



Universiteit  
Leiden  
The Netherlands

## Shi as a complementizer

Yang, Z.

### Citation

Yang, Z. (2024). Shi as a complementizer. *Journal Of Chinese Linguistics*, 52(2), 362-384.  
doi:10.1353/jcl.2024.a929993

Version: Accepted Manuscript

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/3714639>

**Note:** To cite this publication please use the final published version (if applicable).

*Shì* as a complementizer (作为小句标句词的“是”)

Zhaole YANG

Leiden University

Affiliations: Department of Chinese Studies, Leiden Institute for Area Studies (LIAS),

Leiden University

Email: z.yang@hum.leidenuniv.nl

**This version differs somewhat from the one set to be published in the *Journal of Chinese Linguistics*.**

**Acknowledgments:**

The idea for this paper stemmed from a conversation with Rint Sybesma during our co-teaching of a Mandarin typology course in Leiden. I am grateful for his numerous helpful comments and suggestions.

## *Shì* as a complementizer

### Abstract

In this article, it is argued, that a purely functional and semantically bleached *shì* in Mandarin could in many contexts be best analyzed as a complementizer. As such, it is comparable to English *that* and functions as an (optional) introducer of the clausal arguments. Our survey shows that the morpheme *shì* as a complementizer occurs after quite a number of clause introducing operators, for instance, epistemic modals, clausal adverbs, clausal conjunctions, and verbs taking clausal objects like psych verbs and speech-act verbs. Like in some other languages, e.g., the West African language Twi and Dutch, the complementizer can merge with a preceding element and the two elements are turned into one frozen combination involving reanalysis and constituency rebracketing. We argue that Mandarin *shì* has gone through similar developments as complementizer in many frozen combinations such as *shuōshì*. As a piece of additional evidence, we find that in Gangou dialect, a Mandarin variant, *shì* can also take on a complementizer function. In the end, based on the fact that *shì* is used as a demonstrative pronoun in Pre-Qin times, we argue that Mandarin indeed has complementizers from two different sources out of the five major sources across languages listed by Chappell (2017), i.e., *shuō*, the complementizer (as argued by many) originally from a verb of saying, and *shì*, originally a demonstrative.

**Key words:** Complementizer, Mandarin, Grammaticalization, Typology of Sinitic languages

### 摘要

本文提出在汉语普通话中,失去其基本语义而仅具有语法功能的“是”在多种句子类型中均可被分析为小句标句词 (complementizer),用以标记和引导小句论元。通过考察我们发现标句词“是”广泛出现在各种可接引句子的成分之后,比如,知识情态动词,小句层面副词,连结句子的连词以及一些带小句宾语的动词,比如心理类动词和言说类动词。正如西非语言 Twi 和荷兰语,汉语普通话标句词“是”可以通过进一步语法化而失去其独立性,并和其前置成分融合成一种“凝固组合”,比如“说是”。作为旁证,我们也可以在一些汉语方言比如甘沟话中,发现作为标句词使用的“是”。基于“是”在先秦汉语中作为指示词使用,加之“说”在之前文献中被分析为汉语标句词,我们得出结论:汉语小句标句词的来源占 Chappell (2017)归纳的五大类来源中的两类,即言说向标句词“说”和指示向标句词“是”。

**关键词:** 标句词 汉语普通话 语法化 汉语类型学

## 1. Introduction

It has been observed that Mandarin *shì*, which usually functions as a copular verb, can follow modal verbs such as *yīnggāi* ‘should’, *bìxū* ‘must’, and *kěnéng* ‘might’ (Dong 2004, Park 2011, Fang 2018), as demonstrated below:

- (1) *Tā yīnggāi (shì) huì ànshí jiāo zuòyè*  
s3 should SHI will on.time submit homework  
‘It should be the case that he will submit homework on time.’
- (2) *Nǐ bìxū (shì) yí-ge-rén lái, fǒuzé wǒ bú jiàn nǐ.*  
s2 must SHI one.person come otherwise s1 not see you  
‘It must be the case that you come by yourself, otherwise I will not meet you.’
- (3) *Míngtiān kěnéng (shì) huì xiàyǔ.*  
tomorrow might SHI will rain  
‘It might be the case that it will rain tomorrow.’

In these sentences, *shì* displays different properties from its typical copular function (cf. Park 2011: 474). For one thing, its use in these sentences is optional: leaving it out will not have any syntactic or semantic consequences.<sup>1</sup> Furthermore, it cannot take any prosodic stress and nothing can be inserted in between the modal verb and *shì*. Importantly, the constituents introduced by the combination of a modal plus *shì* are generally interpreted with a clausal reading, as reflected in the translations. All these properties suggest that *shì* in (1)-(3) does not function as a copular verb. Rather it is a purely functional (or “grammatical”) morpheme, an element without a lexical meaning.

---

<sup>1</sup> As a copular verb, *shì* is only optional in a very limited number of cases; for an overview, see Cheng (2021).

However, it seems that *shì* cannot always be used after these modals. Consider the following sentences:

(4) *Tā (zuòwéi xuésheng) yīnggāi (\*shì) ànshí jiāo zuòyè.*  
s3 as student should SHI on.time submit homework  
'(As a student), he should submit homework on time.'

(5) *Nǐ bìxū (\*shì) lái!*  
you must SHI come  
'You must come!'

(6) *Míngtiān kěnéng (\*shì) xiàyǔ.*  
tomorrow might SHI rain  
'It might rain tomorrow.'

Park (2011: 474-475) has accurately captured the semantic differences between the phonologically identical modal verbs in the two groups, for instance, *kěnéng* in (3) is used to express the speaker's evaluation or judgement of the possibility of the whole proposition, while in (6), it denotes the "objective possibility of the event or situation"; *Yīnggāi* or *bìxū* in (1) and (2) denote the speaker's speculation about the uttered proposition, while their counterparts in (4) and (5) are used to express the duty or obligation of the subject. In sum, the semantics of the modal verbs in the first three sentences are all speaker-oriented and their counterparts in the latter three are all subject-oriented. However, the differences between the modals in the two groups are not merely semantic. Conventionally, the modals like those in (1)-(3) with a clausal reading are called "epistemic modals" and the modals which relate the subject to the predicate, like those in (4)-(6), are called "root" or "deontic" modals (e.g., Cook 1978; Lin and Tang 1995; Brennan 1997, Tsai 2015). The epistemic/root distinction has a syntactic significance as well.

It has often been argued (e.g., Jackendoff 1972: 89; Cinque 1999: 11) that speaker-oriented adverbs are syntactically higher than subject-oriented adverbs. In the same spirit, a similar hierarchy between modal verbs has been proposed. For instance, Brennan (1997: 192) and Butler (2003) argue that the epistemic modals should be taken as sentential operators which take scope over the subject, and are, thus, structurally higher than [Spec, IP], while the root/deontic ones should be considered as predicate operators and occupy positions higher than vP but lower than IP. In line with Lin and Tang (1995) and Lin (2012) and based on Butler's (2003) modal hierarchy, Yang (2020: 87) proposes a hierarchical structure of Mandarin modals and argues that the modals expressing epistemic necessity like *yīnggāi* and epistemic possibility like *kěnéng* should be located in the CP layer (above the strong/outer subject, respectively in ForceP and FinP) and that root or deontic modals, like those in (4) to (6), occupy a position lower than the subject and above vP. On this basis, we can say that the modals in (1)-(3), with the epistemic reading, should be treated as clausal/higher modals which scope over the whole clause, while those in (4)-(6) should be treated as root/lower modals with scope over the predicate only. The compatibility tests regarding *shì* in these sentences then yield the following result: The *shì* under discussion can follow modals with a clausal/epistemic reading which occupy a position in the CP domain and cannot follow non-clausal root/deontic modals occupying a lower position. As noted above, it does not seem to be the case that *shì* in these sentences behaves like a copular verb. In this article, I argue that this *shì* is a complementizer, which is used to introduce different types of dependent clauses. In the following sections, I will show that the

complementizer *shì* is quite pervasive in Mandarin as it is found following many clausal operators, quite parallel to the complementizers we observe in other languages.

## 2. The distribution of the complementizer *shì*

In general, *shì* as a complementizer can occur immediately after many clause introducing elements, like epistemic modals, clausal adverbs, clausal conjunctions, and verbs taking clausal objects.

### 2.1 *Shì* following clausal modals

As shown in (1)-(6), the semantically bleached *shì* can occur with speaker-oriented modals with a clausal/high interpretation and cannot follow the root modals with a predicative/low interpretation. This generalization is confirmed by (7) and (8):

- (7) a. *Děi (shì) tā lái, (cái néng jiějué wèntí).*  
 have.to SHI s3 come then be.able solve problem  
 ‘It has to be the case that he comes (then the problem can be solved).’

- b. *Tā děi (\*shì) lái.*  
 s3 have.to SHI come  
 ‘He has to come.’

- (8) a. *Huì-bu-huì (shì) tā shēng-bìng le?*  
 will-not-will SHI s3 get.sick PRF  
 ‘Is it possibly be the case that he has got sick.?’

- b. *Tā huì (\*shì) shēng-bìng.*  
 s3 will SHI get.sick  
 ‘He will get sick.’

As shown in (7a) and (8a), the epistemic necessity modal *děi* and the epistemic possibility modal *huì* can be followed by *shì*. Both the combinations of modal plus *shì*

introduce a clause. In contrast, *shì* cannot be inserted after their root counterparts, as can be seen in (7b) and (8b). Note that the combination of epistemic modal plus *shì* in (7)-(8) occurs sentence-initially, before the subject. Note further that all combinations of epistemic modals plus *shì* in (1)-(3) can be placed before the subject as well. This observation lends support to the hypothesis that *shì* could very well be analyzed as a complementizer, comparable to English *that*, which functions as an (optional) introducer of the clausal arguments of the preceding speaker-oriented modal verbs. The variation in word order (the subject follows or precedes the epistemic verb with *shì*) can be explained, if we assume that epistemic modal verbs are base-generated in the sentence-initial position and that the subject of the complement clause optionally moves up. Note that the root modals in Mandarin, which are usually believed to select a vP (or VP), can never occur as high as the epistemic modals and never occur in a sentence-initial position, preceding the subject if there is one. An additional assumption would be that in sentences with a clausal modal, the complementizer *shì* is always there after the modal, overtly or covertly.

## 2.2 *Shì* following clausal adverbials

It was already documented by Chao (1968: 77) that a *shì* can follow some adverbs expressing the speaker's evaluation or attitude. One of his examples is reproduced below:

- (9) *Wūzi-lǐ zhème lěng! Guǎnbǎo shì lúzi miè le.*  
 in.the.room so cold Certainly SHI stove go.out PRF  
 'It is so cold in the room! Certainly, the stove has gone out'

The combination of the epistemic adverb *guǎnbǎo* ‘certainly’ and the desemanticized *shì* is used at the beginning of the sentence and expresses the speaker’s high commitment to the contents of the proposition. Here too, *shì* seems to behave like a complementizer. Like the cases with sentential modals, *shì* is optional in (9) and the combination of the epistemic adverb plus *shì* can alternatively occur after the subject. It has been argued (e.g., Jackendoff 1972:89, Cinque 1999: 11, Ernst 2004: 10) that speaker-oriented adverbs, like evaluative adverbs, epistemic adverbs, and evidential adverbs, occupy a high position in the structure and should be treated as clausal/CP adverbs in the sense that they modify the whole sentence. If *shì* under discussion can occur after different types of clausal adverbs and before the embedded clause, we have all the more reason to argue that *shì* is used as a clausal complementizer. A survey can be seen below:

(10) Evidential adverbs plus *shì*

*Xiǎnrán (shì) /hǎoxiàng (shì) yǒu-rén shòushāng le.*  
 obviously SHI seemingly SHI someone get.injured PRF  
 ‘Obviously/seemingly someone has got injured.’

(11) Evaluative adverbs plus *shì*

*Xìngkuī (shì) tā gǎnshàng le mò-bānchē.*  
 luckily SHI s3 catch PRF last.bus  
 ‘Luckily he managed to get on the last bus.’

(12) Epistemic adverbs plus *shì*

*Dàgài (shì) /yīdìng (shì) /yěxǔ (shì) tā shēng-bìng le.*  
 probably SHI surely SHI perhaps SHI s3 get.sick PRF  
 ‘Probably/Surely/Perhaps he has got sick.’

As these examples show, all three types of speaker-oriented adverbs can be followed by *shì* and occur sentence-initially. As before, the combination of the clausal adverbial plus *shì* can alternatively occur after the subject as well and *shì* can be omitted without causing semantic loss or ungrammaticality. The general acceptance of *shì* after the clausal/high adverbs forms a contrast to the fact that *shì* cannot be used after low VP adverbs, such as manner adverbs, as shown in (13):

- (13) *Tā màn-mǎr (\*shì) zǒujìn le jiàoshì.*  
 s3 slowly SHI walk.into PRF classroom  
 ‘He walked into the classroom slowly.’

The contrast can be explained if we follow what is generally assumed, namely that speaker-oriented adverbs, just like the epistemic modals discussed in the previous section, are base-generated in a position higher than the subject and located in a specifier position of some (covert) clausal functional heads (in line with Cinque’s assumption), which selects a clausal argument introduced by the complementizer *shì*. The complementizer can optionally occur in the surface structure. Low adverbs do not select for a full-fledged sentence/clause.

### 2.3 *Shì* following clausal conjunctions

The semantically bleached *shì* can also occur after most clausal conjunctions in Mandarin and introduce different types of subordinate clauses. Such cases have been discussed in Dong (2004). According to Dong (2004: 35), *shì* in these sentences is always unstressed and there is no pause between the conjunction and *shì*. Crucially, *shì* in these sentences is not used to assert or emphasize the proposition. Therefore, *shì* in

these sentences does not function as a copular verb or focus particle, as it can do in other contexts. As before, the combination of conjunction plus *shì* follows or precedes the subject of the clause introduced by it.

### 2.3.1 *Shì* used after adverbial conjunctions

*Shì* can be optionally used after conditional conjunctions, such as *zhǐyào* ‘as long as’, *zhǐyǒu* ‘only if’, *chúfēi* ‘unless’, *bùguǎn/wúlùn* ‘no matter’, or *rúguǒ* ‘if’:

- (14) *Wúlùn /bùguǎn (shì) háizi xiǎng-yào shénme, tā dōu mǎnzú.*  
 no.matter SHI kids want what s3 all satisfy  
 ‘He will satisfy anything that the kids want.’

- (15) *Zhǐyào /rúguǒ (shì) tā yǒu shíjiān, tā jiù xué Zhōngwén.*  
 as.long.as if SHI s3 have time s3 then learn Chinese  
 ‘As long as/if he has time, he will learn Chinese.’

- (16) *Chúfēi /zhǐyǒu (shì) nǐ dé le xīngguān, nǐ cái kěyǐ bù lái.*  
 only.if SHI s2 get PRF covid s2 then can not come  
 ‘You don’t have to come only if you get covid.’

*Shì* can also be optionally used after concessive or concessive conditional conjunctions, such as *suīrán/jǐnguǎn* ‘although’, *nǎpà /jǐbiàn* ‘even if’:

- (17) *Nǎpà /jǐbiàn (shì) tiānqì zài-chà, tā dōu chūqù pǎobù.*  
 Even.if SHI weather worse s3 all go.out run  
 ‘Even if the weather is worse, he goes out for running.’

- (18) *Suīrán /jǐnguǎn (shì) tā qù guo nàli, tā hái xiǎng zài qù.*  
 although SHI s3 go EXP there s3 still want again go  
 ‘Although he has been there once, he would like to go there for another time.’

*Shì* can also occur after conjunctions expressing reason or cause, as shown below:

- (19) *Jìrán (shì) nǐ cuò le, nǐ wèishénme bù dàoqiàn?*  
 Since SHI s2 wrong PRF s2 why not apologize

‘Since it is your fault, why don’t you apologize?’

- (20) *Yīnwèi (shì) tiānqì bù hǎo, suǒyǐ wǒmen méiyǒu chūqù.*  
 because SHI weather not good so we not go.out  
 ‘Because the weather was not good, we did not go outside.’

### 2.3.2 *Shì* used after coordinating or correlative conjunctions

The grammatical morpheme *shì* can optionally occur with coordinating or correlative conjunctions signaling addition, contrast, sequence, or alternation. The combination of the clause connector and *shì* has a strong tendency to occur before the subject of the introduced clause.

- (21) *Yàome (shì) /huòzhě (shì) Zhāng Sān qù,*  
 or SHI or SHI Zhang San go  
*yàome (shì) /huòzhě (shì) Lǐ Sì qù.*  
 or SHI or SHI Li Si go  
 ‘Either Zhang San will go, or Li Si will go.’

- (22) *Wǒ xǐhuān zhè-ge chéngshì, dì-yī (shì) zhèli huánjìng*  
 s1 like this-CL city firstly SHI here environment  
*hǎo, dì-èr (shì) zhèli-de rén hěn yǒuhǎo.*  
 good secondly SHI here-ATTR people very friendly  
 ‘I like this city. Firstly, environment here is good; secondly, people here are very friendly.’

- (23) *Yǔqí-shuō (shì) wǒ bāng le nǐ,*  
 rather.than SHI s1 help PRF s2  
*dǎo-bùrú-shuō (shì) nǐ bāng le wǒ.*  
 better.say SHI s2 help PRF s1  
 ‘We better say that you have helped me, rather than that I have helped you.’

- (24) *Tā zhēn dǎoméi, xiān (shì) tā zǎoshang diū le*  
 s3 really unlucky first SHI s3 morning lose PRF  
*qiánbāo, ránhòu (shì) yàoshi yòu diū le.*  
 wallet then SHI keys again lose PRF  
 ‘He is so unlucky. He first lost his wallet in the morning, then he lost his keys later.’

## 2.4 *Shì* used after psych verbs or speech-act verbs

Compared to the extensive use of *shì* as a complementizer after epistemic modal verbs, the use of *shì* as a complementizer after other types of verbs introducing a clausal argument is relatively limited. But we do find a few cases as well.

- (25) *Tā shuō /tīngshuō /zhèngshí (shì) Zhāng Sān tōu le*  
s3 say hear confirm SHI Zhang San steal PRF  
*yáng.*  
sheep  
'He said/heard/confirmed that Zhang San stole his sheep.'

- (26) *Wǒ xiǎng /zhīdào (shì) wǒ bú gòu-hǎo,*  
s1 think know SHI s1 not good.enough  
*dá-bu-dào nǐ-de yāoqiú.*  
be.able.to.reach your requirement  
'I think/know that I am not good enough to meet your requirements.'

*Shì*, introducing a clausal complement, is used in (25) after the speech-act verbs *shuō* 'say', *tīngshuō* 'hear', and *zhèngshí* 'confirm', and in (26) used after the psych verbs *xiǎng* 'think' and *zhīdao* 'know'. Note that, as above, *shì* in (25) and (26) has lost its verbal meaning and does not take any stress.<sup>2</sup> Its function is only to introduce the clausal argument. So, it serves as a complementizer in the above sentences.

## 2.5. Summary and discussion

---

<sup>2</sup> Besides the reading provided here, note that there can be a focal reading of *shì* as well for (25) and (26), where the subject is focused. Then *shì* and its adjacent focus should be bracketed as one constituent. This is a different *shì* from the one discussed here. See section 2.5 immediately below.

In this section, I have examined four contexts in which we find the morpheme *shì*, where it can be analyzed as a primarily functional element, more in particular, a complementizer, introducing a clause, as instantiated by the following representations:

Type 1:  $s[\text{ModP } shì \text{ S}]$  (Section 2.1)

Type 2:  $s[\text{AdvP } shì \text{ S}]$  (Section 2.2)

Type 3:  $s_{ADV}[\text{Conj. } shì \text{ S}] \text{ S}$  or

$s[\text{Conj. } shì \text{ S}], s[\text{Conj. } shì \text{ S}]$  (Section 2.3)

Type 4:  $s[\text{NP V } s[ shì \text{ S} ]]$  (Section 2.4)

As emphasized above, these are the contexts in which we typically find a complementizer in other languages.

One of the reviewers remarks that *shì* in some of the sentences presented above could also be analyzed as a focus particle or could at least be ambiguous between the complementizer reading and the focus reading. We agree with this comment. As pointed out in footnote 2, when *shì* is used after certain psych verbs, the focus reading is also possible. However, the focus reading of *shì* has prosodic consequences; the (constituent) structure of a sentence with focus marker *shì* is also different. If *shì* is analyzed as a focus particle in (25), then we would have the bracketing as presented in (27):

(27) [<sub>NP</sub>[*Tā*] <sub>v</sub>[*shuō* /*tīngshuō* /*zhèngshì*] <sub>S</sub>[<sub>FocP</sub>[(*shì*) <sub>NP</sub>[*Zhāng Sān*]]

s3 say hear confirm SHI Zhang San

*tōu-le yáng*]]

steal-PRF sheep

‘He said/heard/confirmed that it is Zhang San who stole his sheep.’

In (27), the subject in the subordinate clause is part of the constituent FocP, headed by the focus particle *shì*. Prosodically, the focused constituent is stressed. This is different from (25), in which *shì* is an independent morpheme introducing the whole sentence and the subject in the subordinate clause need not to be stressed and does not form a constituent with *shì* (with consequences for movement away from it; see below). In other words, (25) is structurally ambiguous, between a reading in which *shì* is a focus particle and one in which it is not (in which case we propose that it is a complementizer). The same applies to a number of other sentences presented above, such as (11) and (12), where *shì* is used after clausal adverbials, and (21), where *shì* is used in the *yàome...yàome...* construction.<sup>3</sup>

<sup>3</sup> The reviewer also wonders whether it is possible to treat *shì* used after the high adverbials as in (10)-(12) as a focus particle with a broad focus scope, i.e., scoping over the whole clause. We find that this reading can only be achieved by providing a very specific context in which there is a clear clausal contrast, e.g., by adding a follow-up sentence introduced by *búshì*, see (i) based on (12) in the text:

(i) *Dàgài shì tā shēng-bìng le, bú-shì huǒchē chū-le wèntí.*  
 probably SHI s3 get.sick PRF not train happen- PRF problem  
 ‘Probably (because) it is that he got sick, not that the train had a problem.’

Note that *shì* in this contrastive sentence *shì...bú-shì...* is in general not optional, which is also different from the optionality of the complementizer *shì*. Therefore, we can say, despite the existence of ambiguity in some cases, it is not difficult to tell the difference between the two usages.

Importantly, the fact that another analysis is possible, does not exclude the possibility that our proposal that *shì* in these sentences functions as a complementizer is not correct. In fact, there are sentences with *shì*, in which it behaves as in the sentences under consideration in this paper, but for which, crucially, a focus analysis is not possible, like in sentences in which the nominal following *shì* (in our analysis, the subject of the subordinate clause) is an indefinite NP of the type that is hard to bear the focus. This is exemplified in (10) and (17), repeated here as (28) and (29):

(28) *Xiǎnrán (shì) /hǎoxiàng (shì) yǒu-rén shòushāng le.*  
 obviously SHI seemingly SHI someone get.injured PRF  
 ‘Obviously/seemingly someone has got injured.’

(29) *Nǎpà /jìbiàn (shì) tiānqì zài-chà, tā dōu chūqù pǎobù.*  
 Even.if SHI weather worse s3 all go.out run  
 ‘Even if the weather is worse, he goes out for running.’

It is basically impossible to have a narrow focus on the indefinite subject immediately following *shì* in the above sentences, i.e., *yǒu-rén* ‘someone’ and *tiānqì* ‘the weather’. Meanwhile, the fact that the subject NPs following *shì* in these sentences can be moved to the left of the combination of the clausal adverbial/conjunction plus *shì* in (28) and (29) makes the focus analysis even less likely, as the focus constituent of *shì* is in general right-adjacent to the particle as we can observe in (26). Therefore, the focus account of *shì* in these sentences can be ruled out and the complementizer approach seems to be the only possibility.

In sum, although the functional *shì* is under certain circumstances ambiguous between a focus reading and a complementizer reading, it is not difficult to identify the differences with the aid of the prosodic features or contextual elements. The

unavailability of the focus analysis of *shì* in some other sentences reaffirms our treatment of *shì* as a complementizer in these sentences.

### 3. “Frozen” combinations

It has been observed that in some languages the complementizer can sometimes merge with the element preceding it, such as adverbs or conjunctions, and form a complex element, which often involves a change in constituency bracketing. Fused with the preceding element, the formerly independent morpheme serving as the complementizer is not a separate word anymore. The two elements are turned into one “frozen” combination and is “reanalyzed” (cf. Hopper and Traugott 2003: 51). Examples from the Western African language Twi were documented by Lord (1993: 169-171), one of which is reproduced here as (30):

- (30) *Oguanee (efi)sɛ osuro.*  
'He ran away because he was afraid.'

Gloss 1:

<i>s</i> [ <i>Oguanee</i> ]	<i>s</i> [ <i>e-fi</i>	[ <i>sɛ osuro</i> ]].
He-ran way	it-come-from	that he-was-afraid

Gloss 2:

<i>s</i> [ <i>Oguanee</i> ]	<i>ADV</i> [( <i>efi</i> ) <i>sɛ</i>	<i>osuro</i> ]].
He-ran way	because	he-was-afraid

As shown in (30), Lord provides two different glosses for the sentence. The first gloss demonstrates the historical source of the conjunction ‘(*efi*)*sɛ*’, in which *sɛ* is treated as a complementizer used after a semantically explicit verb *fi* ‘come from’. In contrast, the second gloss reflects its present-day meaning and change in constituency and category label. Clearly, the cohesiveness between these elements has been changed in a way that the originally separate complementizer is fused with the preceding elements *e* ‘it’ and

*fi* ‘come from’. The grammaticalization shown in (30) involves the reanalysis of the subject-verb-complementizer *e-fi-se* into a frozen combination serving as an adverb conjunction (*efi*)*se* (from which the former subject pronoun and verb can even be completely dropped). A few other frozen combinations including the former complementizer *se* can be found in Twi: *besi se* ‘until’, *kǎnse* ‘although’, *gye-se* ‘unless’, *anase* ‘or’ (Lord 1993: 172).

Similar examples can be found in Dutch. For instance, the Dutch clausal complementizer *dat* has been merged with preceding prepositions and conjunctions and the new frozen combinations should be reanalyzed as one inseparable adverbial conjunction (Hsieh and Sybesma 2011: 4):

(31) Dutch conjunctions containing a *dat*-complementizer

<i>omdat</i>	/for-that/	‘because’
<i>zodat</i>	/so-that/	‘so as to’
<i>opdat</i>	/on-that/	‘in order to’

The frozen combinations which involve constituency reanalysis can also be found in Mandarin. A case in point which has been discussed by Dong (2004), Chen and Chen (2010) and Fang (2006, 2018), among others, would be *shuōshì* ‘it is said that’ consisting of the verb *shuō* ‘say’ and – in our current analysis – the complementizer *shì*. Chen and Chen (2010: 506-509) argue that the two elements in the combination have gone through different stages of grammaticalization and lexicalization, which can be roughly summarized as below:

(32) Action Verb *shuō* ‘say’ + Copular Verb *shì* ‘to be’ (Stage I)

> Quotative Marker *shuō* + Clausal Linker *shì* (Stage II)

> Evidential Adverb *shuōshì* (Stage III)

The following sentence from their paper are used to illustrate the three different stages in (32):

(33) *Hòulái rén shuō shì Liú Xīnshū shī yě.*  
later people say be Liu Xinshu poetry SFP  
'Later people say that it is the poetry of Liu Xinshu.'

(34) *Qián-liǎng-tiān kàn-dào yì-zé bàodào,*  
days.ago see one-CL report  
*shuō shì Ruìshì jūrán yǒu ge Gélǎngtái jùlèbù.*  
QUO SHI Switzerland surprisingly have CL Grandet club  
'I read a report days ago, which said that surprisingly there was a Grandet club in Switzerland.'

(35) *Kěshì lú yě qí-bu-dé le, shuōshì kǎnr duō.*  
but donkey also can.not.ride PRF it.is.said.that bump many  
'But we can not ride donkey either. It is said that the road is too bumpy.'

The first two stages given in (32) demonstrate the process of grammaticalization along with the loss of verbal semantics and the change from lexical function (as in (33)) towards grammatical function (as in (34)).<sup>4</sup> The second and third stages show the process of lexicalization along with the reanalysis from the combination of two separate morphemes to one adverb expressing evidentiality and the speaker's commitment (as in (35)).

---

<sup>4</sup> The earlier (25) can be seen as an intermediate stage between (33) and (34) since in (25) we have an action verb *shuō* expressing direct speech followed by a grammaticalized *shì*.

In the same spirit, Fang (2018) points out that *shuōshì* has completed the process of lexicalization. Therefore, in some contexts, in addition to its quotative meaning, it can also be used to express the speaker's weak commitment or negative evaluation to the quotative proposition. The lexicalization from *shuō-shì* to *shuōshì* is confirmed by the phonological erosion or assimilation, for instance, *shì* can be pronounced with a neutral tone or *shuō* with a contrastive stress (Fang 2018: 8).

According to Zhang (2003) and Dong (2003, 2004),<sup>5</sup> there are indeed quite a few cases in Mandarin which can be treated as frozen combinations containing one lexical element such as a focus adverb *zhǐ* 'only' or *jiù* 'only' or an adverb conjunction *kě* 'but' or *dàn* 'but', plus the grammaticalized *shì* which somehow loses its independent status. Here are a few more examples:

(36) Frozen combinations reanalyzed as adverbs:

<i>zhǐshì/jiùshì</i>	/only-that/	'It is just that...'
<i>zhǔnshì</i>	/must-that/	'It must be that...'
<i>pàshì</i>	/afraid-that/	'I am afraid that...'
<i>zhēnshì</i>	/real-that/	'truly, really'

---

<sup>5</sup> Zhang and Dong argue that before the completion of the grammaticalization or lexicalization process, *shì* should be treated as a "focal marker", which is different from the "complementizer" approach here.

*zǒngshì/lǎoshì*      /always-that/      ‘always’

(37) Frozen combinations reanalyzed as conjunctions:

*dànshì/kěshì*      /but-that/      ‘but’

*yàoshì/ruòshì*      /if-that/      ‘if’

*háishì/huòshì*      /or-that/      ‘or’

The fact that *shì* has gone through similar developments as complementizers in such diverse languages as Twi and Dutch can be taken as indication that our suggestion that it is a complementizer as well may be correct.<sup>6</sup>

#### 4. Evidence from *Gangou* Mandarin

As we saw in section 2, *shì* as a complementizer is widely used in standard Mandarin or Pǔtōnghuà. Meanwhile, it has been found that *shì* can also be used as a complementizer to introduce different types of adverbial subordinate clauses in a

---

<sup>6</sup> One of our reviewers points at the possibility of treating *shì* in the examples of our section 2 as one part of the lexicalized word, the same as in the frozen combinations in this section. However, by looking at the data we presented in section 2, we believe that the lexicalization of the combinations of these clausal adverbs, modals or conjunctions and *shì* has not occurred yet. *Shì* in those combinations in section 2 is in general optional, hence not so lexicalized (yet). But we leave open the possibility of further development of grammaticalization which leads to the lexicalization of these combinations.

northwest Mandarin dialect Gangou (Zhang and Yang 2017). Gangou is spoken at the border area of the two provinces Qinghai and Gansu (in the northwest). The language is categorized as a Mandarin dialect, based on its phonology and lexicon. However, its grammar is heavily influenced by the neighboring SOV languages, e.g., Tibetan and Monguor. As a result, its basic word order is also SOV and the complementizer *shì*, or in Zhang and Yang’s (2017) terms, the “adverbial clause marker”, also occurs in clause-final position. Below are a few examples from Zhang and Yang (2017: 54-55):

(38) *Shì* in a temporal adverbial clause:

- a. *Wǎnxī dào shì, nǐ-de ā-bà lái le.*  
 evening arrive SHI your uncle come PRF  
 ‘When the evening arrived, your uncle came.’
- b. *Tóu shuì shì, huà béng shuō.*  
 before sleep SHI words do.not speak  
 ‘Don’t talk before you go to sleep.’

(39) *Shì* in a conditional adverbial clause:

- a. *Nǐ méi xīn zǒu le shì, jiù chē zuò le qù.*  
 s2 no heart walk SFP SHI then bus sit PRF go  
 ‘If you don’t want to walk, you can also take a bus to go there.’
- b. *Míngtiān dàn yǔ-xià shì, wǒmen jiù tā bù kàn qù le.*  
 tomorrow if rain SHI we then 3s not see go SFP  
 ‘If it rains tomorrow, we will not go to see her.’

(40) *Shì* in a causal adverbial clause:

- Tā wǒ a tóuqián dǎ le shì, wǒ tā a dǎ le.*  
 s3 s1 EXC before beat PRF SHI s1 s3 EXC beat PRF  
 ‘Because he beat me first, I then beat him.’

(41) *Shì* in a concession (conditional) adverbial clause:

- Míngtiān yǔ-xià shì-a bà-ma, wǒmen tā ha kàn qù li.*

tomorrow rain SHI- EXC even.if pl.1 s3 EXC see go SFP  
 ‘Even if it rains tomorrow, we will still go to see her.’

(42) *Shì* in the consecutive clauses:

*Ní-me shì yātóu yītiān è-de-bùchéng shì, yuánzi-lǐ kàn qù shì,*  
 then SHI daughter one.day very.hungry SHI garden-in see go SHI  
*bāogǔ liǎng-ge zhòng-xià zhe shuō.*<sup>7</sup>  
 corn two- CL grow STA SHUO  
 ‘Then one day the daughter was very hungry, so she went to the vegetable garden  
 and found a few corns growing there.’

As we can see from these sentences, there is no overt conjunction (like what we saw in Standard Mandarin) indicating the clausal relation between the subordinate clauses and the matrix clause. However, the relation can be derived from the word order and context, or sometimes relying on the sentence-final modal particle like *bà-ma* in (41). Therefore, *shì* in these sentences has only taken on a clause linker function and is thus analyzable as a complementizer. The data from other Mandarin variant provides an extra piece of evidence for us to treat *shì* as complementizer in Mandarin.

## 5. Conclusion

---

<sup>7</sup> The complementizer *shì* used in this sentence is comparable to that of (24) in standard Mandarin. Interestingly, in the first clause of the sentence, we can find two instances of *shì*, one at the beginning and the other at the end of the sentence. Another interesting observation is that, in addition to *shì* used as the complementizer in the first two subordinate clauses, there is another complementizer *shuō* literally ‘say’ used in the end of the matrix sentence.

In this paper, I have shown that *shì* is not always a copular verb or focus marker. In certain well-defined environments, namely following clause selecting modals, adverbs, conjunctions and certain speech-act and so-called psych verbs, it displays the behavior that we know from complementizers in other languages.

In its use as a complementizer, *shì* is comparable to the “say-complementizers” we also find in Sinitic languages, such as *shuō* in Mandarin, *kong<sup>l</sup>* in Taiwanese Southern Min, and *waa6* in Cantonese, which can, among other contexts be used as a complementizer after various types of verbs, e.g., verbs expressing a speech-act, cognition, emotion and modal verbs (Cheng 1991, 1997, Hwang 1998, Huang 2003, Wang, Katz and Chen 2003, Fang 2006, Yeung 2006, Chappell 2008, 2017, Hsieh and Sybesma 2011).

Chappell (2017) lists the five major sources of complementizers across languages, namely nouns meaning ‘thing’, ‘fact’ or ‘place’ (e.g., Korean *kes* ‘thing’, Thai *thîi* ‘place’), demonstrative, interrogative and relative pronouns (e.g., English *that*, German *daß* ‘that’, French *que* ‘what’), dative, allative and locative case markers or prepositions (e.g., English *to*), verbs of saying (e.g., Ewe *bé*, Nepali *bhan*) and elements with the meaning of ‘resemble’ or ‘be like’ (e.g., Idoma *bɛ* ‘resemble’). Interestingly, we now find complementizers from two different sources in Mandarin, *shuō*, mentioned above, originally from a verb of saying, and *shì*, which in earlier phases in the development of the Chinese language, was a demonstrative pronoun before it developed into a copular verb (Wang 1980: 351-354; Shi and Li 2001: 12-52; see also Cheng 2021). In Pre-Qin

times (before 221 BC), *shì* was used as a demonstrative pronoun to refer to an antecedent, i.e., a nominal or clausal subject. Here are two examples from Wang (1980):

- (43) *Fù yǔ guì, shì rén zhī suǒ yù yě.*  
 wealth and honor DEM people ATTR NOM desire SFP  
 ‘Wealth and honor, this is what people desire to get.’  
 (From *Lúnyǔ: Lǐrén*)

- (44) *Qiān lǐ ér jiàn wáng, shì yǔ suǒ yù yě.*  
 thousand li to see king DEM s1 NOM desire SFP  
 ‘When I came a thousand li to wait on the king, this was what I desired to do.’  
 (From *Mèngzǐ: Gōngsūn Chǒu* II)

Interestingly, in Mandarin, in certain cases, the two complementizers *shuō* and *shì* are interchangeable without repercussions for the meaning:

- (45) *Bùguǎn (shì/shuō) háizi xiǎng-yào shénme, tā dōu mǎnzú.*  
 no.matter SHI/SHUO kids want what s3 all satisfy  
 ‘He will satisfy anything that the kids want.’

- (46) *Hǎoxiàng (shì/shuō) yǒu-rén shòushāng le.*  
 seemingly SHI/SHUO someone get.injured PRF  
 ‘Seemingly someone has got injured.’

- (47) *Tā zhèngshí (shì/shuō) Zhāng Sān tōu le yáng.*  
 s3 confirm SHI/SHUO Zhang San steal PRF sheep  
 ‘He confirmed that Zhang San stole his sheep.’

This means that it makes sense to analyze them in the same way, one more argument, for treating *shì* as a complementizer, in addition to the ones given above.

## REFERENCES

- BRENNAN, Virginia. 1997. Modalities. Unpublished manuscript. Vanderbilt University.
- BUTLER, Jonny. 2003. A minimalist treatment of modality. *Lingua*, 113(10): 967-996.
- CHAO, Yuen Ren. 1968. *A spoken Grammar of Chinese*. Berkeley: University of California Press.
- CHAPPELL, Hilary. 2008. Variation in the grammaticalization of complementizers from verba dicendi in Sinitic languages. *Linguistic Typology* 12 (1): 45–98.
- CHAPPELL, Hilary. 2017. Say-complementizers in Chinese languages. in *Encyclopedia of Chinese Language and Linguistics*, edited by Rint Sybesma et al, Leiden: Brill.
- CHENG, Hang. 2021. *All structures great and small: on copular sentences with shì in Mandarin*. Doctoral dissertation, Leiden University, Leiden.
- CHENG, R.L. 1997. *Tái-yǔ yǔ Táiwānhuáyǔ-li de zǐjù-jiégòu-biāozhì “kong” yǔ “khuann”* (The sentence structure markers *kong* and *khuann* in TSM and Taiwan Mandarin). *Taiwanese and Mandarin Structures and their developmental trends in Taiwan II: Contacts between Taiwanese and Mandarin and restructuring of their synonyms*, 105-132. Taipei, Yuan-Liou. [Originally published in 1991]
- CHEN, Ying and Yi Chen. 2010. *Gùhuà jiégòu “shuōshì” de yǎnhuà jīzhì jīqí yǔyòng gōngnéng* (The evolution mechanisms and pragmatic functions of the fixed phrase *shuōshì*). *Shijie Hanyu Jiaoxue* 4: 505-513.
- CINQUE, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-linguistic*

- Perspective*, New York: Oxford University Press.
- COOK, Walter A. 1978. Semantic structure of English modals. *TESOL Quarterly* 12: 5–17.
- DONG, Xiufang. 2003. “X+shuō” de cíhuì-huà (On the Lexicalization of “X+shuō”). *Yuyan Kexue* 2 (2): 46-57.
- DONG, Xiufang. 2004. “Shì” de jìnyībù yǔfǎ-huà: yóu xūcí dào cí-nèi chéngfèn (Further Grammaticalization of shì: From Functional Word to Word-internal Element). *Dangdai Yuyan-xue* 1: 35-44.
- ERNST, Thomas. 2004. *The Syntax of Adjuncts*. Cambridge: Cambridge University Press.
- FANG, Mei. 2006. Běijīng-huà-lǐ “shuō” de yǔfǎ-huà (The grammaticalization of shuō in Pekingnese). *Zhongguo Fangyan Xuebao* 1, 107-121.
- FANG, Mei. 2018. “Shuōshì ” de huàyǔ gōngnéng jí xiāngguān cíhuì-huà wèntí (The discourse function of shuōshì and its lexicalization) *Zhongguo Yuyan Xuebao* 18:1-14.
- HUANG, Shuan-fan. 2003. Doubts about complementation: a functionalist analysis”. *Language and linguistics* 4 (2): 429-455.
- HOPPER, Paul J., and Elizabeth Closs Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- HWANG, Jya-Lin. 1998. A comparative study on the grammaticalization of saying verbs in Chinese. *NACCL 10: 10th North American Conference on Chinese Linguistics*, ed. by Chaofen Sun, 574-584.

- HSIEH, Feng-fan and Rint Sybesma. 2011. On the linearization of Chinese sentence-final particles: Max spell out and why CP moves. *Korea Journal of Chinese Language and Literature* 1: 53-90.
- JACKENDOFF, Ray, 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA and London: MIT Press.
- LIN, Jo-wang and Chih-Chen Jane Tang. 1995. Modals as verbs in Chinese: A GB perspective. *The Bulletin of the Institute of History and Philology* 66 (4): 53-105.
- LIN, Tzong-hong Jonah. 2012. Multiple-modal constructions in Mandarin Chinese and their finiteness properties. *Journal of Linguistics* 48: 151–186.
- LORD, Carol. 1993. *Historical change in serial verb constructions*. Vol. 26. Amsterdam: John Benjamins Publishing.
- PARK Hae-Kyoung. 2011. *Cíhuì huà xíngshì “gāopín shuāngyīnjié néngyuàn dòngcí + shuō / shì”* (On the lexicalization form of ‘high -frequently used disyllabic optative verb + shuō/shì’), *Shijie Hanyu Jiaoxue* 4: 470-478.
- SHI, Yuzhi and Charles N. Li. 2001. *Hànyǔ yǔfǎ-huà de lìchéng* (A history of Grammaticalization in Chinese). Beijing: Peking University Press
- TSAI, Wei-tien Dylan. 2015. *Cóng wēiguān dào hōngguān: Hànyǔ yǔfǎ de shēngchéng shìyě* (From microscopic view to cosmic perspective: generative landscape of Chinese syntax). Beijing: Shangwu Yinshuguan.
- WANG, Li. 1980. *Hanyu Shi Gao* [Manuscript of Chinese Language History] Reprint: Beijing: Zhonghua Shuju, 2002.
- WANG, Yu-Fang, Aya Katza, and Chih-Hua Chen. 2003. Thinking as Saying: Shuo

- ‘say’ in Taiwan Mandarin Conversation and BBS Talk, *Language Sciences* 25: 457-488.
- YANG, Zhaole. 2020. *Yě, yě, yě : on the syntax and semantics of Mandarin yě*.  
 Doctoral dissertation, Leiden University, Leiden.
- YEUNG, Kawaii. 2006. On the status of the complementizer *waa6* in Cantonese.  
*Taiwan Journal of Linguistics* 4 (4): 1-48.
- ZHANG, Jingting and Yonglong Yang. 2017. *Qīnghǎi mínhé gāngōu-huà de zhuàngyǔ cóngjù biāoji “shì” jīqí láiyuán* (The adverbial clause marker *shì* in Gangou dialect and its source). *Yuwen Yanjiu* 2: 53-59
- ZHANG, Yisheng. 2003. “*Fùcí + shì*” *de lishi yǎnhuà hé gòngshí biànyì* (On the Connections Between Diachronic Changes and Synchronic Variations of “adverb+*shì*” in Chinese). *Yuyan Kexue* 2 (6): 34-49.