analogy has until recently largely been considered to be too unconstrained to be analytically useful (Givón 1991), or has been explicitly downplayed. Indeed Lehmann (2004:161) considers “pure grammaticalization without analogy” to be the norm. However, drawing on Anttila’s (2003:438) observation that “[h]umans are simply analogical animals”, Fischer (2007) and De Smet (2009) argue that analogy, not reanalysis, is the major driving force in change, including grammaticalization. Because construction grammar concerns sets and

schemas, analogy is highly consistent with a constructional approach to language change, and because constructional change affects parts of constructions, it can give a principled account of both analogical thinking and analogization.

3. Data and methodology

This paper makes use of data from the searchable Internet version of the CCL Classical Chinese Corpus⁵ created and managed by Peking University. The data for the quantitative analysis of the development of the copula *shì* construction in Section 5 comprises four books: two Archaic Chinese books *Lunyu* (479–400 BCE) and *Mengzi* (385–303 BCE); *Shiji. Liezhuan* (104–90 BCE), written by the West Han historian Sima Qian (145–90 BCE) at the transition between Archaic Chinese and Medieval Chinese; and one Early Medieval Chinese book *Shishuoxinyu* (432–444 CE). The date of *Lunyu* is taken from Wang (1987), that of *Mengzi* from Takahito (2004), and that of *Shishuoxinyu* from Wei (2002).

The four books are selected because they are widely considered to be standard representatives of the periods of Archaic Chinese and Early Medieval Chinese. This set of four major books is limited compared to the large number of the texts available in the CCL Classical Chinese corpus from these periods. However, the hand coded counts presented in Section 5 clearly illustrate the changes and are therefore informative, although ideally more texts should be searched to provide a finer-grained analysis. In the four books, the first author hand-counted the tokens of *shì*, *shì* ‘to be’, *wéi*, *wéi* ‘to be’, the string [(XP) *shì* XP (PTCP)] and its related variables including adverbs preceding *shì*, the final particle distribution,⁶ and the simple NPs in the string. We use the occurrence of adverbs and the simple NPs (pronouns and proper nouns) preceding *shì* as the evidence for use of *shì* as the copula ‘to be’ (see Section 5.1 for details). In Archaic Chinese, *wéi* was a verb with a variety of meanings including both ‘to do’ (transitive) and ‘to be’ (copula)

5. CCL was built in 2009 and includes data in both Modern Chinese and Classical Chinese. The CCL Classical Chinese corpus contains a list of Chinese texts from the East Zhou Dynasty (Spring and Autumn and Warring states periods) (around 500 BCE) to Republic of China (1911 or so), covering 1059 texts including standard records of history issued by royal family, historical narrations and their commentaries, collective quotes from Masters of a Hundred Lineages, poems and prose, Buddhist and Daoist texts, drama, short stories and philosophy notes (http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp).

6. A sentence final particle denoting illocution: either declarative or question, was frequently found in a copular construction in Archaic Chinese. As the copula *shì* developed, the final particles became less frequent. See Section 5.3 for details.

in English (see Section 4 for details). Occurrence of *wéi* in predicate rather than transitive syntax was used to determine copula use.

Hilpert & Gries (2016) point out that quantitative studies of how units of linguistic structure change across corpus periods can address questions of more general linguistic interest, including: When and how does a given change happen? How can one detect the dynamics of that change? How can one track cases of language variation over time? In this paper we have incorporated modest quantitative analysis within a qualitative study, in keeping with Traugott & Trousdale's (2013: 238) suggestion that qualitative and quantitative approaches are complementary for work in historical linguistics.

4. The development of the copula *shì*

Most research to date on the development of copula *shì* has focused on dating and evidence of grammaticalization, with attention only to the atomic word *shì*. Wang (1937) suggests the copula *shì* developed from the demonstrative pronoun *shì* and that the change did not occur until late Western Han (206 BCE–25 CE) and early Eastern Han (25 CE–220). Peyraube & Wiebusch (1994) develop Wang's proposal and hypothesize that the copula *shì* emerged at the latest in the Qin Dynasty (around 180 BCE). Hong (1957) and Yen (1986) propose that the copula *shì* evolved from the affirmative response *shì*, equivalent of English 'yes/so it is.' Feng (1993) speculates that an overt pause obligatorily occurred in the topic-comment structure between the topic phrase and the comment clause [XP pause [*shì* XP]] and argues that the copula *shì* emerged because of the weakening of the anaphoric function of the demonstrative pronoun *shì* when the pause ceased to be used (Feng 1993: 301). Using the framework of constructionalization, we argue instead that the copula *shì* emerged much earlier with the development of the copular construction. This involved neoanalysis of the topic-comment construction [(XP) [*shì* XP (PTCP)]], as well as analogization to the extant copula *wéi* construction in Archaic Chinese.

In this section, we give a detailed account of the development of copula *shì* from Archaic to Modern Chinese.⁷ We start with an account of the outcome of

7. Periodization for written Chinese is as follows (Chappell 2001:19, taken from Peyraube 1996):

Pre-Archaic Chinese: language of the oracle bone inscriptions	14th – 11th c. BCE
Early Archaic Chinese:	10th – 6th c. BCE
Late Archaic Chinese:	5th – 2nd c. BCE
Pre-Medieval: (transition period)	1st c. BCE – 1st c. CE

these changes: the copula *shì* in Modern Chinese (Section 4.1). In Section 4.2 we give a detailed account of the copula in Archaic Chinese. Section 4.3 suggests an account of the constructionalization of the copula *shì* construction. Probable motivations and mechanisms for the constructionalization are discussed in Section 4.4.

4.1 The copula *shì* in Modern Chinese

The prototypical form of the copular construction in Modern Chinese is [NP1 *shì* NP2], as in (1b) in Section 2.1 above, repeated here as (3) for convenience with structural notation added. In this construction the copula *shì* functions to link NP1 with a predicate NP2 (e.g. Narahara 2002; Zhan & Sun 2013):

- (3) 他 是 老师
tā shì lǎoshī
 3SG COP teacher
 [NP1 COP NP2]
 ‘He is a teacher.’

In (3) *tā* is the subject, *lǎoshī* is the nominal predicate, and the copula *shì* links them together. It has predicational meaning as the nominal predicate *lǎoshī* is a property of the subject ‘he’.

Although the copula is required in standard Mandarin Chinese, it is optional in casual spoken language (as it was in Archaic Chinese, see Section 4.2):⁸

- (4) 今天 星期 五
jīntiān xīngqī wǔ
 today Friday
 [NP1 NP2]
 ‘Today (is) Friday.’

Early Medieval:

2nd – 6th c. CE

Late Medieval:

7th – mid-13th c. CE

Pre-Modern: (transition period)

mid-13th – 14th c. CE

Modern:

15th – mid-19th c. CE

Contemporary:

mid-19th – 20th c. CE

As indicated in Section 2.2, in this paper we use the term Archaic Chinese for Late Archaic Chinese.

8. Examples were searched in Google, accessed 4 October 2015.

- (5) 她 長 頭髮
tā cháng tóufà
 3SG long hair
 [NP1 NP2]
 ‘She (has) long hair/(is) long-haired.’

4.2 The situation in Archaic Chinese

It is generally known that, although a copular construction with *wéi* was optionally available (see Section 4.3 below), copular clauses in Archaic Chinese need not contain any copula verb (e.g. Feng 1993; Wang 1937). There are five basic patterns of copular construction without the copula verb in Archaic Chinese.⁹ Four of these have the simple structure [NP XP (*yě*)], as in (6) to (9).

- (6) unmarked copular sentence:
 窈窕 淑女 君子 好 逑
yáo tiáo shūnǚ jūnzǐ hǎo qíu
 beautiful lady gentleman good spouse
 [NP NP]
 ‘A beautiful lady (is) a gentleman’s good spouse.’
 (Zhouan, *Shijing*, ca. 1000 BCE)
- (7) ...*zhě* ...:
 天 下 者 高祖 天 下
tiān xià zhě gāozǔ tiān xià
 heaven under NMLZ emperor heaven under
 [NP NP]
 ‘The world under heaven (is) emperor’s world.’
 (The biography of Duke Qi and Duke Wu’an from Wei, *Shiji*, 104–90 BCE)
- (8)*yě*:
 仲尼 日 月 也
zhòngní rìyuè yě
 Confucius sun moon PTCF
 [NP NP yě]
 ‘Confucius (is) the sun and the moon.’ (Book 19, *Lunyu*, 479–400 BCE)

9. Thanks to an anonymous reviewer for suggesting that we cite the five basic copular patterns and for Examples (6), (7) and (9).

- (9) ...zhě...yě:
 弓 矢 者 器 也
gōng shǐ zhě qì yě
 bow arrow NMLZ instrument PTCP
 [NP NP yě]
 ‘Bows and arrows (are) instruments.’ (Xici II, *Yijing*, ca. 1050 BCE)

Examples (6) and (7) exemplify the structure [NP NP]. Examples (8) and (9) have the form [NP NP yě]. *Zhě* in (7) and (9) is a nominalizer. As a generalization, the structure of (6) to (9) is [NP NP (yě)]. All examples encode predicational and/or specificational meaning.¹⁰ We call them “classical copular clauses” (CCCs).

Examples (10) and (11) show that a VP or a small clause (S) can also occur in the CCC construction.

- (10) 政 者 正 也
zhèng zhě zhèng yě
 politics NMLZ upright PTCP
 [NP VP yě]
 ‘Governors/politicians (are) upright.’ (Book 12, *Lunyu*, 479–400 BCE)
- (11) 陳良 楚 產 也
chénliáng chǔ chǎn yě
 Chenliang Chu produce PTCP
 [NP s yě]
 ‘Chengliang (is) one that Chu produced (was born in Chu).’
 (Lord Tengwen I, *Mengzi*, 385–303 BCE)

Based on (6) to (11), the syntactic structure of CCCs can be generalized as [NP NP/VP/S (yě)]. If NP/VP/S is further generalized as XP, the syntactic structure of CCCs in Archaic Chinese can be schematized as [NP XP (yě)].

Some linguists, e.g. Shi & Li (2001), propose that the declarative sentence final particle *yě* was required in CCCs in Archaic Chinese and consider it to be the grammatical marker for CCCs in the period. However, (6) and (7) above show that *yě* is optional in CCCs. Moreover, *yě* is not the only final particle found in CCCs. Other final particles also occurred in this position, such as *yǐ* and *ér* in declaratives, *zāi*, *hū* and *yú* in questions, but *yě* was the most frequent. It is therefore more correct to say that a sentence final particle denoting illocution, either declarative or question, was frequently found in a CCC in Archaic Chinese (see (8) to (11) above).

10. An anonymous reviewer suggested that “function” would be preferable to “meaning” here as the copula is zero. In our analysis it is the whole CCC construction that has meaning.

The semantic relation in [NP XP (PTCP)] between NP and XP is a linking one. More specifically, the linking relations of CCC are:

- i. predicational (e.g. Blom & Daalder 1977; Declerck 1986; Zhan & Sun 2013), e.g. (8) and (10), where *rì yuè* ‘the sun and the moon’ and *zhèng* ‘being upright’ are the description of *Zhòngní* ‘Confucius’ and *zhèng zhě* ‘governors’ respectively.
- ii. specificational (e.g. Blom & Daalder 1977; Declerck 1986; Higgins 1979; Patten 2012; Zhan & Sun 2013), e.g. (11), where *Chéngliáng* is the referential member of the non-referential but restricted set of *chǔ chǎn* ‘one that Chu produced’.

The semantic relation in CCC between NP and XP in Archaic Chinese can be formulated as in (12):¹¹

- (12) Classical copular clause Construction (Archaic Chinese)
 [NP_i XP_j (PTCP)] ↔ [SEM_i (linking:predicational/specificational) SEM_j (Illocution)]

Here “linking: predicational/specificational” specifies that the linking relationship is either predicational or specificational.

The CCC is also found in a topic-comment structure in Archaic Chinese, as in (13). This use has been considered to be the fifth basic copular pattern (e.g. Feng 1993; Peyraube & Wiebsuch 1994; Wang 1937).

- (13) ...*shì* ...*yě*:
 富 與 貴 是 人 之 所 欲 也
fù yú guì shì rén zhī suǒ yù yě
 wealth and nobility these people ASSOC thing want PTCP
 [NP [ANA NP *yě*]]
 ‘Wealth and nobility, these (are) the things people want.’

(Book 4, *Lunyu*, 479–400 BCE)

Archaic Chinese was a topic-oriented language (Li & Thompson 1976, 1977) and a topic-comment construction [XP_i [S_j]] ↔ [Topic_i [COMMENT_j]] can be posited for (13). It is generally agreed that *shì* here is a demonstrative pronoun functioning as an anaphor (ANA) referring to the preceding topic phrase (see e.g. Chang 2006; Feng 1993; Peyraube & Wiebsuch 1994; Pulleyblank 1995; Shi & Li 2001; Wang 1937). If so, (13) has the syntactic structure [NP [ANA NP *yě*]] with the demonstrative pronoun *shì* referring anaphorically to the topic NP *fù yú guì* ‘wealth and

11. In Section 4.4 below, we will see that the CCC construction and the copula *wéi* construction are subschemas of the copular construction in Archaic Chinese.

nobility'. In the comment part, the subject¹² *shì* and the nominal predicate *rén zhī suǒ yù* 'the things people want' form a CCC with zero copula.

Example (14) is also a sentence with a topic-comment construction. Here the topic is a conditional clause and the comment is a CCC.

- (14) 如棄 德 不讓 是廢 先君 之舉
rú qì dé bú ràng shì fèi xiānjūn zhī jǔ
 if abandon moral not yield this abolish former emperor ASSOC behavior
 [S [ANA VP yě]]
 也
yě
 PTCP

'If (you) abandon the moral and don't yield, this (is) abolishing the behavior of the former emperor.' (3rd year of Lord Yin, *Zuozhuan*, 403–389 BCE)¹³

Examples (15) to (19) are additional examples of the topic-comment construction with a CCC in the comment.

- (15) 吾無行 而不與二三子者 是丘也
wú wú xíng ér bù yú èr sān zǐ zhě shì qiū yě
 1SG NEG behavior CONN NEG tell PL2 NMLZ this Qiu PTCP
 [NP [ANA NP yě]]

'Having nothing that I cannot tell you guys, this (is) Qiu.'
 (Book 7, *Lunyu*, 479–400 BCE)

- (16) 既欲其生 又欲其死 是惑也
jì yù qí shēng yòu yù qí sǐ shì huò yě
 not only want 3SG live but also want 3SG die this confuse PTCP
 [S [ANA VP yě]]

'Not only to want him to live, but also to want him to die, this (is) confusing.'
 (Book 12, *Lunyu*, 479–400 BCE)

12. Many scholars from both generative and functional traditions agree that topic and subject co-exist in Chinese, e.g. Huang (1982), Li & Thompson (1976, 1981), Shi & Li (2001), Xu & Liu (1998). Li & Thompson (1981: 86–87) characterize topic as what the sentence is about, and subject as the NP that has a “doing” or “being” relationship with the verb in that sentence. Xu & Liu (1998) classify the topic and subject from the perspective of formal structural position. They propose that NP1 in the location of Spec of TopP can be classified as topic while NP2 in Spec of IP is regarded as subject: [NP1[NP2 XP]IP]TopP.

13. The date of *Zuozhuan* is based on Yang (1981).

- (17) 無 父 無 君 是 禽 獸 也
wú fù wú jūn shì qínshòu yě
 NEG father NEG lord this animal PTCP
 [VP [ANA NP yě]
 ‘Those who have no fathers and no lords, they (are) animals.’
 (Lord Tengwen II, *Mengzi*, 385–303 BCE)
- (18) 曰： 是 魯 孔 丘 與？¹⁴
yuē shì Lú Kǒng qīū yú
 say this Lu Kongqiu Q(PTCP)
 [(NP) [ANA NP yú]]
 ‘(Changju) said, ‘(Someone), (is) this Kongqiu from the state of Lu?’
 (Book 18, *Lunyu*, 479–400 BCE)
- (19) 然 而 不 勝 者 是 天 時 不 如 地 利
rán ér bú shèng zhě shì tiān shí bù rú dì lì
 this but NEG win NMLZ this heaven time NEG compare to terrain advantage
 [NP [ANA s yě]]
 也
yě
 PTCP
 ‘Those who (had good timing) but did not win, this (is) that their timing could not compete with the terrain advantage (the city wall).’
 (Gongsun Chou II, *Mengzi*, 385–303 BCE)

Examples (15) to (19) exemplify the syntactic structures [NP [*shì* NP *yě*]], [S [*shì* VP *yě*]], [VP [*shì* NP *yě*]], [() [*shì* NP *yú*]], [NP [*shì* s *yě*]] respectively. The forms of the particular topic-comment constructions exemplified by (13) to (19)¹⁵ can be summarized as [(XP) [*shì* XP (PTCP)]]]. This particular topic-comment construction can be schematized as:

14. As we will show in (34) below, (18) presents potential ambiguity in that it occurs in parallel with a *wéi* construction, which could possibly bring about analogization of *shì* to the extant copula *wéi* construction. Therefore, (18) may have another syntactic analysis: [(NP) COP NP *yú*]. In our view, pragmatic ambiguity cannot be determined in historical texts, it can only be hypothesized. Although Examples (13) to (17) and (19) are standard topic-comment sentences that have been analyzed as involving a CCC, it is possible that some speakers may have had a pragmatically inferred [XP COP XP PTCP] analysis for them.

15. An anonymous reviewer points out that Stassen (1997) predicts class membership predications originate in the zero-coding of identity statements. However, identity statements and class statements co-exist in (13) to (19), and they occur at the same time period. Therefore, it is hard to use these examples as evidence for Stassen’s prediction.

- (20) Topic-comment construction with demonstrative (Archaic Chinese)
 [(XP)_i] [*shì* XP_j (PTCP)] ↔
 [Topic_i [COMMENT(Anaphor (linking:predicational/specificational)
 SEM_j(Illocution))]]

As for the topic-comment construction as a whole, Li & Thompson (1981:95) suggest that the relationship between the topic and the comment is wide open. That is to say, as long as the comment expresses something about the topic in the perception of the speaker and the hearer, the sentence will be meaningful. The topic-comment relationship under discussion here is more restricted. The demonstrative pronoun *shì* frequently¹⁶ functions as the anaphor referring to the complex topic of the topic-comment construction; it entails the semantics of the topic. In this case, the demonstrative pronoun as the anaphor links the semantic meaning of the topic and the predicate part of the comment. The link between Topic and XP in (13) to (15) is pragmatically understood as specificational, the one in (18) as equational/specificational. Here, *shì* introduces a comment on an unexpressed topic (someone) that was introduced in the previous context, see (34) below. The link between Topic and XP in (16), (17) and (19) is pragmatically understood as predicational. In (19), the comment is used to give an explanation / reason. It is descriptive rather than specificational / pure equational, therefore it is predicational.

In anticipation of the next subsection, note that specificational and predicational meanings are also typical of copulas. In the case of the demonstrative anaphor, they are pragmatic inferences from context while in the case of the copula they are semantic (for the semanticization of pragmatic implicatures/inferences see e.g. Eckardt 2006; Traugott & Dasher 2002).

4.3 The emergence of the copula *shì* construction

Evidence for the emergence of copula *shì* is provided by the appearance of examples with adverbs. Peyraube & Wiebusch (1994), among others, propose that since adverbs are words which modify a verb, occurrences of adverbs preceding and modifying *shì* suggest that the demonstrative pronoun had already been reinterpreted as a verb. Example (21) is found in *Mozi* (end of 5th C–4th C BCE). Here adverb *bì* ‘definitely’ occurs preceding and modifying *shì*.

16. Demonstrative pronoun *shì* also occurs in other contexts with other functions. For example: *guó zhī yǒu shì* ‘state Assoc have this, the case of the state of having this’, in which the demonstrative pronoun *shì* occurs in the object position. So anaphoric use is one of the functions of *shì*, but it is a frequent one.

- (21) 不孝子必是怨其亲矣
bú xiào zǐ bì shì yuàn qí qīn yǐ
 not worthy son definitely COP grudge his parents PTCP
 [NP ADV shì VP yǐ]
 ‘An unworthy son definitely is one who grudges his parents.’
 (Jiezhang II, *Mozi*, end of 5th C–4th C BCE)

Example (21) is one of the earliest occurrences of ADV+*shì* in the CCL Classical Chinese Corpus. The structure is no longer a topic-comment construction like those in (13) to (19). In (21) *bú xiào zǐ* ‘an unworthy son,’ is a subject, not a topic, the VP predicate consists of (i) the copula *shì* modified by the adverb *bì*, (ii) the VP *yuàn qí qīn* ‘grudge his parents,’ and (iii) the sentence final declarative marker *yǐ*. Although the copula interpretation is new, the meaning link is not. Semantically, (21) has a predicational linking meaning like (17) and (19), as the verbal predicate ‘to grudge his parents’ is presented as a characteristic of the subject ‘an unworthy son.’ Examples (13) to (19) and (21) are attested at about the same time period, which suggests that the anaphoric demonstrative with pragmatic implicatures of predication and specification coexisted with the newly emerged copula, in which predication and specification had been semanticized.

Example (21) is one of the first occurrences of the copula *shì* in our data. Since it appears about a hundred years earlier than other examples, it can be considered to be an innovation that is a precursor of the construction. Evidence of the conventionalization and therefore constructionalization of this new use is (22), found in *Mengzi* (385–303 BCE). Here the adverb *jūn* ‘totally’ precedes *shì*.

- (22) 公都子問曰：「鈞是人也或為大人，或
gōng dōu zǐ wèn yuē: jūn shì rén yě huò wéi dà rén, huò
 Gongdouzi ask said: totally COP human PTCP Some COP big people, some
 [ADV shì NP yě] [NP1 wéi
 為小人，何也？」
wéi xiǎorén, hé yě
 COP small people, why PTCP
 NP2] [NP2 wéi NP2]
 ‘Gongdouzi asked, ‘(They) totally are human; some of them are good people,
 some of them are bad people; why is that?’ (Gaozi I, *Mengzi*, 385–303 BCE)

In (22), the adverb *jūn* occurring preceding and modifying *shì* shows that *shì*, just as in (21), was already part of the predicate, a copula. In (22), the subject is unspecified, the VP predicate consists of (i) the copula *shì* that is modified by the adverb *jūn*, and (ii) the nominal predicate *rén* ‘human’; *yě* is the final declarative marker. The *shì* clause has a linking meaning, specifically predicational meaning as ‘being human’ is a characteristic of the unspecified subject ‘all of them.’ Furthermore, the

shì clause is parallel to two copula *wéi* clauses. *Wéi* is the conventional copula at this period. In Section 4.4 we will discuss the “competition” between the extant copula *wéi* construction and the emerging copula *shì* in some detail.

We propose that *shì* has been neoanalyzed as one of the instantiations of the copular construction schema on the basis of analogization with the extant copula *wéi* construction, as discussed in the next section.

4.4 Probable motivations and mechanisms for constructionalization of *shì* as a copula

In this section, we consider why and how the change from (20) to (21)–(22) might have come about. We argue that analogical thinking and analogization were the motivation and mechanism for the constructionalization and the replacement of the *wéi* copula by the *shì* copula.

An analysis of the development of the copula *shì* drawing on analogy was earlier proposed by Shi & Li (2001). They argue that the demonstrative pronoun *shì* underwent a process of analogy, and was modeled after the standard transitive verb in Archaic Chinese. They suggest that *shì* occupies the position typically held by a verb in the SVO pattern, hence it was recategorized as a verb. We agree that as the copula *shì* evolved from the demonstrative pronoun it underwent a process of analogization. However, we see several problems with Shi & Li’s analysis:

- i. The analysis pays attention primarily to surface NP *shì* NP order, not to meaning.
- ii. There is no evidence that SVO word order is the exemplar, as the copula *shì* differs from the standard transitive verbs in some syntactically significant ways (e.g. Chao 1968; Li & Thompson 1981), for example, unlike other verbs, it does not occur with aspect markers, and cannot be negated by *méi(yǒu)*.
- iii. Topic-comment structure is ignored.
- iv. The structure of extant copulas is ignored.

We hypothesize that the model was the extant copula *wéi* construction. In Archaic Chinese, *wéi* was a verb with a variety of meanings including both ‘to do’ (transitive) and ‘to be’ (copula) in English. It had the syntactic structure [(XP) *wéi* XP]. For example:

- (23) 為 政 以 德, 譬 如 北 辰
wéi zhèng yì dé pìrú běi chén
 do politics use virtue like north star
 [V NP V NP]
 ‘Doing politics with virtue (is) like the north star.’
 (Book 2, *Lunyu*, 479–400 BCE)
- (24) 民 為 貴, 君 為 輕
mín wéi guì jūn wéi qīng
 people be precious, lords be unimportant
 [NP COP VP]
 ‘People are precious, and lords are less important.’
 (Jinxin II, *Mengzi*, 385–303 BCE)
- (25) 克 己 復 禮 為 仁
kè jǐ fù lǐ wéi rén
 control self restore ritual be benevolence
 [VP VP COP NP]
 ‘Controlling self and restoring rituals are benevolence.’
 (Book 12, *Lunyu*, 479–400 BCE)

Examples (23) to (25) illustrate the major uses of the verb *wéi* in Archaic Chinese. In (23) *wéi* means ‘to do’ and is transitive. In (24) and (25), *wéi* is a copula. Example (24) has the structure [NP COP VP] with the predicational meaning ‘are precious’, and (25) has the structure [VP VP COP NP], in which the two parallel VPs *kè jǐ* ‘control self’ and *fù lǐ* ‘restore rituals’ make up the subject of the sentence, and the VP predicate consists of the copula *wéi* and NP *rén* ‘benevolence’. It has predicational meaning: ‘controlling self and restoring rituals’ are the characteristics that describe the nominal predicate ‘benevolence’.

Examples (26) to (29) are further examples of the copula *wéi* construction from Archaic Chinese.

- (26) 孝 弟 也 者 為 其 仁 之 本 與
xiào tì yě zhě wéi qí rén zhī běn yú
 filial piety PTCP NMLZ be his benevolence ASSOC root PTCP
 [NP COP NP yú]
 ‘The behaviors of filial piety are the root of the benevolence.’
 (Book 1, *Lunyu*, 479–400 BCE)

- (27) 中國 於四 海 內則 在東南 為 陽
zhōngguó yú sì hǎi nèi zé zài dōngnán wéi yáng
 central state at four seas in then at southeast be Yang
 [S COP NP]
 ‘The central state in the world being in southeast is at the side of Yang.’
 (Tianguan Shu, *Shiji*, 104–90 BCE)
- (28) 凡 人 之 欲 為 善 者 為 性 惡 也
fán rén zhī yù wéi shàn zhě wéi xìng è yě
 all people ASSOC want do good NMLZ be nature evil PTCP
 [NP COP S yě]
 ‘The reason that all the people want to do good things is that their nature is evil.’
 (Xing’e, *Xunzi*, end of 3rd C BCE)¹⁷
- (29) 子路曰: [為 孔 丘]
zǐlù yuē: wéi kǒng qiū
 zilü said: be Kongqiu
 [() COP NP]
 ‘Zilü said, ‘(It) is Kongqiu.’
 (Book 18, *Lunyu*, 479–400 BCE)

Example (26) is a specificational copula, (27) and (28) are predicational, and (29), with unspecified subject, is equational.¹⁸ Sentences (24) to (29) have the structure [(XP) wéi XP (PTCP)], and the linking meaning is specificational (including equational) or predicational. The copula *wéi* construction can therefore be schematized as:

- (30) [(XP)_i wéi XP_j (PTCP)] ↔
 [(SEM)_i linking: predicational/specificational SEM_j (Illocution)]

The copula *wéi* is a member of the copular schema, from which it inherits “predicational/ specificational”. The abstract schema can be represented as:

- (31) Copular construction schema (Archaic Chinese)
 [(XP)_i (COP) XP_j (PTCP)] ↔
 [SEM_i linking: predicational/specificational SEM_j (Illocution)]

This schema has persisted throughout the history of Chinese, except that, in standard Chinese, COP is now obligatory, both XPs are usually instantiated by a simple NP, and PTCP is now obsolete.

17. The date of *Xunzi* is taken from Pines (2002).

18. As shown in (34) below, the unspecified subject is “the one who drives the wagon”. Both “the one who drives the wagon” and “Kongqiu” are referential, hence (29) is equational.

In Archaic Chinese, the copular construction schema has two subschemas: the copula *wéi* construction (30) and the CCC (see Section 4.2). Recall that the latter has the structure in (12), repeated here as (33) (see Section 4.1).

- (32) Classical copular clause Construction (Archaic Chinese)
 [NP_i XP_j (PTCP)] ↔
 [SEM_i (linking: predicational/specificational) SEM_j (Illocution)]

The copula *wéi* construction in (30) has similarities to the topic-comment construction in (20), repeated here as (33). Specifically, the structure is similar, and the specificational and predicational semantics given in (31) are consistent with the specificational and predicational pragmatics associated with (33).

- (33) Topic-comment construction with demonstrative (Archaic Chinese)
 [(XP_i) [shì XP_j (PTCP)]] ↔
 [Topic_i [COMMENT(Anaphor (linking:predicational/specificational) SEM_j (Illocution))]]

It is reasonable to hypothesize that the similarity in structure, combined with a pragmatics similar to the semantics of the *wei* copula, made it easy for analogical thinking to take place. “The more closely the pattern looks like a pattern that is already in use, the more easily it may gain ground” (Petré 2014:16).

Now consider the following example from *Lunyu*, in which a possible precursor of the copula *shì* construction appears as (34c) (cited as (18) above). The represented turns are each given a separate line designation (a, b, etc.). Example (34c) is hypothesized to be ambiguous.

- (34) a. 長沮 曰: [夫 執 輿 者 為 誰?]
Chángjū yuē: fū zhí yú zhě wéi shéi
 Changju said: PTCP drive wagon NMLZ be who
 [NP COP NP]
 ‘Changju said, ‘who is the one who drives the wagon?’
- b. 子路 曰: [為 孔 丘]
zǐlù yuē: wéi kǒng qiū
 zilü said: be Kongqiu
 [COP NP]
 ‘Zilü said, ‘It is Kongqiu.’

- c. 曰： [是 魯 孔丘 與?]
yuē: *shì* *lǔ* *kǒngqiū* *yú*
 said: this Lu Kongqiu Q(PTCP)
 [(NP) [NP NP *yú*]]
 [(NP) COP NP *yú*]
 ‘Changju said, ‘(Is) this Kongqiu from the state of Lu?’

(Book 18, *Lunyu*, 479–400 BCE)

Example (34a) has the form [NP₁ *wéi* NP₂] with equational (a subclass of specificational) meaning. *Shéi* ‘who’ activates a topic, but the response is topic-less and subject-less. The form is [(NP) *wéi* NP], and the semantics is equational. The verb *wéi* in both (34a) and (34b) is used as the copula. (34c) has the form [(NP) [*shì* NP *yú*]]. As the topic Kongqiu has been activated in the previous discourse, it is not repeated. The demonstrative pronoun *shì* refers to the unrealized topic and has an equational semantic relation with the nominal predicate of the comment clause *Lǔ Kǒngqiū* ‘Kongqiu from the state of Lu’, and the final particle *yú* marks the clause as a question. The parallel form and meaning between (34b) and (34c) in this cohesive discourse provides an appropriate potentially ambiguous context for analogical thinking to occur: [(NP_i) *wéi* NP_j] ↔ [(SEM_i) linking:equational SEM_j] could be analogically interpreted as [(NP_i) *shì* NP_j] ↔ [(SEM_i) linking:equational SEM_j].

It is possible that either the hearer/reader or the speaker/writer of (34c) neo-analyzed (re-bracketed, re-categorized) the topic-comment construction [(NP_i) [*shì* NP_j *yú*]] ↔ [(Topic_i) [Anaphor SEM_j Q]] as [(NP_i) *shì* NP_j *yú*] ↔ [(SEM_i) linking:equational SEM_j Q] by means of analogization using [(NP_i) *wéi* NP_j] ↔ [(SEM_i) linking:equational SEM_j] as a model. Semantically, when the original demonstrative *shì* was neoanalyzed as the copula, the meaning of anaphoric reference bleached but the equational link between the unspecified subject and the nominal predicate was strengthened.

While (34c) may have been either an actual innovation and precursor, or a potential for such innovation, by the time of *Mengzi*, approximately a hundred years later, unambiguous examples of *shì* are attested, such as (22) in Section 4.3. As discussed there, the adverb *jūn* ‘totally’ preceding and modifying *shì* shows that *shì* was no longer a demonstrative pronoun, but already a verb. We can conclude that the constructionalization of the copula *shì* construction occurred in the 4th C BCE.

As mentioned in Section 2.2, we adopt Traugott & Trousdale’s (2013: 22) definition of constructionalization as the development of signs/constructions, i.e. of conventionalized form_{new}-meaning_{new} pairs at either the substantive or the schematic level. Constructionalization requires both form and meaning change.

The constructionalization of the copula *shì* construction can be represented as in (35), where ‘>’ is short for ‘is analogized and neoanalyzed as’.

- (35) Constructionalization of the copula *shì* construction (Late Archaic Chinese)
 = (20/33) [(XP_i) [*shì* XP_j (PTCP)]] ↔
 [Topic_i [Anaphor (linking:predicational/specificational) SEM_j (Illocution)]]]
 >
 [(XP_i) *shì* XP_j (PTCP)] ↔ [SEM_i linking:predicational/specificational SEM_j
 (Illocution)]

The form changes involve rebracketing such that XP_i is reinterpreted as the subject.

Note that, as is typical of constructionalization (and grammaticalization), it is only copula (not transitive) *wéi* that enabled the change. This also leads to the formal properties of the copula verb *shì* in Modern Chinese that are sufficiently different from the classificatory verbs. The first occurrences of the copula *shì* construction had the structure [(XP) COP XP PTCP].

A demonstrative or demonstrative pronoun changing into a copula is a common grammaticalization process cross-linguistically (e.g. Heine & Kuteva 2002; Pustet 2003; Stassen 1997). For example, in Egyptian, *pw* ‘this’ was a proximal demonstrative, and it was reinterpreted as a copula verb. Hengeveld (1992:250) observes that this evolution “goes hand in hand with a reinterpretation of the theme-clause construction as a subject-predicate construction”. Diessel (1999) discusses the demonstrative-to-copula path of evolution and maintains that the reanalysis of demonstratives as copulas originates in a topic-comment construction. He classifies demonstratives cross-linguistically into four categories: pronominal, adnominal, adverbial, and identificational. In Diessel’s terminology, the demonstrative pronoun *shì* described in this paper is an identificational demonstrative that functions as anaphor. Our discussion of the development of the copular construction that emerged with the reanalysis of the topic-comment construction as a subject-predicate construction gives further support to these earlier findings. Since its emergence, the copula *shì* construction has undergone host-class expansion. [NP1 *shì* NP2] became more and more frequent and acquired new members (see the next section). There were also syntactic and semantic expansions, such as the development of the cleft construction (Zhan & Traugott 2015).

5. Expansion of the copula *shì* construction

In this section we consider ways in which the copula *shì* construction expanded and present quantitative data collected as described in Section 3. First we discuss increase in token frequency with respect to both the copula *shì* and the copula *wéi* (Section 5.1), then increase in adverb types licensed by the copular construction (Section 5.2), generalization of copula without final particle (Section 5.3), changes in the type of pre-copula NP (Section 5.4), and finally the emergence of nominalized predicates (Section 5.5).

5.1 The gain of the copula *shì* construction and reduction of the copula *wéi* construction

The spread of the copula *shì* construction was gradual. As the copular construction came to be used more, use of demonstrative *shì* as an anaphor in the topic-comment construction declined. While there are no unambiguous examples of the copula *shì* in our earliest data, five examples appear in *Liezhuan. Shiji*, and 39 in *Shishuoxinyu*. These represent 0.8% of all uses of *shì* in *Liezhuan* and 12% in *Shishuoxinyu*, as shown in Table 1.

Table 1. The token frequency of *shì* and its use in [NP1 *shì* NP2]

	Token <i>shì</i>	[NP <i>shì</i> NP]
Lunyu	47	0
Mengzi	113	0
Liezhuan. Shiji	635	5 (0.8%)
Shishuoxinyu	326	39 (12%)

The development of the copula *shì* construction affected not only use of the demonstrative *shì* but also that of the pre-existing copula *wéi*. Over time, speakers came to prefer the copula *shì* 'to be' over *wéi* 'to be'. Our method of using the adverbs that occurred preceding *shì* as the criterion for determining copular use of *shì* presumably returned a lower number of the copula *shì* than is actually attested in the data. However, our distinction between *wéi* 'to do' and *wéi* 'to be' based on their semantics in the contexts, presumably returned a fairly accurate assessment of usage of the relevant periods, at least as represented by the texts analyzed.

Table 2 shows that in *Lunyu*, 13% of the verb *wéi* was used as the linking verb 'to be', but none of the *shì* was used as 'to be'. However in *Shishuoxinyu*, 19.6% of the *shì* was used as a copula 'to be', but use of the copula *wéi* was reduced to 6.6%.

This shows that *shì* ‘to be’ significantly increased and *wéi* ‘to be’ decreased by Early Medieval Chinese.

Table 2. Comparison of the use of the copula *wéi* construction with that of the copula *shì* construction

	‘to be’ <i>wéi</i>	Total <i>wéi</i>	‘to be’ <i>shì</i>	Total <i>shì</i>
Lunyu	15 (13%)	115	0 (0%)	47
Mengzi	27 (16.5%)	164	2 (0.2%)	113
Liezhuan. Shiji	376 (29.3%)	1285	6 (0.9%)	635
Shishuoxinyu	45 (6.6%)	686	64 (19.6%)	326

Table 3 organizes the same data to demonstrate the proportion of copula *shì* to copula *wéi*. It shows the decline of *wéi* ‘to be’ in favor of *shì* ‘to be’. It also shows that *shì* ‘to be’ significantly outnumbered *wéi* ‘to be’ by Early Medieval Chinese.

Table 3. The proportion of *wéi* ‘to be’ to *shì* ‘to be’

	‘to be’ <i>wéi</i>	‘to be’ <i>shì</i>	Total ‘to be’
Lunyu	15 (100%)	0 (0%)	15
Mengzi	27 (93%)	2 (7%)	29
Liezhuan. Shiji	376 (98%)	6 (2%)	382
Shishuoxinyu	45 (41%)	64 (59%)	109

Starting with *Liezhuan. Shiji*, *wéi* ‘to be’ became specialized. One significant pattern that appears quite frequently in this text is *wéi...suǒ...*, as in (36).

- (36) 大月氏 王 已 為 胡 所 殺
Dàyuèzhī wáng yǐ wéi hú suǒ shā
 Dayuezhi King already be barbarians whom kill
 ‘King of Dayuezhi has been the one whom the barbarians killed/ King of
 Dayuezhi has been killed by the barbarians.’ (*Liezhuan. Shiji*, 104–90 BCE)

Example (36) shows the pattern *wéi...suǒ...*, in which *wéi* can still be analyzed as ‘to be’, linking two NPs ‘King of Dayuezhi’ and ‘the one whom Hu killed’, and *suǒ* is a pronoun equivalent to English ‘the one’.¹⁹ Increase in use of *wéi ... suǒ...* over *wéi* alone is shown in Table 4.

19. This analysis follows Ma (1898). Other researchers, e.g. Wang (1958), Tang (1991) and one of our anonymous reviewers, consider the pattern to be one of the passive structures in Archaic and Medieval Chinese. Wang (1958) suggests *wéi* also functioned as one of the passive markers in Archaic and Medieval Chinese, and the pattern *wéi...suǒ...* was one of its representations.

Table 4. The proportion of *wéi...suǒ...* to *wéi*

	<i>wéi...suǒ...</i>	<i>wéi</i> 'to be'	Total <i>wéi</i>
Lunyu	0 (0%)	15 (13%)	115
Mengzi	0 (0%)	27 (16.5%)	164
Liezhuan. Shiji	52 (4%)	376 (29.3%)	1285
Shishuoxinyu	25 (3.6%)	45 (6.6%)	686

Table 4 shows that the proportion of occurrences of *wéi...suǒ...* to those of *wéi* increased significantly from *Liezhuan. Shiji* to *Shishuoxinyu*, which means *wéi* 'to be' was gradually confined to the specific *wéi...suǒ...* pattern, while *shì* 'to be' was preferred in other uses.

In Tables 2 to 4 we see that the first three texts show gradual increase in *wéi* before a decisive drop in *Shishuoxinyu*. We speculate that the abrupt-looking drops in *Shishuoxinyu* may be due to the fact that there are 400 years between *Liezhuan. Shiji* and *Shishuoxinyu*, but only 100 between the first three texts. Also, in terms of text size, *Shishuoxinyu* is much smaller than that of *Liezhuan. Shiji*.

5.2 The increase of adverbs preceding *shì*

The appearance of adverbs preceding and modifying *shì* in the string [(XP) *shì* XP (PTCP)] has been used to show that *shì* was no longer a demonstrative pronoun, but a copula (Section 4.3). The increase over time of the token use of adverbs preceding and modifying *shì* in both its demonstrative and copular uses is shown in Table 5.

Table 5. The token frequency of adverb preceding *shì* in the string [(XP) *shì* XP (PTCP)]

	without adv.	with adv.	Total
Lunyu	13 (100%)	0 (0%)	13
Mengzi	29 (93.5%)	2 (6.5%)	31
Liezhuan. Shiji	27 (93%)	2 (7%)	29
Shishuoxinyu	66 (72%)	26 (28%)	92

Table 5 shows that in *Lunyu*, no adverbs precede *shì*, whereas in the Early Medieval Chinese text *Shishuoxinyu*, 28% of the occurrences had an adverb preceding and modifying *shì*. The increase in token frequency of adverbs suggests *shì* was coming to be more and more entrenched as a copula verb. The adverb types licensed by the copula *shì* construction also increased. In our data, the earliest adverb occurring in this environment is *bì* 'definitely' (see (21) above, cited

from *Mozi*, end of 5th C–4th C BCE). Other adverbs that came to be used in the environment include *dìng* ‘definitely’, *jūn, jiē, bìng* ‘totally, all’, *zì* ‘by itself’, *guǒ* ‘as expected’. This is increase in adverb collocates (“host-class expansion”, Himmelmann 2004).

5.3 Generalization of the copula *shì* construction without final particle

As discussed in Section 4.2, a final particle, e.g. *yě, yǐ* and *ér* in declaratives, *zāi, hū* and *yú* in questions, was frequently found (especially *yě*) in a CCC in Archaic Chinese, as in (8) to (11) and (13) to (19) above. After the constructionalization of the copula *shì* construction, the sentence final particle became less and less frequent. Table 6 shows the proportion of uses of *shì* in both its demonstrative and copular meanings.

Table 6. The final particle distribution in the string [(XP) *shì* XP (PTCP)]

	with PTCP	without PTCP	Total
Lunyu	13 (100%)	0 (0%)	13
Mengzi	30 (97%)	1 (3%)	31
Liezhuan. Shiji	26 (90%)	3 (10%)	29
Shishuoxinyu	5 (5%)	87 (95%)	92

Table 6 shows that in *Lunyu*, 100% occurrences with the structure [(XP) *shì* XP (PTCP)] had the sentence final particle, whereas in the Early Medieval Chinese text *Shishuoxinyu*, 95% of the occurrences did not have the final particle. From the perspective of final particle use, there is a decrease by the Early Medieval period, but from that of the copula *shì* there is increase in use of the “bare” copula without particle in the form [(XP) *shì* XP].

5.4 Changes in the type of NP in [NP1 *shì* NP2]

As mentioned in Section 4.2, the pre-copula XP in the string [(XP) *shì* XP (PTCP)] could be a complex NP, VP, S or optional. After constructionalization, more and more simple NPs (pronouns and proper nouns) came to be used in the pre-copula position, and eventually the pre-copula complex phrases gave way to simple NPs, as shown in Table 7. Table 7 also shows that in more than 90% of the occurrences where simple NPs occurred in the pre-copula position of [(XP) *shì* XP (PTCP)], the post-copula XP is an NP as well, i.e. [NP *shì* NP (PTCP)].

Table 7 shows that, in *Lunyu* and *Mengzi*, all the strings occurring in the pre-copula position of [(XP) *shì* XP (PTCP)] are complex, however, in *Shishuoxinyu*,

Table 7. The frequency of simple NPs in the string [(XP) *shì* XP (PTCP)]

	complex pre-copula phrase (including unspecified)	simple pre-copula NP	simple post-copula NP	Total
Lunyu	13 (100%)	0 (0%)	0	13
Mengzi	31 (100%)	0 (0%)	0	31
Liezhuan. Shiji	24 (83%)	5 (17%)	5	29
Shishuoxinyu	53 (58%)	43 (42%)	39	92

close to half are simple NPs. Most of the simple NPs were pronouns, such as *cǐ* ‘this’, *wǒ* ‘I’, *bǐ* ‘he/they’, and proper nouns, such as names of people and places. It should be noted that in *Liezhuan. Shiji*, for the five occurrences where simple NPs occurred in the pre-copula position of [(XP) *shì* XP (PTCP)], the post-copula XP is also an NP. In *Shishuoxinyu*, there are 43 examples with a simple NP preceding *shì*; of these 39 are followed by an NP. As shown above in Section 5.3, along with the constructionalization, use of the final particle gradually decreased, so the structure [NP1 *shì* NP2] became more and more frequent and fixed, and eventually became the prototypical structure of the copular construction in Modern Chinese (Zhan & Sun 2013).

However, as [NP1 *shì* NP2] became entrenched in Early Medieval Chinese, a new, more complex, nominalized NP2 type began to be attested. In Archaic Chinese, the nominalization is marked by *zhě* (see *zhèng zhě* ‘people who are associated with politics (governors, politicians)’ in (10) above, repeated here as (37).

- (37) 政 者 正 也
zhèng zhě zhèng yě
 politics NMLZ upright PTCP
 [NP VP yě]
 ‘Governors (are) upright.’ (Book 12, Lunyu, 479–400 BCE)

In *Shishuoxinyu* there are two examples of this type of nominalization used as NP2 in the copula *shì* construction, one of which is cited in (38). This means the structure [NP *shì* NMLZ] had developed (a morphosyntactic host-class expansion).

- (38) 讓 是 殺 我 侍 中 者 ， 不 可 宥 ！
Ràng shì shā wǒ shìzhōng zhě, bù kě yòu
 Rang COP kill my servant NMLZ, NEG can forgive
 ‘Rang is the one who killed my servant; he cannot be forgiven!’
 (*Shishuoxinyu*, 32–444 CE)

Although only two tokens of [$_{NP}$ *shì* NMLZ] appear in *Shishuoxinyu*, they show the further development of the copula *shì* construction. The process of recruitment of the nominalization involves both host-class expansion and syntactic expansion. Later on, in Late Medieval Chinese, the nominalizer *zhě* was replaced by *de*, and further semantic-pragmatic expansion occurred. Together with syntactic expansion this eventually gave rise to the cleft construction with specificational + contrastive meaning in Late Medieval Chinese around 900 CE (see Zhan & Traugott 2015 for details).

In sum, over time the internal structure of [$_{NP1}$ *shì* $_{NP2}$] changed. The type of $_{NP1}$ was expanded from complex NP to both complex and simple NPs, and $_{NP2}$ was expanded to sanction NMLZ.

5.5 The intertwining of expansion and reduction

Three of the expansions discussed above are accompanied by loss. From a grammaticalization as reduction perspective it is likely that they would be framed as reductions not only internally within the demonstrative *shì*, and copula *wéi* and *shì* constructions individually, but also in terms of paradigmatic options. In summarizing Section 5.1 with focus on reduction, one could say that use of demonstrative *shì* in topic-comment construction was reduced; likewise the token frequency of the copula *wéi* decreased and its distribution was reduced in that it became increasingly restricted to the *wéi...suǒ...* pattern. In summarizing Section 5.3, one could say use of the final particle decreased, and summarizing Section 5.4, one could say use of the more complex XPs was reduced. But if one frames the changes in terms of expansion, one could equally well say, as we did above, that:

- a. as the token frequency of the copula *shì* increased, that of the demonstrative *shì* and the copula *wéi* decreased,
- b. as the use of the bare copula *shì* construction without particle increased, the potential token use of particles decreased,
- c. as simple NPs increased as instantiations of $_{NP1}$, the use of more complex XPs was reduced.

Framing the changes in terms of expansion as well as reduction is more in keeping with constructionalization and increase in sets and schemas than with grammaticalization as a reduction. Furthermore, the expansion of adverbs in pre-*shì* position (Section 5.2) can hardly be framed as a reduction.

Expansions in type frequency are expansions in productivity (Barðdal 2008). Increased productivity led to increased schematicity. Originally, the copular schema in Archaic Chinese only included the copula *wéi* construction and the CCC. When the copula *shì* first emerged and came to be entrenched as the copula

shi construction, it was recruited into the copular schema. This increased the schematicity of the construction. Furthermore, when the copula *shi* construction was first constructionalized, the structure was [(NP) *shi* XP], but over time the form [NP1 *shi* NP2] came to be prototypical of the construction. By the 10th century CE, a new construction-type, the cleft construction came into being and eventually became a subschema of the prototype copula. Increase in schematicity is associated with increase in abstractness of meaning and decrease in compositionality. With the emergence of the cleft, the meaning of the nominalization as a whole was bleached. It no longer denoted an entity, but cued a presupposition made explicit by contrastive focus. Figure 2 demonstrates the development of the copular construction (final ‘-’ signals that the construction still persists):

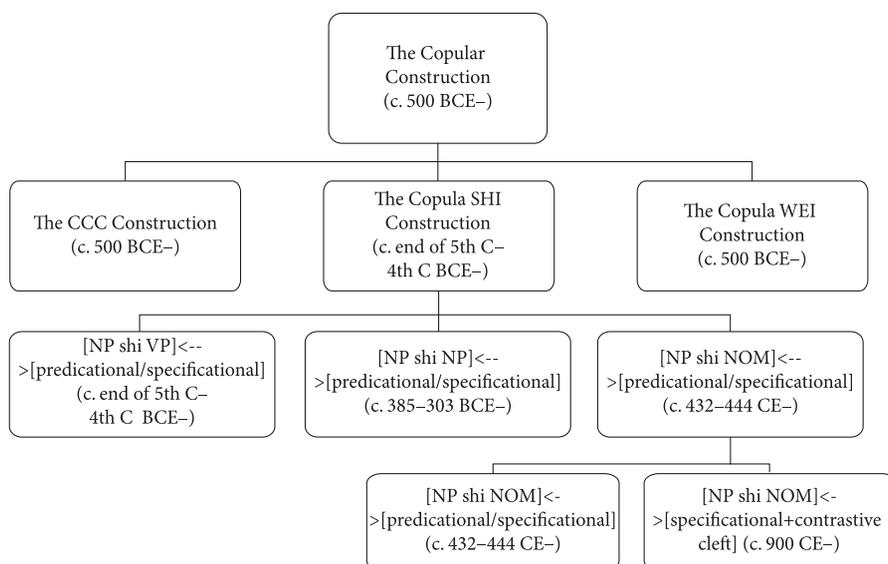


Figure 2. The development of the copula *shi* construction (based on Zhan & Traugott 2015:480)

The emergence of the subject-predicate copula *shi* construction out of the topic-comment construction, plus the decrease of the sentence final particle, led to the decrease of the compositionality of the topic-comment construction. Its compositionality was further decreased as the focal cleft construction emerged out of the copula *shi* construction.

In this section, we have presented quantitative evidence for the gradual development of the *shi* construction, in particular for increased productivity and schematicity as it became increasingly entrenched.

6. Conclusion

In this paper we have given a detailed account of the formation and development of the copula *shì* construction in terms of constructionalization rather than grammaticalization. This has encouraged us to look at the development of *shì* in the context of constructions rather than as a single item out of context. We argue that the onset morphosyntactic context that enabled the constructionalization was the Archaic Chinese topic-comment construction $[(XP_i) [shì XP_j (PTCP)]] \leftrightarrow [(Topic_i) [Anaphor SEM_j (Illocution)]]$. The construction including the demonstrative pronoun *shì* that originally functioned as an anaphor was constructionalized into a copular construction that was incorporated into the copular schema. The mechanisms involved were analogization to the extant copula *wéi* construction $[(XP_i) wéi XP_j (PTCP)] \leftrightarrow [(SEM_i) linking SEM_j (Illocution)]$ as well as neo-analysis (recategorization of *shì* and rebracketing of the clause-internal relations).

The data in Section 5 demonstrate the small steps in the process of constructionalization: the types of adverbs preceding and modifying *shì* increased, the final particles gradually decreased, simple NPs gradually outnumbered complex NPs, and *wéi* 'to be' was gradually replaced by *shì* 'to be'. The changes are discrete, but occur step-wise, micro-step by micro-step (Traugott & Trousdale 2013).

There is evidence for processes of reduction as well as of expansion: the topic-comment sequence was syntacticized as a subject-predicate clause; the final sentence particle became less and less frequent. However, the adverbs preceding and modifying *shì* became more and more frequent; more and more simple NPs occurred in the pre-copula position showing host-class expansion and leading to more and more tokens of $[NP shì NP]$. The recruitment of the nominalization to the post-copula position also shows host-class expansion and syntactic expansion, which eventually led to semantic-pragmatic expansion as well as further syntactic expansion – the emergence of the cleft construction. This shows that increase in productivity and schematicity, and reduction in compositionality are directional; but they are not simply unidirectional in the way traditional work on grammaticalization suggests.

Acknowledgements

Many thanks to two anonymous reviewers for insightful comments and suggestions. We would also like to thank Chaofen Sun for his stimulating feedback at the early stage of this work.

Abbreviations

ADV	adverb
ANA	anaphor
AP	adjective phrase
ASSOC	associative
CCC	classical copula clause
COP	copula
Cxn	construction
NEG	negative
NMLZ	nominalization
NP	noun phrase
PL	plural
PTCP	sentence final particle
Q	question
S	clause
SEM	semantics
SG	singular
VP	verb phrase

References

- Andersen, Henning. 2001. Actualization and the (uni)directionality of change. In Henning Andersen (ed.), *Actualization: Linguistic change in progress*, 225–248. Amsterdam: Benjamins. <https://doi.org/10.1075/cilt.219.11and>
- Anttila, Raimo. 2003. Analogy: The warp and woof of cognition. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 425–440. Oxford: Blackwell. <https://doi.org/10.1002/9780470756393.ch10>
- Baayen, R. Harald. 2001. *Word frequency distributions*. Dordrecht: Kluwer. <https://doi.org/10.1007/978-94-010-0844-0>
- Barðdal, Jóhanna. 2008. *Productivity: Evidence from case and argument structure in Icelandic*. Amsterdam: Benjamins. <https://doi.org/10.1075/cal.8>
- Barðdal, Jóhanna, Elena Smirnova, Lotte Sommerer & Spike Gildea (eds.). 2015. *Diachronic Construction Grammar*. Amsterdam: Benjamins. <https://doi.org/10.1075/cal.18>
- Bergs, Alexander & Gabriele Diewald (eds.). 2008. *Constructions and language change*. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110211757>
- Bisang, Walter. 2010. Grammaticalization in Chinese: A construction-based account. In Traugott & Trousdale (eds.), *Gradience, gradualness and grammaticalization*, 245–277. Amsterdam: Benjamins. <https://doi.org/10.1075/tsl.90.13bis>
- Blom, Alied & Saskia Daalder. 1977. *Syntaktische theorie en taalbeschrijving*. Muiderberg: Coutinho.
- Booij, Geert. 2010. *Construction morphology*. Oxford: OUP.

- Börjars, Kersti, Nigel Vincent & George Walden. 2015. On constructing a theory of grammatical change. *Transactions of the philological society* 113(3). 363–382. <https://doi.org/10.1111/1467-968X.12068>
- Bybee, Joan. 2003. Mechanisms of change in grammaticization: The role of frequency. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 602–623. Oxford: Blackwell. <https://doi.org/10.1002/9780470756393.ch19>
- Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago, IL: University of Chicago Press.
- 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, CA: University of California Press.
- Chappell, Hilary. 2001. *Sinitic grammar: Synchronic and diachronic perspectives*. Oxford: OUP.
- Chen, Mianhua. 2009. Some issues about construction grammar. *Foreign Language Education and Studies* 5. 337–400.
- Croft, William. 2001. *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: OUP. <https://doi.org/10.1093/acprof:oso/9780198299554.001.0001>
- De Smet, Hendrik. 2009. Analysing reanalysis. *Lingua* 119. 1728–1755. <https://doi.org/10.1016/j.lingua.2009.03.001>
- Diewald, Gabriele. 2006. Context types in grammaticalization as constructions. *Constructions* SV, 1–9.
- Diessel, Holger. 1999. *Demonstratives: Form, function and grammaticalization*. Philadelphia: Benjamins. <https://doi.org/10.1075/tsl.42>
- Eckardt, Regine. 2006. *Meaning change in grammaticalization: An enquiry into semantic reanalysis*. Oxford: OUP. <https://doi.org/10.1093/acprof:oso/9780199262601.001.0001>
- Feng, Shengli. 1993. The copula in classical Chinese declarative judgment sentences. *Journal of Chinese Linguistics* 21. 277–311.
- Fillmore, Charles J. & Paul Kay. 1997. Berkeley Construction Grammar. Available online at <http://www.icsi.berkeley.edu/~kay/bcg/ConGram.html>
- Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions. *Language* 64. 501–538. <https://doi.org/10.2307/414531>
- Fischer, Olga. 2007. *Morphosyntactic change: Functional and formal perspectives*. Oxford: OUP.
- Givón, Talmy. 1979. *On understanding grammar*. New York, NY: Academic Press.
- Givón, Talmy. 1991. The evolution of dependent clause morpho-syntax in Biblical Hebrew. In Elizabeth Closs Traugott & Bernd Heine (eds.), *Approaches to grammaticalization* Vol. II., 257–310. Amsterdam: Benjamins. <https://doi.org/10.1075/tsl.19.2.14giv>
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: OUP.
- Goldberg, Adele E. 2013. Constructionist approaches. In Thomas Hoffmann & Graeme Trousdale (eds.), 15–31.
- Haspelmath, Martin. 1998. Does grammaticalization need reanalysis? *Studies in Language* 22. 315–351. <https://doi.org/10.1075/sl.22.2.03has>

- Haspelmath, Martin. 2004. On directionality in language change with particular reference to grammaticalization. In Olga Fischer, Muriel Norde & Harry Perridon (eds.), *Up and down the cline: The nature of grammaticalization*, 17–44. Amsterdam: Benjamins.
<https://doi.org/10.1075/tsl.59.03has>
- Heine, Bernd. 1997. *Cognitive foundations of grammar*. New York, NY: OUP.
- Heine, Bernd, Ulrike Claudi & Friederike Hünne Meyer. 1991. *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.
- Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: CUP.
<https://doi.org/10.1017/CBO9780511613463>
- Hengeveld, Rob. 1992. *Dynamic biogeography*. New York, NY: CUP.
- Higgins, Roger Francis. 1979. *The pseudo-cleft construction in English*. New York, NY: Garland.
- Hilpert, Martin. 2013. *Constructional change in English: Developments in allomorphy, word-formation and syntax*. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139004206>
- Hilpert, Martin & Stefan Th. Gries. 2016. Quantitative approaches to diachronic corpus linguistics. In Merja Kytö & Päivi Pahta (eds.), *The Cambridge handbook of English historical linguistics*, 36–53. Cambridge: CUP. <https://doi.org/10.1017/CBO9781139600231.003>
- Himmelman, Nikolaus P. 2004. Lexicalization and grammaticization: Opposite or orthogonal? In Walter Bisang, Nikolaus P. Himmelman & Björn Wiemer (eds.), *What makes grammaticalization: A look from its fringes and its components*, 21–42. Berlin: Mouton de Gruyter.
- Hoffmann, Thomas & Graeme Trousdale. 2013a. Introduction. In Thomas Hoffmann & Graeme Trousdale (eds.), 1–12.
- Hoffmann, Thomas & Graeme Trousdale (eds.). 2013b. *The Oxford handbook of Construction Grammar*. New York, NY: OUP. <https://doi.org/10.1093/oxfordhb/9780195396683.001.0001>
- Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization* (revised edn.). Cambridge: CUP. <https://doi.org/10.1017/CBO9781139165525>
- Hong, Cheng. 1957. Lun Nanbeichao yiqian de xici ‘On the copula before the Nanbei Dynasty’. *Yuyan Yanjiu* 2. 1–22.
- Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. Cambridge, MA: MIT Phd Thesis.
- Huddleston, Rodney. 2002. The clause complements. In Rodney Huddleston & Geoffrey K. Pullum, *The Cambridge grammar of the English language*, ch4. Cambridge: CUP. <https://doi.org/10.1017/9781316423530.005>
- Kiparsky, Paul. 2012. Grammaticalization as optimization. In Dianna Jonas, John Whitman & Andrew Garrett (eds.), *Grammatical change: Origins, nature, outcomes*, 15–51. Oxford: OUP.
- Kuryłowicz, Jerzy. 1975 [1965]. The evolution of grammatical categories. In Jerzy Kuryłowicz, *Esquisses linguistiques*, Vol. II. 38–54. Munich: Fink. (Originally published in *Diogenes* 51. 55–71, 1965.)
- Langacker, Ronald W. 1977. Syntactic reanalysis. In Charles N. Li (ed.), *Mechanisms of syntactic change*, 57–139. Austin: University of Texas Press.
- Langacker, Ronald W. 1991. *Foundations of cognitive grammar*, vol. 2. Stanford, CA: Stanford University Press.
- Lehmann, Christian. 1985. Grammaticalization: Synchronic variation and diachronic change. *Lingua e Stile* XX. 303–318.
- Lehmann, Christian. 1995. *Thoughts on Grammaticalization*. Munich: LINCOM EUROPA (2nd, rev. edn. of *Thoughts on grammaticalization: a programmatic sketch*, 1982).

- Lehmann, Christian. 2004. Theory and method in grammaticalization. *Zeitschrift für germanistische Linguistik* 32. 152–187. <https://doi.org/10.1515/zfgl.2004.32.2.152>
- Li, Charles N. & Sandra A. Thompson. 1976. Subject and Topic: A new typology. In Charles N. Li (ed.), *Subject and topic*, 458–489. New York, NY: Academic Press.
- Li, Charles N. & Sandra A. Thompson. 1977. A mechanism for the development of copula morphemes. In Charles N. Li (ed.), *Mechanisms of syntactic change*, 419–444. Austin: University of Texas Press.
- Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Lu, Jianming. 2004. The construction grammar and study of Chinese. *Chinese Language* 5. 412–479.
- Ma, Jianzhong. 1898. *Mashi wentong 'Ma's grammar'*. Beijing: Commercial Press.
- Meillet, Antoine. 1958 [1912]. L'évolution des formes grammaticales. In Antoine Meillet, *Linguistique historique et linguistique générale*, 130–148. Paris: Champion. (Originally published in *Scientia (Rivista di scienza)* XXII, 1912)
- Narahara, Tomiko. 2002. *The Japanese copula: forms and functions*. New York: Palgrave Macmillan. <https://doi.org/10.1057/9780230504530>
- Noël, Dirk. 2007. Diachronic construction grammar and grammaticalization theory. *Functions of Language* 14. 177–202. <https://doi.org/10.1075/fol.14.2.04noe>
- Patten, Amanda L. 2012. *The English IT-cleft: A constructional account and a diachronic investigation*. Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110279528>
- Peng, Rui. 2013. A diachronic construction grammar account of the Chinese cause-complement pivotal construction. *Language Sciences* 40. 53–79. <https://doi.org/10.1016/j.langsci.2013.03.004>
- Petré, Peter. 2014. *Constructions and environments: Copular, passive and related constructions*. Oxford: OUP. <https://doi.org/10.1093/acprof:oso/9780199373390.001.0001>
- Peyraube, Alain. 1996. Recent Issues in Chinese Historical Syntax. In C.-T. J. Huang & Y.-H. A. Li (eds.), *New horizons in Chinese linguistics*, 161–213. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-009-1608-1_6
- 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.), *In Honor of William S.-Y. Wang: interdisciplinary studies on language and language change*, 383–404. Taipei: Pyramid Press.
- Pines, Yuri. 2002. Lexical Changes in Zhanguo Texts. *Journal of the American Oriental Society* 122(4). 691–705. <https://doi.org/10.2307/3217610>
- Pustet, Regina. 2003. Copulas: Universals in the Categorization of the Lexicon. *Oxford studies in typology and linguistic theory*. Oxford: OUP.
- Pulleyblank, Edwin G. 1995. *Outline of classical Chinese grammar*. Vancouver: UBC Press.
- Roberts, Ian. 1993. A formal account of grammaticalization in the history of Romance futures. *Folia Linguistica Historica* XIII. 219–258.
- Sag, Ivan A. 2012. Sign-based construction grammar: An informal synopsis. In Hans C. Boas & Ivan A. Sag (eds.), *Sign-based construction grammar*, 69–202. Stanford, CA: CSLI Publications.
- Shi, Yuzhi & Na Li. 2001. *Hanyu yufahua de lichen – xingtai jufa fazhan de dongyin he jizhi 'The path of Chinese grammaticalization – the motivation and mechanism of morphosyntactic change'*. Beijing: Beijing University Press.

- Smirnova, Elena. 2015. Constructionalization and constructional change: The role of context in the development of constructions. In Jóhanna Barðdal, Elena Smirnova, Lotte Sommerer & Spike Gildea (eds.), *Diachronic construction grammar*, 81–106. Amsterdam: Benjamins. <https://doi.org/10.1075/cal.18.03smi>
- Stassen, Leon. 1997. *Intransitive predication*. Oxford: OUP.
- Takahito, Yamada. 2004. Mengzi de Chengshu Niandai ‘The date of Mengzi’. *Limingguan Dongyang Shixue* No.27.
- Tang, Yuming. 1991. *Zhuming zhongnian yuyanxuejia zixuanji* ‘Self-selected articles from famous middle aged linguists’. Hefei: Anhui Education Press.
- Traugott, Elizabeth Closs. 2011. Grammaticalization and mechanisms of change. In Bernd Heine & Heike Narrog (eds.), *The Oxford handbook of grammaticalization*, 19–30. New York: Oxford University Press.
- Traugott, Elizabeth Closs & Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: CUP.
- Traugott, Elizabeth Closs & Trousdale, Graeme. 2013. *Constructionalization and constructional changes*. Oxford: OUP. <https://doi.org/10.1093/acprof:oso/9780199679898.001.0001>
- Trousdale, Graeme. 2010. Issues in constructional approaches to grammaticalization in English. In Katerina Stathi, Elke Gehweiler & Ekkehard König (eds.), *Grammaticalization: Current views and issues*, 51–72. Amsterdam: Benjamins. <https://doi.org/10.1075/slcs.119.05tro>
- Wang, Hongjun. 1987. Guanyu Mingcihua biaoji “zhi” de Xiaoshi ‘The disappearance of the nominalizer ‘zhe’’. *Yuyanxue luncong* 4. 158–196.
- Wang, Li. 1937. Zhongguo Wenfa zhong de Xici ‘The linking verb in Chinese grammar’. In Li Wang, 1958. *Hanyushi lunwenji*, 212–276. Beijing: Beijing Press.
- Wang, Li. 1958. *Hanyushi lunwenji* ‘The collection of articles on the history of Chinese’. Beijing: Beijing Press.
- Wei, Shimin. 2002. Shishuoxinyu ji Zhu Chengshu Niandai Kao ‘The dates of *Shishuoxinyu* and its commentary’. *Changzhou Shizhuan Xuebao* 3. 19–20.
- Xu, Liejiong & Danqing Liu. 1998. *The structure and functions of Topics*. Shanghai: Education Press.
- Yang, Bojun. 1981. *Chunqiu Zuozhuan zhu* ‘The commentary of Chunqiu Zuozhuan’. Beijing: Zhonghua Shuju.
- Yen, Sian L. 1986. The origin of the copula *shi* in Chinese. *Journal of Chinese Linguistics* 14(2). 227–241.
- Zhan, Fangqiong & Chaofen Sun. 2013. A copula 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. <https://doi.org/10.1075/sl.39.2.06zha>

Address for correspondence

Fangqiong Zhan
East China Normal University
International College of Chinese Studies
3663 North Zhongshan Rd.
200062 Shanghai
China
zhjade2000@gmail.com

Co-author information

Elizabeth Closs Traugott
Stanford University
traugott@stanford.edu

Author Queries

- Please provide a complete reference for the citation '(Bybee 2006), (Bybee & McClelland 2005), (De Smet (2012)), (Declerck 1986), (Meillet (1978[1912]))' in this article.
- Please provide a citation for the reference id "CIT0030 (Goldberg, Adele E. 2013)" since citation is missing in the article.