The constructionalization of the Chinese cleft construction*

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

This paper addresses the emergence and development of the Chinese cleft construction, with particular attention to the period from Early Archaic Chinese through Late Medieval Chinese. Prototype copulas are typically of the form [NP SHI NP], are predicational or specificational, and cue information focus. We trace the gradual development over time of copula clefts in addition to prototype copula constructions. A key factor in their development is the use in Medieval Chinese of nominalizations in post-copula position. Copula clefts typically have the form [NP SHI XP DE] and cue both specificational and contrastive meaning. The study is a contribution to the developing field of constructionalization by making more explicit the way in which individual constructional changes contribute sequentially to constructionalization. It also demonstrates one way in which a complex contrastive cleft construction may come into being.

Keywords: copula, cleft, focus, specification, contrast, constructionalization, analogy, reanalysis, history of Chinese

1. Introduction

There has been a growing body of literature on the historical development of information structuring (e.g. Lehmann 2008, Hinterhölzl & Petrova 2009, Batllori & Hernanz 2011). Recently, a constructional perspective has been taken on the development of clefts in English (see Patten 2012 on IT-clefts, Traugott 2008 on pseudo-clefts) and in Chinese (see Zhan 2012, Zhan & Sun 2013). In this paper we contribute further to constructional perspectives on clefts in Chinese, with focus on their development.

* We are indebted to two anonymous reviewers for their insightful comments and suggestions of references.
In many languages, a cleft involves two finite clauses, as in the case of the English IT-cleft (1a) or WH-pseudo-cleft (1b), or the standard Spoken French focus (1c):

(1)  a. It was Dracula who left.
    b. What I want is a jalapeño ice cream.
    c. Y’a le téléphone qui sonne!
        There-has the telephone which rings
        ‘The telephone is ringing.’ (Lambrecht 1994: 4)

One of the finite clauses is the main clause with a semantically empty subject (it) or object (what), the other an “extra-focal clause,” often a relative, as in the examples in (1) (Lehmann 2008: 212). Chinese clefts are, however, considerably different in form: the subject can be fully contentful, and, although there may be an extra-focal relative clause, it is not required.¹ Yet they have a similar kind of semantic structure to English IT-cLEFTs: the focus is contrastive (exclusive and exhaustive) as well as specificationaL (a specific member of a set is selected; for details see Section 3).

In this paper we investigate the emergence and development of the Chinese cleft construction as in (2):²

(2)  a. 我是去年來美國的
    wǒ shì qùnián lái Měiguó (de)
    1SG SHI last.year come US (de)
    ‘It was last year that I came to the US.’
    b. 這個是用烤箱做的
    zhè-ge shì yòng kǎoxiāng zuò de
    this CL SHI using oven cook DE
    ‘It was by using an oven that this was cooked.’
    c. 是他請我來的
    shì tā qǐng wǒ lái de
    SHI 3SG invite 1SG come DE
    ‘It is he who invited me to come.’

The cleft construction is a special copula sentence type in which the copula shì links a subject NP and a nominal predicate that is a nominalization (NOM)

¹. But see footnote 15 for mention of the fact that the distinction between nominalization and relativization in Chinese is a subject of debate.

². In (2) and throughout we use large caps for variables in slots, e.g. NP, CL. Small caps are used for three specific micro-constructions (i.e. construction types): SHI, DE, and ZHE, to distinguish them from the constructs (i.e. instances or tokens of micro-constructions) shì, de, and zhē, and, in the case of DE, as a cover term for historically different representations: Medieval Chinese dǐ and contemporary Chinese de.
marked by the nominalizer \texttt{DE} (e.g. Li & Thompson 1981, Paris 1979, Cheng 2008, Paul & Whitman 2008, Hole 2011). Adopting a construction grammar model (e.g. Goldberg 2006, Croft 2001) Zhan & Sun (2013) propose a constructional schematic for the prototypical cleft in Standard Modern Chinese which has the form \[\text{NP COP NOM}\] linked to the meaning ‘specificational and contrastive focus.’

The nominalization has the structure \[\text{XP DE}\], where \text{XP} = \text{VP/S}^3 \text{ (see (2a, b, c) respectively).}^4 \text{ Specificational meaning relates a referential NP to a non-referential but restricted set in a member-class relationship (see e.g. Patten 2012: 46–47). Contrastive focus signals selection by the speaker of an alternative NP from a set. In clefts the immediate post-copula element, e.g., \text{qùnián ‘last year’} in (2a), encodes contrastive focus (exclusiveness and exhaustiveness) and asserts what is different from the presupposition ‘I came to the US at some time’ (see Lambrecht 1994: 213). This type of copula sentence is a construction since the pattern is non-compositional: the meaning of the whole is not derivable from the individual parts.

In earlier historical work, much of it conceptualized within a grammaticalization framework that privileges the development of individual grammatical markers or “grams” (Bybee, Perkins & Pagliuca 1994), focus has been on the development of the individual marker \texttt{shi}. It is widely accepted that \texttt{shi} evolved from a demonstrative pronoun to a copula verb in Archaic Chinese\(^5\) (e.g. Wang 1937, Li 3. Zhan & Sun (2013: 782) suggest that \text{XP} here can be a complex NP (with a relative clause) with the sentence final nominalizer \texttt{de} implicit ((\texttt{de}) in the example below). For example:

\begin{quote}
\text{他是去的台北 (的)}
\text{tā shì qù de Táiběi (de)}
\text{he COP go REL Taipei (DE)}
\end{quote}

‘It was Taipei that he went.’

Since it is controversial, we do not discuss this issue in the paper.

4. Not all \[\text{NP SHI NOM}\] are clefts. It requires both the form and the specificational plus contrastive meaning. For example, (i) is not a cleft but a predicational copula clause.

\begin{quote}
\text{我是开车的}
\text{wǒ shì kāichē de}
\text{sg1 SHI drive DE}
\text{‘I am a driver.’}
\end{quote}

5. 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 ; 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
Several scholars hold that after demonstrative *shì* was reanalyzed as a copula, it was reanalyzed as a focus marker through a process of further grammaticalization in Early Medieval Chinese (Shi & Li 2001, Dong 2004, etc.). In this paper we argue that *shì* was in fact not further grammaticalized into a focus marker, but rather continued to be used as a copula in a new contrastive focus cleft construction that emerged in Late Medieval Chinese.

Hypotheses that *shì* underwent secondary grammaticalization as a focus marker assume a largely atomic perspective on change: the history of an individual item, in this case *shì*, is the object of inquiry. However, it is impossible for a particular item, and most especially for a component of information structure, to be grammaticalized in isolation. As Bybee, Perkins & Pagliuca (1994: 297) pointed out, “Everything that happens to the meaning of a gram happens because of the contexts in which it is used” (see also Diewald 2002, Heine 2002, Himmelmann 2004), so a more holistic approach is needed that takes the construction, in this case, the whole cleft, into consideration as the context.

One model of grammar that allows us to take such a holistic view is construction grammar. Instead of looking at how *shì* developed atomically as previous researchers have done, we aim to address how the cleft construction as a whole came into being in the history of Chinese. We use the perspective of constructionalization (Traugott & Trousdale 2013) and Zhan & Sun’s (2013) analysis of the Chinese cleft. From these perspectives, *shì* is still an invariant copula verb.

The paper is structured as follows: Section 2 introduces the concepts of constructionalization and constructional change. Section 3 outlines the historical emergence of the cleft construction in addition to the prototype copula schema with specificational and predicational (descriptive) subschemas. Section 4 suggests a model of the development of the cleft construction. Section 5 explores why and how the cleft construction came into being. Section 6 is the conclusion.

2. Constructions and constructionalization

Constructionalization is a framework that has been explored only in the last decade or so. It has been developed primarily in connection with the history of European languages, e.g. Czech (Fried 2008), Dutch (Norde, De Clerck & Colleman 2014), English (Hilpert 2013, Traugott & Trousdale 2013), and Icelandic (Barðdal 2008). There is, however, a growing body of literature on constructionalization in Chinese, e.g. Bisang (2010), Zhan (2012), and Zhan & Sun (2013), which has tested the

---

6. Thanks to an anonymous reviewer for this reference.
cross-linguistic validity of the framework and refined some of the concepts. In this section we introduce those aspects of Traugott & Trousdale’s (2013) framework of constructionalization framework that are relevant for our analysis of the development of the *shì*-cleft in Chinese.\(^7\) We start with the concept of “construction.”

In construction grammar, a construction is understood as a form-meaning pairing ([FORM] \(\leftrightarrow\) [SEM]). Adapting the formalism for constructions used by Booij (2010), and for the Chinese cleft by Zhan & Sun (2013:759), we propose the formalism in (3) for the cleft that is the focus of the present paper:

\[
(3) \quad [\text{NP}_i \text{ shì} \ [\text{XP}_{\text{DE}}]_j] \leftrightarrow [\text{SEM}_i \text{ specificational + contrastive SEM}_j]
\]

This is to be read as follows: Form is linked to meaning (\(\longleftrightarrow\)) and the semantics (SEM) is indexed to NP and [XP DE], respectively (for further details see Section 3).

Constructions are the basic units of a speaker’s knowledge of language. They are related to each other in a network and at different levels of abstraction ranging from the token utterance (or “construct”) to highly schematic internalized, conventional representation. Traugott & Trousdale (2013) propose a hierarchy of levels of constructions. These may be illustrated by the binominal quantifier construction with members such as *a lot of N* ‘much N’, *a bit of N* ‘little N’ (see Brems 2011) as in (4):

\[
\begin{align*}
(4) & \quad \text{a. Schemas: abstract constructions that are taxonomic generalizations over several more particular constructions, e.g. the binominal quantifier [Ni of a Nj] } \leftrightarrow \text{[Sizei related to SEMj].} \\
& \quad \text{b. Subschemas: subsets of schemas, less abstract, but still schematic constructions, e.g. among the subschemas of the binominal quantifier are: small size (a bit of N), and large size (a lot of N).} \\
& \quad \text{c. Micro-constructions: individual construction types, e.g. members of the small subschema: a bit/shred/jot/iota of N.} \\
& \quad \text{d. Constructs: instances of micro-constructions, tokens of actual use, e.g. They are hacks without a shred of intellectual honesty.}
\end{align*}
\]

The greatest amount of information about a construction is found at the micro-construction level of the categorical taxonomy. Form and meaning pairings subsume subcomponents: on the form side, minimally syntax, morphophonology, and prosody, on the meaning side minimally semantics, pragmatics, and discourse function (e.g. information structuring). Each component can change independently.

\(^7\) For a somewhat different approach to the development of new constructions, see Smirnova (Forthcoming). There constructionalization is understood as the growth and decline of contextual restrictions prior to the development of what Smirnova calls “critical constructions”, the result of changes in “critical contexts” (Diewald 2002).
Constructionalization is the development of micro-constructions over time:

Constructionalization is the creation of form\textsubscript{new}–meaning\textsubscript{new} (combinations) of signs. It forms new type nodes, which have new syntax or morphology and new coded meaning, in the linguistic network of a population of speakers. It is accompanied by changes in degree of schematicity, productivity, and compositionality (Traugott & Trousdale 2013:22).

The mechanism that brings about this form\textsubscript{new}–meaning\textsubscript{new} representation is what is widely known as reanalysis, but is more properly called neoanalysis\(^8\) (Andersen 2001, Traugott & Trousdale 2013:21). Constructional changes differ from constructionalization in that “a constructional change is a change affecting one internal dimension of a construction. It does not involve the creation of a new node.” (Traugott & Trousdale 2013:26) Constructional changes include the development of optional or obligatory complementizer marking, e.g., clitic > inflection, narrowing of meaning from mete `food’ to meat `flesh of an animal used as food.’ There may be several constructional changes on the way to constructionalization, but not all such changes result in constructionalization. Constructional changes that precede constructionalization are called “pre-constructionalization” changes. These typically involve expansion of pragmatics, mismatch between form and meaning, and small distributional changes similar to the “critical contexts” Diewald (2002) identified as enabling grammaticalization. Changes that follow constructionalization are called “post-constructionalization” changes, and typically involve expansion of collocates (Hilpert 2008), phonological reduction, and changes in frequency. Constructionalization is a special case of constructional change in which a new form and meaning pairing results from a sequence of small micro-step adjustments in which form and meaning are neoanalyzed and the new pairing comes to be conventionalized.

All four constructional levels (schemas, subschemas, micro-constructions, and constructs) are of importance at various stages of the constructionalization process. While constructs are the locus of change, they can be innovations only. Over time patterns may emerge leading to conventionalization and constructionalization of a construction type (a micro-construction), and sets of micro-constructions may be organized into schemas and subschemas. It is only when innovations are replicated and become conventionalized that they can be considered to be “changes” (Weinreich, Labov & Herzog 1968).

Particularly important in constructionalization are changes in compositional-ity, schematicity, and productivity. Compositionality concerns the transparency with which form and meaning are matched: the more idiosyncratic a construction,
the less compositional it is. Schematicity concerns the degree of abstraction of a construction, and where a construction is situated in the hierarchy of levels cited in (4) above. Productivity is of two kinds: type productivity, which concerns the degree to which a schema or subschema sanctions new, similar constructions, and token productivity, which concerns the frequency with which individual constructions are used (Barðdal 2008).

3. The rise of the cleft construction

In contemporary Standard Chinese the copula schema has two subschemas. One that has been relatively stable, except with respect to the form of the default copula, since Archaic Chinese is the prototype copula that expresses information focus (see Subsection 3.1). The second subschema is the cleft copula that expresses contrastive focus. This arose in the thirteenth century CE, the outcome of a sequence of changes to individual constructions starting in Archaic Chinese.

We start with precursors and enablers of the development of the cleft: first the crystallization of shì as a standard copula in Early Medieval Chinese (sub-Section 3.1), and then the emergence of the nominalization [XP DE] in Late Medieval Chinese, followed by the development of the combination of the copula with the nominalization, i.e. of the sequence [NP shì XP DE], also in Late Medieval Chinese (sub-Section 3.2). The subsequent emergence in Pre-Modern Chinese (mid-thirteenth century) of the contrastive cleft is the topic of sub-Section 3.3.

3.1 Shì: the copula in Early Medieval Chinese

A copula attested in early Archaic Chinese is wéi ‘to be’ (as early as 5th century BCE). In the later Archaic Chinese period a new copula shì arose (in 4th century BCE such as in Mengzi and Mozi) and by about 5th century CE was generalized as the default copula verb. Shì is generally agreed to have originated in a proximal/distal demonstrative as in (5):

9. There have been rich analyses on the dates for the emergence of the copula shì. For example, Wang (1937) suggests it did not occur until late Western Han (206 BCE-25 CE) and early 151 Eastern Han (25 CE-220); Peyraube & Wiebusch (1994) suggest it started at the latest in the Qin Dynasty (ca.180 BCE); Shi (2002) argues that the copula shì was introduced to the language around 100 BCE; Chang (2006) indicates that the copula shì occurred in pre-Medieval times. We thank two anonymous reviewers for the references.

10. Zhan (2012) provides a statistical analysis of the competition between wéi ‘to be’, and shì ‘to be’ at the time of the rise of the copula shì, and concludes that shì ‘to be’ significantly took over the functions of wéi ‘to be’ in Pre-Medieval Chinese.
(5)  a. 國之有是多矣
   guó zhī yǒu shì duō yǐ
   state assoc have this many PTCL
   ‘States have numerous cases of this kind of thing.’
   Zuozhuan (左傳 403–389 BCE)

   b. 是歲，晉又饑
   shì suì jìn yòu jī
   that year Jin again famine
   ‘That year, the state of Jin again got a famine.’
   Zuozhuan (左傳 403–389 BCE)

It still has these pronominal and modifier functions in Modern Chinese, e.g. 唯利是圖 wéi lì shì tú ‘only profit this attempt (only to attempt the profit)’, in which shì is a demonstrative pronoun referring to lì ‘profit’. The demonstrative pronoun is thought to have been neoanalyzed as a copula in the Archaic Chinese topic-comment construction where the demonstrative pronoun shì functioned as an anaphor referring to the topic phrase.

After the neoanalysis in Archaic Chinese of demonstrative shì as the copula micro-construction that we represent as SHI, a copula construction emerged and is still in use. In the Early Medieval Chinese book, Shi shuo xin yu (5th century CE), there are 66 attested copula sentences. The majority of them have the structure [NP SHI NP] (48 occurrences or 73% out of 66 attested sentences), as exemplified in (6):

(6)  a. 此三人並是高才
   cǐ sānrén bìng shì gāocái
   NP ADV SHI NP
   this three.people totally SHI high.talent
   ‘All these three people are of great talent.’
   Shi shuo xin yu (世說新語 432–444 CE)

   b. 我是李府君親
   wǒ shì lífūjūn-qīn
   NP SHI NP
   1SG SHI Lifunjun.relative
   ‘I am one of Lifunjun’s relatives.’
   Shi shuo xin yu (世說新語 432–444 CE)

The attested copula constructions like (6) encode predicate informational focus: The subject encodes referential given information, usually a topic, and the post-copula predicate as a whole is the informational focus indicating non-referential new information. Both predicational and specificational meanings came to be

11. The date of Zuozhuan is based on Yang (1981).
conventionally associated with the copula construction as it emerged, depending on whether the post-copula NP was descriptive (predicational) or identifying (specificational) (Blom & Daalder 1977, Declerck 1988, Zhan & Sun 2013). (6a) is a predicational copula sentence with the post-copula predicate attributing a property ‘high talent’ to the subject ‘these three people.’ (6b) is a specificational sentence, in which the non-referential but restricted set ‘one of Lifujun’s relatives’ is specified by the unique referential member of the set, the subject ‘I.’ In other words, such clauses involve a “value-variable” relation (Higgins 1979): ‘I’ specifies the value of the variable ‘one of Lifujunj’s relatives.’ Predicational and specificational copulas formed subschemas of the prototype copula early on. It should be noted that both examples in (6) encode information focus but not contrastive focus.\footnote{As an anonymous reviewer points out, shì could still be ambiguous between an anaphoric demonstrative/pronominal and a copula in (6a) if bìng ‘totally’ is a sentence adverb. It is more likely to have been neogrammarialized as a copula since an anaphoric element is not necessary following a simple (one-word) topic NP.}

In addition to the 48 occurrences of [NP shì NP] in Shi shuo xin yu (e.g. (6a)), there are 16 other copula sentences (27% of the 66 copula sentences). [NP shì VP] is exemplified in (7a), and [NP shì S] in (7b). These copula sentences co-exist with the more frequent [NP shì NP] form from inception to Modern Chinese. Semantically, (7a, b) are predicational:

(7) a. 此是爱痴
\[cǐ shì ài qí \text{ shì VP} \]
This shì have love devoted
‘This is having devoted love.’ \textit{Shi shuo xin yu (世說新語432–444 CE)}

b. 此是屋下架屋耳
\[cǐ shì wū xià jià wū ěr \text{ shì S} \]
this shì house under build house just
‘This is just building a house under another house.’ \textit{Shi shuo xin yu (世說新語432–444 CE)}

Shi & Li (2001) argue that shì in (7a, b) is not a copula. It is rather a contrastive focus marker that was further grammaticalized from a copula shì. They say that as a focus marker, shì marks VP \(\text{yǒu qíngchì} \) ‘have love devoted’ in (7a) and NP wū xià ‘house under’ in (7b) and serves like a modern “focus marker.” Shi & Li (2001) base the claim that shì is not a copula on the frequent occurrence in Early Medieval Chinese of shì preceding interrogative WH-words, such as shéi ‘who’, hé ‘what’, i.e. in OV word order as in (8).
Their argument builds on Heine & Reh (1994), who discuss several African languages, in some of which the copula is derived from a demonstrative. Heine & Reh (1994: 179) analyze various kinds of focus and say that “WH-words have an inbuilt focus marker — irrespective of whether it is morphologically present or not.” Shi & Li argue that WH-words in Chinese have inbuilt focus, and the focus is often marked. In declarative clauses, the standard verb phrase in Archaic Chinese had VO word order, e.g.

(9) a. 廢之
fèi zhī
dispose it
‘get rid of it’

b. 享之
xiǎng zhī
‘enjoy it’

By contrast, a verb phrase consisting of a WH word, e.g. shéi ‘who,’ hé ‘what,’ appeared with OV word order, e.g.

(10) a. 誰立
shéi lì
whom crown
‘crown whom?’

b. 何如
hé rú
what like
‘like what?’

Shi & Li (2001: 45) claim that after shì was changed from a demonstrative into a copula verb, due to its low transitivity, WH-words could only occur following shì. Consequently, the focus of the WH-words was no longer marked by word order; instead, speakers chose shì to mark focus. According to Shi & Li (2001: 48), in Early Medieval Chinese, shì served as a focus marker for WH-words and by analogy it started to be used to mark other categorical elements immediately following it as focus, e.g. VP in (7a), and NP in (7b). In that case, shì was further grammaticalized into a focus marker in Early Medieval Chinese and has been in use from that time until now.
The hypothesis that the copula *shì* further grammaticalized into a focus marker in Early Medieval Chinese is debatable on several grounds. With respect to the argument from WH-words, in Archaic Chinese, WH-words occurred preceding regular transitive verbs. However, when they appeared with the word *wéi* ‘to be’, they systematically occurred following it, e.g.

(11) a. 子為誰
   zǐ wéi shéi
   2SG Wei who
   ‘you are who (who are you)?’
   Lunyu (論語479–400 BCE\(^{13}\))

b. 子之子為何
   zǐ zhi zǐ wéi hé
   2SG ASSOC son Wei what
   ‘your son is what?’
   Shiji (史記104–91 BCE)

When copula *shì* ‘to be’ arose, it took over the syntactic template and functions of *wéi* ‘to be’, and its occurrence preceding WH-words was predictable.\(^{14}\) Therefore there is no evidence for the hypothesis that the grammaticalization of *shì* led to a word order change or that it started to be used to mark the focus of WH-words. Secondly, Shi & Li do not explain why low transitivity is correlated in Chinese with OV order. Furthermore, they do not distinguish unmarked informational focus from marked contrastive focus in Chinese. Since their examples (e.g. (7a,b)) do not occur in contrastive contexts, there is no evidence that *shì* was used to mark contrastive focus in Early Medieval Chinese. The development of contrastive focus is connected with the rise of nominalization in Late Medieval Chinese, as discussed in the next sub-section.

### 3.2 Nominalization in Late Medieval Chinese

In this section we argue that crucial to the development of the cleft construction is the prior development of nominalizations with the form [XP + nominalization marker (NOM)], especially *zhē* and *dé*. The earliest examples of [XP *dé*] (known as the “*de* construction” among Chinese linguists) appeared in the Tang Dynasty (618–907 CE) with very low frequency. The most widely cited examples are in (12)


[14. Also in Early Medieval Chinese, in some instances of *shì* preceding a WH-word, *shì* was no longer a free morpheme, but an element coalesced with the WH-word, and therefore an integral part of the WH-word. For example: *rú wéi shí-shéi* 汝為是誰? ‘Who are you?’ *Xianyujing* (賢愚經 500CE). Here, *wéi* is the main verb following the subject and preceding the predicate that is the WH-word *shí-shéi* ‘who.’ Lien (2009) proposes a similar idea by looking at the early Southern Min dialect dated back to the 16th century.]

All rights reserved
and are among the earliest occurrences of nominalization [XP de] in the history of Chinese:

(12) a. 定知幃帽底，儀容似大哥
dìng zhī wéimào dǐ, yíróng sì dàgē
must know hat de, appearance like big.brother
‘(You) must know the hatted one; he looks like (your) big brother.’  
   Chaoye jianzai (朝野僉載700 CE)

b. 張底乃我輩一般人，此終是其坐處
zhāng dǐ nǎi wǒ bèi yībān rén, cǐ zhōng shì qí zuòchù
Zhang de PTCL we normal person, this eventually shì his sit.place
‘The one named Zhang is a normal person like us; this will eventually be his seat.’  
   Suitang jiahua (隋唐嘉話700 CE)

Ota (1958) suggests that wéimào dǐ ‘hat de’ in (12a) is an abbreviation of dài wé-imào dǐ ‘wear hat de (the one who’s hatted).’ Similarly, Zhāng dǐ ‘Zhang de’ in (12b) is a short form of xìng Zhāng dǐ ‘surname Zhang de (the one who’s named Zhang).’ Nevertheless, to date no examples of the corresponding long versions of similar examples have been found. Zhan’s data from the Classical Chinese Corpus confirm that there was no reduction and that the earliest nominalizations had the structure [NP de].

The particle de has several functions. It may be a nominalizer as in (12), but it may also be an associative (or genitive) as in (13a), or an attributive (or relative) as in (13b). How exactly the different functions of de emerged is not relevant to our discussion. It should, however, be mentioned that although there has been extensive debate among linguists about the formation of de, no consensus has been reached. Lü (1984) suggests that all the functions of dì developed from the Archaic Chinese nominalizer zhě. Wang (1958) argues, based on the phonological similarity, that the Archaic Chinese attributive particle zhî was the origin of dì, as in (13c):

(13) a. 水底浪
shuǐ dǐ làng
water ASSOC wave
‘the waves in the water’  
   Dunhuang bianwen (敦煌變文ca.900 CE)

15. The commonalities and distinctions between relativization and nominalization have been a hot topic among Chinese linguists, as relativization and nominalization are marked by the same linguistic element de, and both denote attributive meaning. In this paper, we treat [VP/S de head NP] as relativization, and [NP/VP/S de] as nominalization.

16. Other than the literature cited here, see Aldridge (2009), Yap, Chor & Cheung (2010) and Yap & Wang (2011) for a diachronic analysis of dì, de, and zhē in Chinese, respectively. We thank the anonymous reviewers for these references.
b. 修底行
   xiū  dǐ  xíng
   practice REL behavior
   ‘the behavior that one practices’

   Dunhuang bianwen (敦煌變文 ca.900 CE)

c. 罪我之由
   zuì  wǒ  zhī  yóu
   convict sg1 REL reason
   ‘the reason that you convict me’

   Zuozhuan (左傳 403–389 BCE)

Mei (1988) further develops Wang’s argument stating that the attributive dǐ was first derived from the source zhī, then the nominalizer dǐ emerged due to the influence of zhé, which generally appeared in phrase final position. Jiang (1999) proposes dǐ originally evolved from the localizer dǐ:

(14) 鍋底焦飯
   dāng  dǐ  jiāofàn
   pan  bottom burned.rice
   ‘the burned rice at the bottom of the pan’

   Shi shuo xin yu (世說新語 432–444 CE)

Later the localizer dǐ gradually assimilated the functions of zhē and zhī. Cao (1999) develops Jiang’s claim and argues that the three major functions of dǐ that appeared in Late Medieval Chinese have three different origins: the attributive dǐ developed from zhī; the associative dǐ came from the localizer dǐ; and the nominalizer dǐ had zhē as its source.17

Although different accounts are proposed for the distinct functions of de, most scholars agree that the nominalizer de is, in one way or another, related to the Archaic Chinese zhē. In Archaic Chinese, the micro-construction zhe was a nominalizer that normally occurred in phrase final position serving to nominalize an XP (cf. English ‘the one/thing that…’). The phrase [XP ZHE] appeared in a variety of positions18 including object position, as in (15):

17. This idea is developed in Section 5 below.

18. The use of the nominalizer zhē in Archaic Chinese is complex. It normally occurred in [XP ZHE], but it can be optional. For example, we can find both 上者 shàng zhē and 上 shàng in one text meaning ‘those that have higher position in the court.’ The nominalization in Archaic Chinese is outside the scope of this paper, but see e.g. Zhu (1983), Wang (1958), Cao (1999).
(15) 知之者不如好之者
zhī zhī zhě bù rú huà zhī zhě
know it ZHE NEG compare like it ZHE
‘Those who know it cannot compare to those who like it.’

Lunyu (論語479–400 BCE)

In Early Medieval Chinese, when shì had been frozen and constructionalized as the standard copula SHI, nominalizer ZHE occurred sporadically in the predicate position of a copula sentence, as in (16):

(16) 讓是殺我侍中者，不可宥！
Ràng shì shā wǒ shìzhōng zhě, bù kě yòu
Rang SHI kill my servant ZHE, NEG can forgive
‘Rang is the one who killed my servant; he cannot be forgiven!’

Shi shuo xin yu (世說新語432–444 CE)

Example (16) appears to be a specificational copula sentence with the structure [NP shi NOM]. The context of (16) is: Rang kills one of the king’s servants and another general. Some high-ranked official in the court wants to excuse him. The king becomes furious and says that Rang is the one who killed his servant, and he cannot be forgiven. From the context, the subject Rang is represented as uniquely the one who killed the speaker’s servant. The copula sentence here is background information to support the conclusion that Rang cannot be forgiven. Pragmatically, ‘Someone killed my servant’ is presupposed. Rang is the topic of the copula sentence uniquely selected from the set of those alleged to have killed the king’s servant. Rang is therefore a contrastive topic in (16). This suggests that the copula sentence with the form [NP shi XP ZHE] in Early Medieval Chinese could be used to denote a contrastive. It is not entirely clear at this point how contrastive focus that is associated with the Modern Chinese cleft construction was systematically signaled in Early Medieval Chinese. Examples with [NP shi XP ZHE] are not at all like the Modern Chinese cleft construction. The construct is specificational in that the post-copula [XP DE] conveys a restricted non-referential set (‘those who killed my servant’). Very few occurrences of [NP shi XP ZHE] are found in Early Medieval Chinese. Another example from the same text is (17), which is a specificational copula sentence with no explicit contrast.

(17) 太后入戶，見直侍並是昔日所愛幸者
tàihòu rù hù, jiàn zhíshì bìng shì xīrì
empress.dowager enter room, see servant all shi in.the.past
suǒ àixing zhē
them favor ZHE

‘The empress dowager entered the room and saw all the servants there were
the ones that she favored in the past.’

Shi shuo xin yu (世說新語432–444 CE)

From the time that ZHE was first attested (Archaic Chinese), it was predominantly
used as a nominalizer, as was the pattern [NP SHI XP ZHE]. But around 700 CE, de
came gradually to be used instead of ZHE in nominalizations. The use of ZHE de-
clined in Medieval Chinese; in contemporary Standard Chinese it is considered an
archaism. The earliest instances of nominalization [XP DE] appear to be [NP DE],
as exemplified in (12) above. They are found in the object position, as in (12a), or
in subject position, as in (12b), but no examples are found in post-copula position
at this period of time.

After speakers chose DE over ZHE as the nominalizer, the host-classes of [NP
DE] expanded to include the various categories available in XP, especially VP, S,
and NP. (18) is an example with VP:

(18) shī yuē: shuō-qū xíng bù dé dǐ,
Master say: say do NEG obtain DE,
行取說不得底
Xíng-qū shuō bù dé dǐ
do say NEG obtain DE
‘Master says: “say those(you) cannot obtain through doing, do those (you)
cannot obtain through saying.”’

Yunzhou dongshan wuben chanshi yulu (筠州洞山悟本禪師語錄850 CE)

In (18), both of the nominalizations xíng bù dé dǐ and shuō bù dé dǐ have the
serial verbs plus dǐ with the second (main) verb dé negated and the first verb in-
dicating instrument/method [V Vneg DE]. Xíng bù dé dǐ ‘those (you) cannot ob-
tain through doing’ is contrasted with shuō bù dé dǐ ‘those (you) cannot obtain
through saying.’ The nominalization, somewhat like English restrictive relative

19. Qu here is a particle attaching to the verb shuo and xing, which does not encode any sub-
stantial meaning. For example:

稻花香里说丰年，听取蛙声一片。
dào huā xiāng-lǐ shuō fēng nián, ting-qū wāshēng yī piàn
rice flower aroma-in say harvest.year, listen frog.sound one piece
Say the harvest year in the aroma of the paddy field, and listen to the sounds from the frogs.

XinQiji（辛弃疾 1140–1207）Xijiangyue（西江月）
clauses, provides the semantically restricted set that can be specified by a definite referential member (Patten 2012).

In sum, the nominalizer ZHE was used very frequently in Archaic Chinese, and is attested sporadically in the post-copula nominalized predicate by around 440 CE. As we show in the next sub-section, around 900 CE, the construction [XP DE] begins to be attested sporadically in the predicate position of copula sentences, a crucial step in the development of the cleft construction.

3.3 The emergence of the cleft construction

In what follows, we argue that the emergence of the cleft construction involved two steps: first, the expansion of the post-copula slot to include nominalizations in addition to NPs (Subsection 3.3.1), and second, the expansion of the semantics of the construction to contrastive in addition to specificational meaning (Subsection 3.3.2).

3.3.1 The emergence of [NP SHI XP DE]

As was shown in 3.2, the first occurrences of [XP DE] appear around 700 CE but with no instances in copula sentences. In the later part of the ninth century it occurs sporadically in post-copula predicate position, like [XP ZHE] (see example (16) above). One of the earliest examples of [NP SHI XP DE] is (19), found in a Tang Dynasty Buddhist text20 Zhengzhou linji huizhao chanshi yulu ‘The collective words from Master Linji and Master Huizhao from Zhenzhou’:

(19) 道流，是爾目前用底
Dàoliú shì ěr mùqián yòng dǐ
Daoism SHI SG2 currently practice DE
‘Daoism is the thing you are currently practicing.’
Zhengzhou linji huizhao chanshi yulu (鎮州臨濟慧照禪師語錄880 CE)

Here the nominalization consists of a clause ěr mùqián yòng ‘you are currently practicing’ plus the nominalizer dǐ [S DE] in the predicate position. Semantically, similar to (16), (19) is specificational in that the post-copula nominalization conveys a restricted non-referential set ěr mùqián yòng dǐ ‘the thing you are currently practicing’ and the definite referential subject dàoliú ‘Daoism’ specifies its referent. As in (16), pragmatically, since (19) is found in a Buddhist text, the topic of the

20. An anonymous reviewer points out that “Language and cultural contact often necessitates borrowing or using new grammatical forms, as in the case of the emergence of the [NP SHI XP DE] construction during the Tang period” (when Buddhist texts were massively introduced into the Chinese culture).
sentence dàoliú ‘Daoism’ is in contrast with the Buddhism that is advocated by the Masters. This is an example with contrastive topic just like (16). This shows again that the copula construction with the form [NP shì XP de] in 880 CE expressed contrastive meaning, but it had not yet developed into a cleft sentence.

Although the nominalization was recruited into the predicate position of the copula construction and the new structure [NP shì XP de] was used, the meaning has not changed; like (16), (19) is specificational. Although it has contrastive meaning, the focal contrastive meaning that is characteristic of cleft constructions is not found in (19). Since constructionalization requires a form_new-meaning_new pairing the emergence of (19) was simply a constructional form change; it represents a micro-step in the pre-constructionalization of the cleft construction. It is a critical context for the later development.

3.3.2 The emergence of the cleft construction

The first example known to us of a cleft in which a construct with post-copula [XP de] is both contrastive and specificational appears in (20). The structure of the post-copula nominalization in (20 III) is [S de] (where S= N V):

(20)  I. 莫將浮賄施為,
     mò jiāng fúhuì shǐwéi
     NEG take bribe behave

II. 非是菩薩行藏,
     fēi shì púsà xíngcáng
     NEG shì Buddha behavior

III. 此是俗門作底
     cǐ shì súmén zuò dǐ
     this shì layman do ōn
dunhuang bianwen (敦煌變文ca.900 CE)

‘Don’t execute the behavior of taking bribes; (it) is not Buddha’s behavior; it is laymen who do this.’

The copula sentence in (20 III) has the form [NP shì XP de] with the nominalization [S de] in the predicate position, and may have been developed on analogy with patterns like that in (16) with zhe (see further Section 5). The context of (20 III) is: Vimalakīrti tells one of his disciples not to execute the behavior of taking bribes, because that is not Buddha’s behavior. Semantically, the sentence is specificational as the nominalization [S de] ‘the things that laymen do’ indicates a non-referential set that is specified by the subject cǐ ‘this’ (the bribe behavior). In (20 I), what is different from the presupposition ‘some people execute the bribe behavior’ is asserted in (20 III) by the post-copula NP súmén ‘laymen,’ which is the focus contrasting with púsà ‘Buddha’ in (20 II) immediately preceding it. (20 III)
expresses the focal contrastive meaning, exclusiveness and exhaustiveness, as laymen are the only people who execute the bribe behavior in this context.

If we take a further look at the newly emerging cleft sentences exemplified by (20 III), we can see that the subject ci ‘this’ is semantically co-referential with the implicit object of the nominalization (as in the subject-object coreferential cleft). A later example involves subject-subject coreferentiality as in (21):

(21) I. 天下人總是學得底,
   tiānxiàrén zōng shì xué dé dǐ
di people.under.heaven always shi study obtain de
II. 某甲是悟得底
   mǒujiǎ shì wù dé dǐ
   sg1 shi enlighten obtain de
   ‘It is through study that people under heaven always obtain (the state of Chan); it is through enlightenment that I obtain (it).’

*Chanlin senbao zhuan* (禅林僧宝传 1100 CE)

Example (21) appears in *Chanlin sengbao zhuan* (1100 CE), a Chan Buddhism classic from the Northern Song dynasty (960 CE-1127). The context is about a little monk: One day when the little monk is meditating, he suddenly feels enlightened; he instantly gets up and goes to see the abbot; then he says to the abbot that it is through study that people always obtain the state of Chan, but it is through enlightenment that he himself obtains it. Both the sentences in (21) have the structure [NP shi XP de] with the nominalization [V V de], in which the first of the serial verbs indicates instrument/method. The subjects of the two sentences, i.e. tiānxiàrén ‘people under heaven’ in (21 I) and mǒujiǎ ‘I’ in (21 II), are coreferential with the implicit subjects of the two nominalizations. Both sentences in (21) express specificational meaning as the post-copula nominalizations indicate two restricted non-referential sets ‘those who obtain (Chan) through study’ and ‘those who obtain (Chan) through enlightenment’ and are specified by the referential subjects ‘people under heaven’ and ‘I’ respectively. Both sentences also express contrastive meaning. First the two topics ‘people under heaven’ and ‘I’ are in contrast. Second, the immediate post-copula expression xué ‘through study’ in (21 I) and wù ‘through enlightenment’ in (21 II) are the contrastive foci, marked by the copula shì.

Examples (20) and (21) appear to be innovations, one-offs that are precursors of later conventionalization as a construction. They are two of a total of three examples found so far of specificational contrastive copulas between 900 and 1270 (the third is in *Zhutang ji* from 952 CE). The text *Zhuzi yulei* (朱子語類 1270 CE) suggests that the specificational contrastive copula with nominalizer de had been constructionalized and conventionalized during the thirteenth century.
as there are more than 500 examples in this work. In some contexts the nominalizer 
DE is optional. For example, (22) consists of two cleft sentences with subject-
subject co-referentiality, and contrastive nominalized VPs *fā chūlái-le* ‘discharge 
come out-ASP’ in (22 I) and *shŏuliàn xiànglǐ* ‘retain toward inside’ in (22 II), but 
only the second of the VPs is marked by *dǐ*:

(22) I.  蓋仁是箇發出來了21，
gài rén shì gè fā chūlái-le, 
Alas benevolence SHI CL discharge come.out-ASP,
便硬而強 
bìan yìng ér qiáng 
then hard and strong 

II. 義便是收斂向裏底，
yì biàn shì shŏuliàn xiàng lǐ dǐ, 
Righteousness then SHI retain toward inside DE, 
外面見之便是柔 
wàimiàn jiàn zhī biàn shì róu 
outside see it then SHI soft 

‘Alas, benevolence is the thing that discharges and comes out of (the body), 
and thus it is hard and strong; whereas righteousness is the thing that retains 
toward inside (of the body), and thus is seen to be soft from outside.’

*Zhuzi yulei* (朱子語類1270)

We hypothesize that the copula *shi* was expanded to signal contrastive as well as 
specificational focus in sentences like (20)–(22) between 900 and 1270 CE. In ad-
dition, during 1100–1270 CE, marking of the nominalizer DE became optional 
provided that the subject was agentive and coreferential with the subject of NOM, 
as in (22 I). In this case there were no form-meaning changes involved, and 
accordingly no constructionalizations, only local constructional changes.

Example (22) coexists with (23) in *Zhuzi yulei*. In (23) the copula *shi* signals 
contrastive as well as specificational focus. It helps highlight the fact that the op-
tionality of DE is constrained by subject-subject coreferentiality, not, for example, 
by the aspect marker *le*. Both (22) and (23) have nominalized VPs, and the first, 
like (22 I) has the aspectual marker *le* with the implicit DE. However, in (23 II) the 
post-copula NOM [ADV VP DE] has the aspectual marker *le* with the explicit DE. 
Therefore, when there is subject coreferentiality, DE is optional:

21. Note that the nominal predicate in (22 I) is marked by classifier *ge*, while the nominal predi-
cate in (22 II) is marked by nominalizer *DE*. As one of our reviewers points out, “the language 
has more than one strategy to mark nominal expressions.”
4. Modeling the development of the cleft construction

To summarize the developments discussed in Section 3, the textual record shows the gradual development over time of copula clefts in addition to prototype copulas. The latter are typically of the form \([NP \text{ shi} \ NP]\), are specificational or predicational and cue information focus. Copula clefts, by contrast, have the form \([NP \text{ shi} \ \text{NOM}]\) and cue both specificational and contrastive meaning. In summary form the change can be represented as:

\[
\begin{align*}
\text{Copula Construction} & \quad \leftarrow \rightarrow \text{Cleft Construction} \\
[NP_i \text{ shi} \ NP_j] & \quad \leftarrow \rightarrow [\text{SEMi specificational SEM}_j]
\end{align*}
\]

This is a procedural constructionalization as it involves changes in cueing of information structure. Like other such constructionalizations it arose gradually micro-step by micro-step (see Traugott & Trousdale 2010, 2013). Specifically, the constructionalization of the cleft involves the following main steps (abstracting over minor distributional and token frequency changes):

a. Development of a copula construction with shi, and its use in preference to \(\text{we\text{"}}\): By Early Medieval Chinese, the standard copula already typically had the form \([NP \text{ shi} \ NP]\), and its semantics was equational, predicational, or specificational. Post-copula NPs in specificational copula sentences expressed non-referential but restricted set meaning.

b. Syntactic expansion: development in Early Medieval Chinese of nominalization involving use of \(\text{zhe}\) in the predicate position of a copula sentence.

c. Host-class expansion: recruitment in Late Medieval Chinese of \(\text{de}\) as a nominalizer, initially only in non-copula contexts, but later in post-copula contexts as well. The encoded meaning of \([\text{XP de}]\) expressed a non-referential but restricted set. As indicated in (a), this meaning was already a component of specificational copula sentences. However, in specificational copulas, this
non-referential but restricted set meaning was characteristic only of NPs in post-copula position, and it was not structurally marked by any particular morpheme. In contrastive contexts, [XP DE] came to be used occasionally in post-copula position.

d. Semantic-pragmatic expansion: Contrastive focus emerged in copulas with post-copula [XP DE] and became conventionalized independently of contrastive contexts.

e. Eventually the nominalizer ZHE obsolesced (category variant reduction) and [NP shi XP DE] became the default structure (the equivalent of “obligatorification” in grammaticalization).

These steps are summarized in (25):

\[
\begin{align*}
\text{(25) a.} & \quad \text{[NPi shi NPj]} \leftrightarrow \text{[SEMi specificational SEMj]} \rightarrow \\
\text{b.} & \quad \text{[NPi shi [XP ZHE]j]} \leftrightarrow \text{[SEMi specificational SEMj]} \rightarrow \\
\text{c.} & \quad \text{[NPi shi [XP DE]j]} \leftrightarrow \text{[SEMi specificational SEMj in contrastive contexts]} \rightarrow \\
\text{d.} & \quad \text{[NPi shi [XP DE]j]} \leftrightarrow \text{[SEMi specificational + contrastive SEMj]} \rightarrow \\
\text{e.} & \quad \text{[NPi shi [XP DE]j] (default)} \leftrightarrow \text{[SEMi specificational + contrastive SEMj]}
\end{align*}
\]

Each step of the process was a constructional change, as it involved either form change (b, e), discourse context (c), or meaning change (d) only. The result of the sequence of micro-steps is a form\textsubscript{new}-meaning\textsubscript{new} construction (constructionalization):

Another representation shows the sources of the cleft and the sequential order of the development of the construction in a way relatively familiar from the grammaticalization literature (except for the dual source):

![Figure 1. Main sources of the Cleft Construction](image)

Figure 1 captures the abstract sequential process of constructionalization of the cleft construction in Chinese. Like similar sequential arrays in work on grammaticalization, it gives no sense of the persistence of older constructions (most notably of the predicational and equational copulas), or of the extensive expansion...
involved in change. Figure 2 below is a constructional taxonomy that attempts to capture both of these factors including the several constructional changes discussed above.

![Constructional Taxonomy Diagram]

**Figure 2.** The development of the constructional schematic taxonomy of the prototypical Chinese copula construction

5. **Factors leading to and correlated with the rise of clefts**

We turn to consideration of the motivations and mechanisms involved in the constructionalization of the cleft construction \([NP_i \text{ shi} \ [XP \text{ de}]_j] \leftrightarrow [\text{SEMi specificational+contrastive } \text{SEM}_j]\), as represented in (25) above, specifically what are traditionally known as analogy and reanalysis. We also consider the roles of productivity, schematicity, and compositionality.

5.1 **Motivations: analogical thinking and discourse strategy**

We have argued that a key step in the development of the cleft in Chinese was the use of the nominalization construction \([XP \text{ de}]\) in post-copula position. We hypothesize that this change was enabled by analogical thinking motivated by two main factors. One is the semantic relatedness of expressions denoting a non-referential but restricted set. Recall that the encoded meaning of \([XP \text{ de}]\) expressed a non-referential but restricted set and that this meaning was already available in non-marked specificational post-copula NPs. However, in the earlier periods \([XP \text{ de}]\) did not appear in this position. This factor alone might have enabled the generalization of \([XP \text{ de}]\) to post-copula position. However, there was a second,
structural, factor that we hypothesize strengthened this analogical thinking: the use of the nominalizer zhe in post-copula position. zhe also denoted a non-referential but restricted set. At this stage the copula was specificational.

By the tenth century CE, speakers started using the string [NP shi XP de] sporadically in contrastive contexts to introduce contrasting discourse referents into the non-referential sets. Over time, contrastive meaning came to be associated with [NP shi XP de] independently of a concomitant contrastive clause, as in (26).

(26)  
I. 那馬在家歇他兩日  
那马在家歇他两日  
that horse at home rest a couple of days

II. 這馬是昨日東京翟雲峰親家送來的  
zhèmǎ shì zuórì dōngjīng zháiyúnfēng qīnjia sòng lái de  
this horse shi yesterday Dongjing ZhaiYunfeng in-law bring over de

III. 是西夏劉參將送他的  
shì xīxià liú cānjiàng sòng tā de  
SHI Xixia Liu general give him de

‘(Let) that horse rest at home for a couple of days. It was yesterday that ZhaiYunfeng’s in-law from Dongjing brought over the horse, and it was General Liu from Xixia who gave him (the horse)’

Jinpingmei (金瓶梅 1600 CE)

(26) is specificational with contrastive focus. In I, an indefinite short stay (“couple of days”) is suggested. In II, the focus is the specific day “yesterday” and the alternatives are other dates. In III, the focus is “General Liu from Xixia” in contrast to other possible people. Use of [NP shi XP de] independently of a concomitant contrastive clause is a shift from meaning associated with context to coded semantic meaning that is similar to the “context-absorption” which Kuteva (2001: 151) associates with the semanticization of inferential pragmatics in grammaticalization.

Therefore, the steering factors that motivated the emergence of the cleft are by hypothesis essentially speakers’ analogical thinking combined with communicative strategies in context.

---

22. The absence of [XP de] in post-copula position in texts like (16) and (17), but presence of [XP ZHE] in this position during the same period, suggests that DE replaced ZHE, and was not derived from it.
5.2 Mechanisms: analogization and neoanalysis

We propose that the analogical thinking that motivated the development of [NP *shi* *XP* *de*] enabled the mechanism of analogization of [XP *de*] to [XP *zhe*]. As we have seen, the latter was less constrained since it could be used in copula constructions around 440 CE, whereas the former could not when it first occurred in 700 CE. The hypothesis is that language users made a partial match between the two nominalizations, and the more versatile one, i.e. [XP *zhe*], was taken as the model. While the recruitment of the nominalization [XP *de*] is plausibly a case of analogization modeling the Early Medieval Chinese string [NP *shi* *XP* *zhe*], at the same time it is a syntactic reconfiguration or neoanalysis of the form ([NP *shi* NP]→[NP *shi* XP *de*]).

A few words may be in order about the relationship of analogization to neoanalysis. Mechanisms of change are hypotheses about how one mental representation of a given expression can give rise to a different one. Among mechanisms of change often cited are analogy (our “analogization”) and reanalysis (our “neoanalysis”) (Harris & Campbell 1995). While analogization involves pattern match and results in greater similarity, neoanalysis involves differentiation. In essence they are two sides of the same coin. When a pattern is used in a way more similar to that of another one (analogization) some of its former characteristics are changed or lost; this is neoanalysis. So all analogizations are neoanalyses (Kiparsky 2012 using the terms “analogy” and “reanalysis”, also Traugott & Trousdale 2013). That the hypothesized analogization of form of the copula [NP *shi* NP] to [NP *shi* XP *zhe*] is at the same time a neoanalysis is captured by the difference in form representation.

A later neoanalysis is the rise of optional use of *de* in contrastive clefts if the subject of the copula is semantically an agent and coreferential with the subject of NOM (e.g. (22) (23)). In this case the neoanalysis appears not to be analogical. We hypothesize that a motivation for this change may have been the use of the agent. This strengthened the event meaning of the nominalization, and by hypothesis allowed for loosening of the required presence of the nominalizer *de* (see Zhan & Sun 2013:781, who hypothesize that the presence of the nominalizer *de* in the clefts may be due to semantic and pragmatic factors).

---

23. We use the term “analogization” to contrast the mechanism of analogy from the motivation of analogical thinking (also traditionally referred to as “analogy”).

24. See also Simpson & Wu (2002) and Yap, Matthews & Horie (2004) for more discussion of the functions of the nominalizer *de*.
5.3 Productivity

Traugott & Trousdale (2013) suggest that the most relevant dimensions of constructionalization are productivity, schematicity, and compositionality. Here we discuss productivity and mention some implications for schematicity and compositionality.

Increase in productivity is related to frequency and to generalization of use and meaning. With respect to frequency, Bybee (2003) distinguishes token frequency from type frequency. Token frequency is a mechanism that enables and brings about change at the first place and is also the outcome of change, whereas type frequency is the key to entrenchment or storage, which helps the outcome of change to be frozen, fixed, and conventionalized in a community.

Bybee (2003:602) also argues that repetition and frequency of use is one of the mechanisms of change: “frequency is not just a result of grammaticalization, it is also a primary contributor to the process.” Traugott & Trousdale (2013) point out that the enabling effect of token frequency is debatable. Their evidence is that according to the historical texts they have examined, several grammatical changes started with very low frequency and sometimes continue to be used with low token frequency (see also Hoffmann 2005). Our observations of the emergence of the Chinese cleft construction are consistent with the observation that change does not require prior token frequency, as illustrated in Table 1:

<table>
<thead>
<tr>
<th></th>
<th>Dunhuang bianwen (ca. 900 CE)</th>
<th>Zhutang ji (952 CE)</th>
<th>Zhuzi yulei (1270 CE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Token dǐ</td>
<td>12</td>
<td>230</td>
<td>4560</td>
</tr>
<tr>
<td>[NP shi XP de]</td>
<td>3</td>
<td>1</td>
<td>606</td>
</tr>
</tbody>
</table>

There are 12 tokens of the particle dǐ in Dunhuang bianwen (ca. 900 CE), but only 3 tokens of the nominalization [XP de] (Cao 1995, Wu 1997), and only one of [XP de] in post-copula position (our example (20)). In Zhutang ji (952 CE), there are 230 tokens of the particle dǐ of which only 26 (11.3%) are instances of the nominalization (Feng 2000:428), but only one occurs in post-copula position. After the cleft construction emerged, it continued to occur with low frequency even in Southern Song (1127–1279). In Zhuzi yulei (1270 CE), a text written after the constructionalization of the cleft, the token number of [NP shi XP de] is still fairly low — there are 606 examples of [NP shi XP de] (13.3% out of 4,560 tokens of particle de) (Zhu 1991). From Zhuzi yulei (1270) on, as [NP shi XP de] became more and more frequent, the cleft construction was entrenched, integrated, and spread through the language system and was conventionalized in the language.
community. It appears that the increased token frequency of use of [XP de] in post-copula position is correlated with or is the outcome of constructionalization, but there is no evidence that it was an enabling factor.

Although there is little evidence of token frequency increase prior to the constructionalization of the cleft, there is evidence of type frequency increase. Increase in type productivity (host-class expansion) is exemplified by the expansion of the categories that became available as foci (host-class expansion). For example, in (20)–(23) we find the post-copula position occupied by an NP, V, VP, or ADV, + nominalizer. The increased productivity in available foci contributed to micro-step type frequency changes at the micro-constructional level. As clefts with coreference between the subject of the copula and the subject of the nominalization (which may be implicit) began to be used, the nominal entity meaning of the nominalization became more bleached and generalized, and the verb of the nominalization started to involve temporal and aspectual events and situations. As we saw in (22) and (23), both of the VPs in the nominalizations fā chūlái-le ‘discharge come out-ASP’ and yǐ sàn-le ‘already scatter-ASP’ were cliticized by the aspectual marker le. Accordingly, the nominalizations fā chūlái-le (dī) ‘the thing that has come out’ and yǐ sàn-le dī ‘the thing that has scattered’ expressed situations involving aspectuality.

Additional evidence for increase in type frequency is provided by the appearance in the seventeenth century of examples like (26), with temporal and spatial nominal in NOM. In the earliest examples the nominal in NOM denotes humans, but by the seventeenth century we begin to find use of inanimate, especially time and place nominals as focus (e.g. example (26)). The appearance of temporal and spatial nominal in NOM is presumably an example of analogization. It is also an example of generalization.

Increased productivity led to increased schematicity. With the development of new micro-constructions, the schema came to have new construction-types, and the rising cleft construction became a subschema in the constructional schematic taxonomy of the prototypical Chinese copula construction. Increase in schematicity is associated with increase in abstractness of meaning. 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.

The contrastive focus indicated by the immediate post-copula element within the nominalization and the presupposition cued by the rest of the copula sentence gave rise to the decrease of the compositionality of the copula construction. Furthermore, its compositionality also decreased as the nominal entity meaning of the nominalization became bleached and generalized, leading to the optionality of the nominalizer.
6. Conclusion

This paper has addressed the constructionalization processes of the cleft construction in the history of Chinese. We argued that the constructionalization of the cleft involved the recruitment of a nominalization pattern to the predicate position of the prototype copula construction through analogization, which gives rise initially to pragmatic modulation and later to semantic and syntactic neoanalysis. This process involves the development of complex clause syntax. Analogical thinking and use in discourse contexts were the enabling factors. Analogization and neoanalysis were the major mechanisms for the change.

The study is a contribution to the developing field of constructionalization in that it has made more explicit the way in which individual constructional changes contribute sequentially to procedural constructionalization. It has also shown how a complex contrastive cleft construction may come into being. Lehmann (2008: 211) has suggested that cross-linguistically a cleft structure is “[t]he most explicit syntactic strategy of contrastive focus.” According to Lehmann, a cleft construction is an ideal syntactic information structuring strategy because:

[1]he separation of focus and presupposition by a two-clause structure … reflects the attention cline between focus and presupposition by the asymmetric syntactic status of main clause and dependent clause. Finally, it puts the focus expression, and nothing else, into the predicate of the main clause, thus assigning the focus the canonical syntactic function for new information. (Lehmann 2008: 212)

He goes on to show that in some languages this two-clause structure is the source of various, often monoclausal expressions. The developments outlined in this paper do not conform to Lehmann’s prototype for contrastive clefs. For one, not all contrastive copulas with post-copula [XP de] in Chinese are biclausal, although some have VP or S in XP. Furthermore, the only kind of reduction that figures in the development of the Chinese cleft is the development of optionality in the use of de when the subject is agentive and coreferential with the subject of the nominalization.

Lehmann does not discuss how clefting structures come into being as he is primarily interested in showing how “[a]s always in grammaticalization, the degree of complexity shrinks from the text level via the sentence and clause levels down to the phrase level” (Lehmann 2008: 227), but he does acknowledge that there is probably a pre-cleft stage since he refers to “a stage not too long past their genesis” (Lehmann 2008: 227). The constructionalization process of the Chinese cleft that we have discussed suggests one way in which the separation of focus and presupposition by a two-clause structure can come into being. It reflects the expansion
of the copula construction over time by recruitment of a nominalization to the predicate position, which may or may not result in a complex clause structure.\textsuperscript{25}

**Abbreviations**

The following abbreviations are used: ASP = Aspectual marker; ASSOC = Associative; CL = Classifier; COP = Copula; FM = Focus marker; NEG = Negative; NOM = Nominalization; NP = Noun phrase; PTCL = Particle; RC = Relative clause; REL = Relativizer; S = Clause; SEM = Semantics; 1SG = First person singular; 2SG = Second person singular; 3SG = Third person singular; VP = Verbal phrase

**Data source**


**References**


\textsuperscript{25}. For more nominalization studies, see Yap & Matthews (2008), Yap, Choi & Cheung (2010), and Yap & Grunow-Hårsta (2010) for details.

All rights reserved


Harris, Alice & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511620553


Hilpert, Martin. 2013. *Constructional change in English: Developments in allomorphy, word-formation and syntax*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139004206


All rights reserved
The constructionalization of the Chinese cleft construction


Author’s addresses

Ms. Fangqiong Zhan  
East China Normal University  
International College of Chinese Studies  
109 Office Building  
Shanghai 200062  
China  
fqzhan@hanyu.ecnu.edu.cn

Elizabeth Closs Traugott  
Department of Linguistics  
Margaret Jacks Hall, Building 460  
Stanford University, Room 127  
Stanford, CA 94305–2150  
USA  
traugott@stanford.edu