

A Pilot Corpus Study of the Grammaticalization Status of *Shuō* as a Complementizer in Mandarin Chinese

Liwei GONG

Abstract

Although traditional grammars widely accept that Mandarin Chinese lacks an overt complementizer like that in English, recent literature shows that a ‘say’ verb is undergoing grammaticalization, serving as a semi-complementizer in a wide range of variants or dialects of Chinese. This study conducts a pilot corpus analysis of the grammaticalization status of 说 *shuō* ‘say’ using the most recent natural speech data from Mandarin podcasts. Contrary to most previous studies, the results indicate that 说 *shuō* ‘say’ in Mandarin can introduce a complement clause in various semantic contexts, notably including specific emotion and factive verbs, which signifies a high level of grammaticalization.

[Keywords: corpus / complementizer / grammaticalization / Mandarin Chinese]

1. Introduction

It is well-established in traditional reference grammars and syntactic studies that Mandarin Chinese (henceforth Mandarin) lacks overt complementizers like *that* in English or *que* in French (Li and Thompson 1981: 598-602; Ramsey 1987: 86). However, more recent research suggests that the ‘say’ verbs in several variants and dialects of Chinese have grammaticalized into a quotative marker or a complementizer to varying degrees (e.g., Wang et al. 2003 for 说 *shuō* in Taiwanese Mandarin; Fang 2006 for 说 *shuō* in Beijing Dialect; Yeung 2006 for 話 *waab* in Cantonese; Huang 2016 for 呶 *tā* in Jieyang Dialect; see also Chapell 2008 for 講 *kóng* in Southern Min and a systematic analysis of this grammaticalization process in Sinitic languages and dialects).¹ The grammaticalization pathway SAY VERBS > QUOTATIVE MARKERS AND COMPLEMENTIZERS is widely observed across a wide range of languages, including African languages like Bemba (Givón 1980) and Ewe (Saxena 1995; Lord 1993), Austronesian languages, such as Buru and Tukang Besi (Klamer 2000), and various other languages listed in Kuteva et al. (2019: 375-379). It is argued that in these languages, the ‘say’ verbs first cooccur with various communicative verbs in serial verb constructions. The possible contexts then gradually extend

to cognition and perception verbs, or even modal and factive verbs in later stages. The ‘say’ verbs are grammaticalized into complementizers along the process.

However, there are also linguists who oppose the complementizer analysis of ‘say’ verbs in variants of Chinese. Paul (2014) provides counterexamples from a generative perspective, challenging the complementizer analysis of *shuō* in Mandarin. For example, the embedded clause introduced by *shuō* cannot be fronted or used as a sentential subject, and *shuō* cannot be stranded (Paul 2014: 98). Long and Deng (2023) question the complementizer status of *shuō* from a functional perspective. They argue that *shuō* in Mandarin is more likely to be a quotative marker introducing direct quotations instead of a complementizer marking complement clauses.

Apart from the theoretical discussions of this phenomenon in various dialects of Chinese from different perspectives, corpus studies have become a significant approach towards language changes, or more specifically, the grammaticalization status of an evolving grammatical marker. However, no corpus studies on the grammaticalization of *shuō* in Standard Mandarin have been conducted since Chappell (2008), which only surveyed dialogues in one single novel written in 1993. In order to address this gap and investigate the most recent usage of *shuō* in Mandarin, I conduct a corpus analysis using a self-compiled corpus of Mandarin podcast transcripts, encompassing a significant volume of conversational speech produced in 2023. Specifically, I focus on instances where the term *shuō* follows another event-denoting verb and precedes a complement clause. I analyze the attested collocation patterns and further examine the occurrence of certain unreported collocations in The Chinese Web Corpus (zhTenTen), a large-scale corpus of web language. In addition to the possible contexts of *shuō*, I also briefly discuss three other major parameters of grammaticalization as defined by Narrog and Heine (2022): Desemanticization, Decategorization, and Erosion.

This paper is organized as follows. Section 2 introduces the methodology of this study, including the data collection process and some theoretical prerequisites. Section 3 presents major findings of the corpus study and the relevant discussions. Section 4 addresses Long and Deng’s (2023) recent arguments against the complementizer analysis of *shuō* in Mandarin.

2. Methodology

2.1 Data Collection

I collected natural speech data from nine episodes of open-access Mandarin podcasts from nine different podcast channels. The episodes selected were among the most popular in September 2023, played over 50,000 times within the two weeks following their release.² The length of each episode ranges from 16 minutes to 111 minutes, and the constructed corpus (henceforth “the podcast corpus”) comprises 11 hours and 29 minutes of conversations in total.

The audio data was transcribed using Ifyrec's speech transcription service³, and the transcripts contain a total of around 224,000 Chinese characters. Subsequently, I manually identified all the instances where *shuō* follows another primary verb and is followed by a complement clause.

To ensure that these usages were not idiosyncratic to specific speakers, all collocations found in the podcast corpus were further verified using The Chinese Web Corpus (zhTenTen)⁴. There are two main reasons for using zhTenTen for the secondary analysis. First, since it is likely that *shuō* as a complementizer is still in preliminary grammaticalization status in Mandarin (spoken in Mainland China), its usages are more likely to occur in informal and casual contexts such as web languages. Second, zhTenTen is the most recent and largest corpus of Mandarin web language to the best of my knowledge. The corpus was compiled in 2017 and comprised of 13.5 billion Chinese words using simplified characters.

Additional features of each relevant instance are further annotated. First, the main verbs cooccurring with *shuō* are classified into four major semantic classes according to the semantic contexts of the four grammaticalization statuses proposed by Chapell (2008, 2017): communicative speech act verbs, cognition and perception verbs, emotion and stative verbs, and modal and factive verbs. Second, I categorized the complements in the relevant instances into three types: direct reports, indirect reports, and unclear cases. The distinction between direct report complements and indirect report complements in Chinese is not always conspicuous, given that there are no overt and obligatory tense markings and the data sources in this study are recordings. Nevertheless, the classification is possible for a part of the instances which contains hints from person indexations or discourse particles. This distinction is relevant in that an indirect report generally represents a higher degree of grammaticalization of the complementizer than a direct report, since the indirect complement is better embedded under the matrix clause and better bonded with the complementizer (cf. Long and Deng's 2023 argument that a direct report/quotation is not a complement clause). Finally, instances where *shuō* appear to be shortened or unstressed are also noted, as these cases may represent some form of erosion and thus a high level of grammaticalization (see also Chapell 2008: 56 for cases where *kóng* 'say' is shortened in Southern Min).

All selected instances and categorizations of the aforementioned features are presented in the Supplementary Materials.⁵

2.2 The grammaticalization of 'say' verbs as complementizers

Chapell (2008, 2017) proposes several hypothetical stages on a continuum of the 'say' verb grammaticalization.

(1) Stages of the ‘say’ verb grammaticalization (adapted from Chapell 2008: 58-62)

Stage I (Quotative constructions):

NP_{subject} V_{say} [quotative verb]: [QUOTATION]

Stage II (Serial verb constructions):

NP_{subject} VI[speech act verb] V2_{say} [quotative/semi-complementizer]: [QUOTATION]

Stage III (Complement constructions with cognition verbs):

NP_{subject} VI[cognition/perception verb] V2_{say} [complementizer] [COMPLEMENT]

Stage IV (Complement constructions with emotion and stative verbs):

NP_{subject} VI[emotion/stative verb] V2_{say} [complementizer] [COMPLEMENT]

Stage V (Conventionalized complement constructions):

NP_{subject} VI[modal/factive verb] V2_{say} [complementizer] [COMPLEMENT]

As introduced in Section 2.1, I categorized the “V1” that occurs in each of the relevant instances into one of the four semantic classes, corresponding to Stage II through Stage V in Chapell’s framework. In this way, the attested collocations in the podcast corpus reflect the grammaticalization status of *shuō* as a complementizer in the most up-to-date usage.

In addition to the context extension, I will also briefly discuss the other three major parameters of grammaticalization defined by Heine and Narrog (2022) in Section 3, i.e. Desemanticization, Decategorization, and Erosion. Finally, I will explore the implications of this corpus study for Long and Deng’s (2023) recent analysis.

3. Results and Discussion

3.1 Semantic contexts of *shuō* complement constructions

Within the constructed podcast corpus, a total of 129 instances were found where *shuō* follows another lexical verb and precedes a complement clause.⁶ The frequencies of the co-occurring verbs, categorized by semantic context, are presented in Table 1, while Table 2 aggregates these frequencies according to semantic classes.

Table 1: Frequencies of main verbs followed by *shuō* ‘say’ and a complement clause.

Cooccurring main verbs	Frequencies	Semantic classes
觉得 <i>juéde</i> ‘feel’ / ‘think’	32	cognition/perception
问 <i>wèn</i> ‘ask’	12	communicative speech act
讲 <i>jiǎng</i> ‘speak’	9	communicative speech act
写 <i>xiě</i> ‘write’	5	communicative speech act
发现 <i>fāxiàn</i> ‘discover’ / ‘find’	5	cognition/perception
希望 <i>xīwàng</i> ‘hope’ / ‘wish’	4	cognition/perception
想到 <i>xiǎngdào</i> ‘think of’	4	cognition/perception
谈 <i>tán</i> ‘talk’ / ‘discuss’	3	communicative speech act

Cooccurring main verbs	Frequencies	Semantic classes
讨论 <i>tāolùn</i> 'discuss'	3	communicative speech act
知道 <i>zhīdào</i> 'know'	3	cognition/perception
告诉 <i>gàosù</i> 'tell'	2	communicative speech act
讲到 <i>jiǎngdào</i> 'mention'	2	communicative speech act
聊 <i>liáo</i> 'chat'	2	communicative speech act
骂 <i>mà</i> 'scold'	2	communicative speech act
关注 <i>guānzhù</i> 'pay attention to'	2	cognition/perception
怀疑 <i>huáiyí</i> 'doubt'	2	cognition/perception
看到 <i>kàndào</i> 'see'	2	cognition/perception
认为 <i>rènwéi</i> 'believe' / 'think'	2	cognition/perception
在乎 <i>zàihu</i> 'care about'	2	cognition/perception
安慰 <i>ānwèi</i> 'comfort' / 'console'	1	communicative speech act
...		
(another 22 communicative verbs with a frequency of 1)	1	communicative speech act
...		
确定 <i>quèdìng</i> 'confirm'	1	cognition/perception
想 <i>xiǎng</i> 'think'	1	cognition/perception
想过 <i>xiǎngguò</i> 'have thought of'	1	cognition/perception
想象 <i>xiǎngxiàng</i> 'imagine'	1	cognition/perception
好奇 <i>hàoqí</i> 'be curious about'	1	emotion
怕 <i>pà</i> 'be afraid of'	1	emotion
导致 <i>dǎozhì</i> 'cause'	1	general factive
证明 <i>zhèngmíng</i> 'prove'	1	general factive
Summary	129	

Table 2: Frequencies of main verbs by semantic classes.

Semantic classes	Frequencies
communicative speech act	63
cognition/perception	62
emotion	2
general factive	2

The results indicate that, within the podcast corpus comprised of conversations, *shuō* most frequently cooccurs with communicative speech act verbs, exemplified by *jiǎng* 'say' in (2) and *wèn* 'ask' in (3).

(2)	他	就	讲	说	在	这个	活动	上
	<i>tā</i>	<i>jiù</i>	<i>jiǎng</i>	<i>shuō</i>	<i>zài</i>	<i>zhège</i>	<i>huódòng</i>	<i>shàng</i>
	3SG.M	then	say	say	on	this	event	on
	他	一直	想要	跟	乌尔善	讲话		
	<i>tā</i>	<i>yízhí</i>	<i>xiǎngyào</i>	<i>gēn</i>	<i>wūěershàn</i>	<i>jiǎnghuà</i>		
	3SG.M	always	want	to	W.	talk		

'He then said that on this event, he has always wanted to talk to W.'

- (3) 很多 人 发私信 问 我 说 你 怎么
hěnduō rén fāsīxin wèn wǒ shuō nǐ zěnmē
 a lot people send-message ask 1SG say 2SG how
 在 舞台上 那么 自如
zài wǔtáishàng nàme zìrú
 on on-stage so easy
 'A lot of people sent me messages and asked me (say) "how can you be so easy on the stage?"'

Although the cooccurrence of *shuō* with communicative verbs in serial verb constructions is well-documented in traditional grammars (e.g., Li and Thompson 1981: 601-602), the nearly equal frequency of *shuō*'s occurrence with cognition/perception verbs in Mainland Mandarin has been scarcely noted in existing literature. This corpus investigation demonstrates that *shuō* in Mandarin can actively appear after various cognition/perception verbs, such as *juéde* 'feel' / 'think' in (4), *fāxiàn* 'discover' in (5), and *xīwàng* 'hope' / 'wish' in (6).

- (4) 她 觉得 说 给 角色 穿上衣服 的 过程
tā juéde shuō gěi juésè chuānshàngyīfú de guòchéng
 3SG.F think say for character get dressed REL process
 对她来说 是 一 个 具身化 的 过程
duìtāláishuō shì yī gè jùshēnhuà de guòchéng
 for her COP one CL embodiment ATTR process
 'She_i thinks that the process of getting the characters dressed is a process of embodiment for her_i.'⁷

- (5) 我 就 发现 说 其实 没 有 太多 的 区别
wǒ jiù fāxiàn shuō qíshí méi yǒu tàiduō de qūbié
 1SG then discover say actually NEG exist much ATTR difference
 'I then discovered that there is actually not much difference.'

- (6) 创作者 还是 希望 说 自己 是
chuàngzuòzhě háishi xīwàng shuō zìjǐ shì
 creators still hope say self FOC
 有 自己 的 主体性
yǒu zìjǐ de zhūtíxìng
 have self POSS individuality
 'The creators still hope that they can have their own individuality.'

In addition, there are also instances where *shuō*, introducing a complement clause, appears after emotion verbs such as *pà* ‘be afraid of’ in (7), or factive verbs such as *dǎozhī* ‘cause’ in (8).

- (7) 你 也 很 怕 说 他 会 违心
ní yé hěn pà shuō tā huì wéixīn
 2SG also very afraid say 3SG.M would against one’s will
 地 做 一个 他 不 喜欢 的 东西
de zuò yīgè tā bù xǐhuān de dōngxi
 ADV create one 3SG.M NEG like REL thing
 (you would hope that his work can satisfy our requirements and expectations, but at the same time...)

‘you are also afraid that he would make something he does not like against his will’

- (8) 所以 导致 说 这个 东西 它 的 运作
suóyǐ dǎozhī shuō zhège dōngxi tā de yùnzuo
 so cause say this thing 3SG.N POSS operation
 周期 会 比 预想当中 要 长 非常多
zhōuqī huì bǐ yùxiǎngzhōng yào cháng fēichángduō
 period will CMP expectation will long a lot

‘so (the situation) caused that this thing, its operation period will be much longer than expected.’

According to Chapell’s (2008: 58-63) hypothetical stages of grammaticalization of ‘say’ verbs, the presence of emotion and factive verbs as semantic contexts represents the final two stages of the grammaticalization and the “onset of conventionalization of the complementizer usage”. It has not been reported in any previous studies that *shuō* in Mainland Mandarin has reached this stage.

I further examined the collocations attested in the podcast corpus using zhTenTen, a large-scale Chinese web language corpus. All collocations attested in the podcast speech were also observed in zhTenTen, including more occurrences where *shuō* cooccurs with emotion verbs (e.g. *hàoqí* ‘be curious about’ in [9a], *pà* ‘be afraid of’ in [9b]) and factive verbs (*dǎozhī* ‘cause’ in [9c] and *zhèngmíng* ‘prove’ in [9d]).

- (9) a. 后来 我们 好奇 说 自己 去 可以 吗
hòulái wǒmén hàoqí shuō zìjǐ qù kěyǐ mā
 later 1PL curious say self go can Q

‘Later, we were **curious say**that if we can go by ourselves.’

- b. 微波炉 很多 人 会 怕 说 辐射
wēibōlú *hěnduō* *rén* *huì* *pà* *shuō* *fúshè*
 microwave.oven many person will afraid say radiation
 比较 大
bǐjiǎo *dà*
 quite strong

'As for microwave ovens, many people **are afraid say**that the radiation could be quite strong... (so it is even more necessary to choose a reliable brand)'

- c. 你 这 个 应该 不 会 导致 说
nǐ *zhè* *ge* *yīnggāi* *bú* *huì* *dǎozhì* *shuō*
 2 this CL probably NEG will cause say
 温度 特别 高
wēndù *tèbié* *gāo*
 temperature very high

'This [problem] of yours will probably not **cause say**that the temperature to be very high... (it mainly depends on the temperature of the CPU and background processes)'

- d. 那 证明 说 它 不 是 一 个
nà *zhèngmíng* *shuō* *tā* *bù* *shì* *yī* *ge*
 then prove say 3SG.N NEG COP one CL
 有效 的 方法
yǒuxiào *de* *fāngfǎ*
 effective ATTR method

'(if there is one method that can solve whatever problem...) then it **proves say**that it is actually not an effective method.' (zhTenTen)

In summary, the results show that *shuō* in Mandarin Chinese may have undergone major context extension in the past one or two decades as it can now introduce complement clauses following various cognition/perception verbs, certain emotion verbs, and even factive verbs.

3.2 Other parameters of grammaticalization

In examples (4) through (9), the literal meaning 'say' is largely incongruent with the semantic contexts where the complement clauses describe certain wishes, beliefs, themes of emotions, or even factive events. There is no actual speech being uttered or conveyed in these contexts, indicating a noticeable level of desemanticization of 'say'.

The 'say' verb also exhibits signs of decategorization since it is not marked for aspect or

modality markers in all relevant instances found in the podcast corpus. The loss of aspect/modality marking is particularly evident in semantic contexts involving emotion verbs or factive verbs. For example, it is impossible to mark any aspect or modality on *shuō* in constructions like *pà shuō* ‘be afraid say’ or *dǎozhī shuō* ‘cause say’.

- (10) 你 也 很 怕 (*能) 说 (*了) [他 会 ...] complement
*nǐ yě hěn pà (*néng) shuō (*le) tā huì ...*
 2SG also very afraid (*can) say (*PFV) [3SG.M would ...]
 Intended: ‘you can be afraid that [complement]’
 or ‘you have been afraid that [complement]’

Yeung (2006) reported a similar pattern on the Cantonese *waa6* ‘say’: it loses its ability to take aspect markers or undergo verb-doubling when introducing a complement clause.

Finally, I observed that *shuō* in several instances is either unstressed or pronounced as /zuo/ or /uo/ instead of the usual /ʃuo/.

- (11) 底下 都 骂 他 说 白酒 文化 真 是 糟粕
dǐxià dōu mà tā shuō báijiǔ wénhuà zhēn shì zāopò
 down all scold 3SG.M say liquor culture really COP dregs
 ‘People down there all scolded him (saying) that the liquor cultures are really dregs.’

In (11), *shuō* was shortened to /uo/ in speech and was unstressed. This may represent a certain degree of erosion and loss of morphological freedom. However, as Chapell (2008: 56) noted for the shortening of *kóng* in Southern Min, it is also possible that this is simply an effect of fluent speech. I have not conducted sufficient analyses to address this phonological process in depth, and I will leave this aspect of the grammaticalization process for future studies.

4. Addressing Long and Deng’s (2023) analysis of *shuō* as a quotative marker

Long and Deng (2023) is one of the latest functional analyses concerning the grammaticalization of *shuō* in Chinese. Contrary to the majority of prior research that considers *shuō* as a complementizer, they argue that in Standard Mandarin, *shuō* functions as a quotative marker, introducing a quotative parenthetical clause rather than a complement clause. Specifically, they argue that a complement clause must be an indirect report without illocutionary force. The actual structure of constructions such as 他们认为说 *tāmén rènwéi shuō [complement]* ‘3PL think say [COMPLEMENT]’ consists of two coordinated events, ‘think’ and ‘say’. The subject of *shuō* is omitted and the following clause is actually a direct quotation

which can take sentence final particles and thus have illocutionary force, as in examples like (12) and (13).

- (12) 他们 就 认为 (他们) 说, 这 得 政府 给
tāmén jiù rènwéi tāmén shuō zhè dě zhèngfǔ gěi
 3PL thus think they say DEM should government for
 我们 解决 (啊)
wǒmén jiějué a
 1PL solve SFP

‘They thus thought and (they) said, the government should solve the problem for us.’ (Long and Deng 2023: 147)

- (13) 我 记得 说 她 小时候 很 爱 哭 也
wǒ jìdé shuō tā xiǎoshíhòu hěn ài kū yě
 1PL remember say she young time very love cry
 很 爱 笑 (呀)
hěn ài xiào ya
 very love smile SFP

‘I remember (saying) that she cried a lot and smiled a lot in her childhood.’ (Long and Deng 2023: 148)

I intend to present counterarguments to this analysis partially based on the findings from this corpus investigation.

First, the specific analysis of examples (12) and (13) presents two significant issues. If the proposed subject of ‘say’, 他们 *tāmén* ‘3PL’, is restored, (12) cannot have the intended meaning suggested by the translation without a significant prosodic pause between 认为 *rènwéi* ‘think’ and 他们 *tāmén* ‘3PL’. The sentence would instead be understood as “they thought that they said [complement]”. In addition, the presence of sentence-final particles does not confirm that the purported quotative clause possesses illocutionary force, as these particles can also operate at the matrix clause level. This is evident in example (14), where an indirect report complement clause appears with the sentence-final particle *a* (cf. Examples [5] and [7] in Long and Deng 2023).

- (14) - 他 来 吗? - 他 说了 他 不 来 啊
tā lái ma tā shuō-le tā bù lái a
 3SG.M come Q 3SG.M say-PFV 3SG.M NEG come SFP

‘Will he come?’ – ‘He did say that he would not come’

Second, the results of this corpus study constitute conspicuous counterexamples to Long and Deng’s (2023) argument that the clauses following *shuō* ‘say’ in Mandarin are quotative parenthetical clauses and *shuō* is a quotative marker/verb in those cases. As briefly mentioned in Section 2, although Mandarin lacks overt tense marking and many other inflectional verbal categories, person indexation clearly identifies some of the instances in question as indirect reports. The frequencies of direct reports, indirect reports and unclear cases are presented in Table 3.

Table 3: Frequencies of direct reports and indirect reports following V *shuō* constructions.

Complement types	Frequencies
Indirect Report	15
Direct Report	51
Unclear	62
Summary	129

Fifteen instances of indirect reports following the V *shuō* construction are observed in the podcast corpus, as illustrated by, for example, (4) and (14).

The remaining relevant examples are available in the Supplementary Materials (see Section 2.1). The *shuō* (semi-)complementizer is capable of introducing not only quotative direct reports but also indirect reports or propositions.

Third, the quotative marker analysis might still be acceptable in the semantic contexts involving cognition verbs given that the metaphor “thinking is speaking” is widely encoded in human languages. However, this analysis becomes less acceptable in the contexts of perception verbs like ‘see say that’, and completely untenable in factive contexts like ‘prove say that’ and ‘cause say that’, as illustrated in (8), (9c), and (9d). In these situations, nothing is uttered or even thought of. The complement clauses denote factive events in the reality and such usages signify a relatively high degree of grammaticalization of *shuō* into a complementizer as proposed in Chapell (2008: 61-63).

5. Conclusions

This study aims to explore the current grammaticalization status of *shuō* ‘say’ as a complementizer in Mandarin Chinese, using a self-constructed corpus of recent podcast conversations. The results show that *shuō* ‘say’ can introduce a subordinate clause across various semantic contexts, including those involving emotion and factive matrix verbs, although communicative verbs and cognition verbs remain the most frequent contexts. In addition to context extension, *shuō* ‘say’ in Mandarin has also undergone some degree of desemanticization,

deategorization, and presumably erosion. It appears that *shuō* ‘say’ in Standard Mandarin has made significant progress on the grammaticalization continuum, when compared with its status described in Chappell (2008). This study also shows the significance of a corpus linguistic approach within the field of grammaticalization, capturing recent usages of *shuō* in natural speech which may be neglected in major theoretical discussions.

References

- 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: *shuō* 說, *waa3* 話, *kong1* 講. In Rint Sybesma (ed.), *Encyclopedia of Chinese language and linguistics*. London: Brill.
- Fang, Mei. 2006. 北京话里“说”的语法化－从言说动词到从句标记 (*Beijingshua-li ‘shuo’de yufahua-cong yanshuo dongci dao congju biaoji*) [Grammaticalization of *shuō* (say) in Beijing Mandarin: from lexical verb to subordinator]. *中国方言学报 Zhongguo Fangyan Xuebao* 1. 107-121.
- Givón, Talmy. 1980. *On understanding grammar*. New York: Academic Press.
- Heiko Narrog & Bernd Heine. 2021. *Grammaticalization*. Oxford University Press.
- Huang, Yanxuan. 2016. 揭陽方言言說動詞“呔”的語法化 (*Jiēyáng fāngyán yánshuō dòngcí de yǔfàhuà*) [Grammaticalization of the SAY verb *da* (呔) in the Jieyang dialect]. *中国語文 Zhōngguó Yǔwén* 6. 686-694.
- Klamer, Marian. 2000. How report verbs become quote markers and complementizers. *Lingua* 110. 69-98.
- Kuteva, Tania, Bernd Heine, Bo Hong, Haiping Long, Heiko Narrog & Seongha Rhee. 2019. *World lexicon of grammaticalization*, 2nd edn. Cambridge: CUP.
- Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley and Los Angeles: University of California Press.
- Long, Haiping & Chunlin Deng. 2023. Do ‘say’ verbs really grammaticalize into complementizers through clause combination?: Evidence from Chinese *shuō* ‘say’. *Functions of Language* 30(2). 137-158.
- Lord, Carol. 1993. *Historical change in serial verb constructions*. Amsterdam Benjamins.
- Paul, Waltraud. 2014. Why particles are not particular: Sentence-final particles in Chinese as heads of a split CP. *Studia Linguistica* 68(1), 77-115.
- Ramsey, S. Robert. 1987. *The languages of China*. Princeton, NJ: Princeton University Press.
- Saxena, Anju. 1995. Unidirectional grammaticalization: Diachronic and synchronic evidence. *Sprachtypologie und Universalienforschung* 48. 350-372.
- Wang, Yu-Fang, Aya Katz & Chi-Hua Chen. 2003. Thinking as saying: *Shuo* (‘say’) in Taiwan Mandarin conversation and BBS talk. *Language Sciences* 25. 457-488.
- Yeung, Kawaii K. Y. 2003. The grammaticalization of complementizers from verbs of ‘saying’: The case of Cantonese *waa6*. *North American Conference on Chinese Linguistics* 15. 487- 503.

Note

- 1 The Chinese characters for 说 *shuō* and other words will be omitted in the remaining text for the sake of simplicity.

- 2 I referred to the public statistics of Mandarin podcasts available on <https://xyzrank.com/#/hot-podcasts>.
- 3 <https://www.iflyrec.com/>.
- 4 <https://www.sketchengine.eu/zhtenten-chinese-corpus/>.
- 5 https://docs.google.com/spreadsheets/d/1eSA8cTxe9r-T0G84_u2EM1PNB_j0vJ5rTX8tJX_beHU/edit?usp=sharing
- 6 The cooccurring lexical verb may appear directly before *shuō* or is positioned with one intervening word before *shuō*. The purpose of including the latter situation is to include communicative verbs that take an addressee argument together with a complement argument, such as “he told me say [complement]”.
- 7 The two pronouns in (4) refer to the same referent, indicating that the complement clause is an indirect report. This will be further discussed in Section 4.

(Liwei GONG International Graduate Program in Language Sciences, Graduate School of
International Cultural Studies)