

The semi-complementizer *shuō* and non-referential CPs in Mandarin Chinese

Jiahui Huang*

Abstract. The empirical focus of this paper is the syntactic status of the semi-complementizer *shuō* grammaticalized from verbs of saying, in Mandarin Chinese. Such elements have been shown to exhibit atypical patterns compared to *that* in English, which triggers discussions of whether *shuō* should be analyzed as a complementizer (Paul, 2014; Huang, 2018). This paper presents novel data surrounding the distributional patterns of *shuō* and argues that *shuō* is a C head that introduces a subtype of CPs called non-referential CPs, following de Cuba (2017).

Keywords. complementizer; grammaticalization; CP; referentiality; Mandarin Chinese; finiteness; syntax

1. Introduction. The grammaticalization of verbs of saying (e.g. ‘say’) into complementizers, has been observed in a variety of languages across language families, including Indo-Iranian, Austronesian, Nakh-Daghestanian, Uralic languages, and Sino-Tibetan languages (Hock 1982, Klammer 2000, Heine & Kuteva 2002, Chappell 2008, Burukina 2021). In Mandarin Chinese¹, a well-documented candidate is the verb *shuō* (說), which not only retains its lexical meaning of ‘say’ as the main predicate (1a), but also exhibits behavior similar to that of the complementizer *that* in English in that it can occur in a position that follows matrix verbs and precedes clausal complements (1b) and (1c). Due to the preservation of the concrete lexical meaning of *shuō*, I called the *shuō* in Mandarin as a ‘semi-complementizer’ in this paper. (see Chappell (2008) for a detailed documentation of the grammaticalization of complementizers from verbs of saying in Sinitic languages, including *waa6* in Cantonese, *koŋ⁵¹* in Taiwanese Southern Min.)

- (1) a. 張三 先 說
 Zhāngsān xiān shuō
 Zhāngsān first say
 ‘Zhāngsān says first.’
- b. 我 總是 覺得 說 生活-裡 缺-了 點兒 什麼
 Wǒ zǒngshì juéde shuō shēnghuó-lǐ quē-le diǎnr shénme
 I always think say life-in lack-PERF little something
 ‘I’ve always felt that there is something a little lacking in my life.’ (Fāng 2006: 109)
- c. 我 打算 說 到 英國 留學
 Wǒ dǎsuàn shuō dào Yīngguó liúxué
 I plan say arrive UK study.abroad
 ‘I planned to study abroad in the UK.’

* Thanks to Edith Aldridge, Shuheng Liu, Lu Shi, Chun Zheng, Karen Zagana, and audiences at LSA 2021 for helpful questions and comments on previous versions of this work. All errors are mine. Author: Jiahui Huang, University of Washington at Seattle (huangjh@uw.edu).

¹ Mandarin Chinese here refers to Beijing Mandarin specifically. Chappell (2008) reports that *shuō* is attested in Beijing Mandarin and Taiwanese Mandarin, but not in Standard Mandarin or *Pǔtōnghuà*. This paper focuses on the non-standard varieties of Mandarin in which the non-lexical usage of *shuō* has been more widely documented.

On the basis of the distributional pattern of *shuō* in (1b), there has been a tradition in analyzing *shuō* (and their counterparts in other Sinitic languages) as a complementizer (Yeung 2006; Hsieh & Sybesma 2008). However, this is by no means an undebatable conclusion, which has been challenged by some problematic cases where *shuō* cannot occur in clausal subjects, which is an environment that should be possible if *shuō* is a genuine complementizer (Paul 2014). Recent analyses make a further observation that *shuō* occurs not only in finite clauses (1b), but nonfinite clauses (1c), and defend the position of *shuō* as a complementizer by analyzing *shuō* as a verbal suffix at PF (N. Huang 2018), which can arguably cover some atypical patterns compared to genuine complementizers.

In this paper, I claim that *shuō* is a complementizer (C), and crucially that *shuō* introduces a specific type of CP called non-referential CPs, in the sense of de Cuba (2017). By looking into the patterns of *shuō* in more detail, I show that there exist environments in which *shuō* is consistently ruled out, which suggests that *shuō* only introduces a subtype of clauses (including finite and non-finite clauses), that are non-referential. This proposal then argues that *shuō* is not entirely optional (cf. Chappell 2008) and that the two variants of *shuō* that occur in both finite and nonfinite clauses can be unified (cf. Huang 2018).

The paper is organized as follows. Section 2 reviews recent proposals about the syntactic status of the *shuō*. Section 3 presents an overview of the data patterns in more detail, with the special focus on the environments where *shuō* cannot occur, which has not been discussed in previous analyses. Section 4 discusses the notion of (non)referential CPs in the sense of de Cuba (2017) and their attestation in Mandarin Chinese. Section 5 discusses the optionality of *shuō*. Section 6 proposes the syntactic position for *shuō* and its implications for the atypical patterns of *shuō* as a complementizer. Section 7 discusses the predictions that this proposal makes towards the counterparts of *shuō* in other Sinitic languages. Section 8 concludes the discussion.

2. Is *shuō* a complementizer? There has been a tradition in analyzing *shuō* as a complementizer, which can be evidenced by the syntactic tests such as aspect marking (2a) and constituency test (2b), borrowing Yeung's diagnostics for the Cantonese semi-complementizer *waa6* (Yeung 2006).

- (2) a. 張三 以為 說-(*過) 李四沒 來
 Zhāngsān yǐwéi shuō-guo Lǐsì méi lái
 Zhāngsān believe say-ASP Lǐsì NEG come
- b. 張三 想-著 說 今天 不-去 並且 說 明天 也 不-去
 Zhāngsān xiǎng-zhe shuō jīntiān bù-qù bìngqiě shuō míngtiān yě bú-qù
 Zhāngsān think-DUR say today NEG-go and say tomorrow also NEG-go
 'Zhāngsān is thinking that he is not going today and that he is not going tomorrow.'

First, the fact that the *shuō* following the verb *yǐwéi* 'believe' cannot be aspectually marked in (2a) suggests that *shuō* lost its verbal property and acts as a complementizer. Second, while the constituency test in (2b) cannot directly target the complementizer status of *shuō*, the fact that *shuō* can form a constituent with the following complement in conjunction falls out naturally if *shuō* is analyzed as a complementizer that introduces clausal complements.

Paul (2014), on the other hand, provides counterexamples that constitute serious challenges to the complementizer approach (this part of review is modified based on N. Huang's 2018 paper). First, *shuō* cannot appear in fronted embedded clauses in afterthought constructions (3a); Second, *shuō* cannot head a sentential subject (3b); Third, *shuō* cannot be stranded (3c); Finally,

there is often a pause after *shuō*, but crucially not between *shuō* and the preceding verb (cf. (2b)). N. Huang further observes that *shuō* cannot appear in the clausal complement of a noun (3d).

- (3) a. *(說) 生活-裡 缺-了 點-兒 什麼 我 總是 覺得
 Shuō shēnghuó-lǐ quē-le diǎn-r shénme wǒ zǒngshì juéde
 say life-in lack-PERF little what I always feel
 Intended: ‘That something is missing in life, I have always thought so.’
- b. *(說) 李四 不-會 出席 會議 使 大家 感到 很 驚訝
 Shuō Lǐsì bú-huì chūxí huìyì shǐ dàjiā gǎndào hěn jīngyà
 say Lǐsì NEG-will attend meeting make all feel very surprised
 Intended: ‘That Lisi won’t be attending the meeting surprised everyone.’
- c. * 生活-裡 缺-了 點-兒 什麼 我 總是 覺得 說
 Shēnghuó-lǐ quē-le diǎn-r shénme wǒ zǒngshì juéde shuō
 life-in lack-PERF little what I always feel say
 (Paul 2014: 98)
- d. 總裁 否認-了 (*說) 公司 將 裁員 的 謠言
 Zǒngcái fǒurèn-le say gōngsī jiāng cáiyuán de yáoyán
 CEO deny-PERF say company will lay.off DE rumour
 ‘The CEO denied rumours that the company will lay off workers.’
 (N.Huang 2018: 369)

This set of examples on the surface undermines the complementizer approach since a genuine complementizer like English that does not exhibit this restriction, as shown in the translation in (3a, 3b, 3d). N. Huang (2018) recently proposes an analysis in defense of *shuō* as a complementizer while aiming to explain the problematic cases in (3a-3d). First, he proposes that there exist two variants of *shuō* that occur in finite (e.g., complements introduced by verbs such as *yǐwéi* ‘believe’) and non-finite contexts (e.g. complements introduced by verbs such as *dǎsuàn* ‘plan’) respectively, as shown in (1b) and (1c), due to different sets of syntactic properties related to either of the contexts respectively. Second, he specifically analyses *shuō* as a verbal suffix at PF, following Bošković & Lasnik’s (2003) analysis of null C in English, which shows similar patterns in being disallowed in clausal subjects or in clausal complements of nouns. With *shuō* being a verbal suffix, it must immediately follow a verb when syntactic structures are linearized at PF. This proposal then accounts for the ungrammaticality in (3a) and (3b) with *shuō* at the sentential-initial position², because in these two cases there are no preceding verbs that the *shuō* can attach to, as well as the pro-sodic pattern that *shuō* forms a unit with the preceding verb. For (3d), Huang argues that there exist some functional elements in the embedded left periphery that prevents *shuō* from suffixing onto the preceding verb.

Two further implications of Huang’s analysis are (i) *shuō* can basically co-occur with all kinds of verbs as a suffix at PF so long as there is not intervening material between the verb and *shuō*, and (ii) the two variants of *shuō*s in finite and non-finite clauses are two different elements syntactically. This paper proposes an alternative approach by obtaining a more comprehensive picture of data surrounding *shuō*. It will be shown that *shuō* consistently fails to occur in a set of environments, which is left unexplained in the existing analyses. The renewed look at the pattern

² Note that Paul (2014) uses the fact that *shuō* cannot be stranded (3c) to argue against the complementizer analysis. However, as Huang (2018: 371) notices that ‘there appears to be a general constraint against standing complementizers (Abel 2003)’, even for English *that*. Thus, I set aside this restriction in the following discussion in this paper.

of *shuō* in turn points to a unified analysis of the two *shuō* that occur in finite and nonfinite contexts, which then explains the atypical patterns of *shuō* shown in (3) as a complementizer.

3. Data Patterns in more detail. Previous literature offers a large corpus documenting the environments where *shuō* can occur, which is conditioned by the degree of grammaticalization. Chappell (2008, 2017) offers an implicational hierarchy of verb classes that can co-occur with semi-complementizers in Sinitic languages (including *shuō* in Mandarin), which can be divided into five stages (4).

- (4) Stage I (Quotative constructions) → Stage II (Speech act verbs, e.g. ‘ask’, ‘tell’) → Stage III (Cognition verbs, e.g., ‘think’) → Stage IV (Perception, emotion, and stative verbs, e.g. ‘be worried’) → Stage V (Modal verbs, e.g. ‘be necessary’).

Chappell (2008, 2017) argues that the *shuō* in Beijing Mandarin has reached Stage IV, by which *shuō* can occur not only in quotative constructions (the least grammaticalized stage) and speech act contexts (5a), but crucially in complements introduced by cognition verbs such as the verb *juéde* ‘feel’ (1b), repeated in (5b), and perception, emotion, or stative verbs, such as *dānxīn* ‘be worried’ (5c), which renders the lexical usage of *shuō* ‘say’ highly unlikely. Chappell also notices that the *shuō* in Taiwanese Mandarin exhibits even a higher degree of grammaticalization by which it can co-occur even with modal verbs, signaling that it has reached Stage V, but no examples are provided³.

- (5) a. 他就告訴說他姑姑來了
Tā jiù gàosù shuō tā gūgu lái le
he then tell say he aunt come ASP
‘He then told that his aunt had come.’
- b. 我總是覺得說生活裡缺了點兒什麼
Wǒ zǒngshì juéde shuō shēnghuó-lǐ quē-le diǎnr shénme
I always feel say life-in lack-PERF little something
‘I’ve always felt that there is something a bit lacking in my life.’
- c. 他就會擔心說這個孩子以後會怎麼樣
Tā jiù huì dānxīn shuō zhè-ge háizi yǐhòu huì zěnmeyang
he then will be.worried say this-CL child afterwards will how
‘He is then worried that how this child will become afterwards.’

(Beijing Mandarin from Fāng (2006))

To summarize, *shuō* in Mandarin (including both Beijing and Taiwanese Mandarin) can co-occur with a wide range of verbs. These distributional patterns form the basis for the existing syntactic analyses (see Section 2). However, no special attention has been paid to the environments where *shuō* is consistently *not* allowed. The only place where the discussion of the environment where *shuō* cannot occur is one note made by Chappell in which she claims that when the grammaticalization of semi-complementizers reaches upon the Stage V (e.g., modal

³ In Chappell’s documentation of the *shuō* counterparts in other non-Mandarin Sinitic languages, Hakka *koŋ⁵¹* has reached Stage II, Cantonese *waa6* has reached Stage III, Taiwanese Southern Min *koŋ⁵¹* has reached Stage V, which is the most highly grammaticalized complementizer of all the Sinitic languages, as shown by the well-attested co-occurrence between *koŋ⁵¹* and modal verbs (i).

(i) I thâu-náu hó bô-it-tēng koŋ⁵¹ I toh gâu chò seng-lí.
3SG brains good NEG.necessary say 3SG then clever do business
‘Having good brains does not necessarily mean that you are good at business.’ (Chappell 2017)

complements), ‘expansion of verb classes taking the complementizer to other kinds of factive verbs may take place, with the potential of broadening the scope of further verb classes in an unrestricted manner (Chappell 2008: 62)’. In other words, factive verbs seem to behave as a cutting point along the scale of grammaticalization, after which the grammaticized complementizer can finally behave like a genuine complementizer. Interestingly, existing literature similarly shows the deficiency in detailed documentations of *shuō* occurring in factive complements.

This paper tries to fill the gap by exactly offering a more detailed look at the environments where *shuō* cannot occur. My data collection of Beijing Mandarin indicates that the *shuō*s occurring in complements introduced by factive verbs such as emotive factives like *gāoxìng* ‘be happy’/*shāngxīn* ‘be sad’ and *hòuhuǐ* ‘regret’/*fǎngǎn* ‘resent’⁴, are generally degraded compared to all the attested contexts (Stage I -V), which are all non-factive, as shown in (6) and (7).

- (6) 張三 很 高興/ 傷心 (??說) 李四 得了 第一名
 Zhāngsān hěn gāoxìng/shāngxīn (??shuō) Lǐsì dé-le dì-yī míng
 Zhāngsān very happy/sad say Lǐsì gain-PERF first.prize
 ‘ZS. is happy/sad (*that) LS gained the first prize.
- (7) 張三 很 後悔/ 反感 (??說) 他 沒 認真 讀書
 Zhāngsān hěn hòuhuǐ fǎngǎn (shuō) tā méi rènzhēn dúshū
 Zhāngsān very regret resent say he NEG diligently study
 ‘Zhangsan regrets/resents (*that) he didn’t study diligently.’

In other words, the occurrence of *shuō* is not random. In particular, *shuō* is incompatible with factive contexts and all the attested examples are those *shuō* that occur in non-factive contexts. It is unclear how the existing approach can capture this distributional discrepancy of *shuō*. In particular, N. Huang’s approach that analyzes *shuō* as a C and a verbal suffix predicts that *shuō* is allowed to appear so long as there exists a verb that it can attach to in LF. In Section 4, I will argue that the very fact *shuō* is ruled out in factive contexts suggests that *shuō* heads a specific type of CP, which in turns explains the problematic cases Paul (2014) reports in (3) and the two variants of *shuō* in finite and non-finite contexts (1b) and (1c).

4. Referentiality of CPs. Recall from Section 2 and Section 3 that the attested instances of complementizer *shuō* mainly occur in non-factive contexts, *shuō* occurring in factive contexts are generally degraded. de Cuba & Ürögdi (2009), de Cuba & MacDonald (2013) and de Cuba (2017) argue that factivity does not operate in syntax. Instead, it is the *referentiality* of the CP that distinguishes the so-called factive and non-factive clauses. The definitions of referential CPs and non-referential CPs, in the spirit of de Cuba (2017), are recounted in (8). The factive clauses in traditional sense are replaced by ‘Referential CPs’ in this view.

- (8) a. Referential CPs: an accepted (or pre-established) proposition in the existing discourse which has no illocutionary force.
 b. Non-referential CPs: a speech act which introduces a proposition (or an open question) which is not yet accepted (or pre-established) in the existing discourse.
 (de Cuba: 2017: 27)

⁴ Note that my data collection indicates that *shuō* is compatible with complements introduced by semifactive verbs such as ‘know’. I will leave the distinction between emotive factives and semifactives with regard to their behavior in co-occurring with *shuō* for future research.

One immediate question of adopting this proposal is that whether there exists independent evidence for the referentiality distinction in the clausal domain in Chinese. de Cuba & MacDonald (2013) present two diagnostics in distinguishing referential CPs from non-referential CPs, one is whether there exists a difference between the two subtypes of clause in requiring discourse context: Referential CPs requires discourse context and cannot be used in out-of-the-blue-contexts; By contrast, non-referential CPs can be used in out-of-the-blue-contexts. This contrast is attested in Mandarin Chinese (9) where a context in which a child stole a book from the library is provided (assuming this is a context where a teacher is meeting with the child's parents).

- (9) Context: 你的孩子今天偷了一本書
 Nǐ-de háizi jīntiān tōu-le yì-běn shū
 your child today steal-PERF one-CL book
 'Your child stole a book today.' (uttered by a teacher)
- a. 我很抱歉/遺憾我的孩子今天偷了一本書
 Wǒ hěn bàoqiàn/yíhán wǒ-de háizi jīntiān tōu-le yì-běn shū
 I very sorry/regret my child today steal-PERF one-CL book
 'I am sorry/regret that my child stole a book today.'
- b. 我覺得/以為我的孩子今天偷了一本書
 Wǒ juéde/yǐwéi wǒ-de háizi jīntiān tōu-le yì-běn shū
 I feel/believe my child today steal-PERF one-CL book
 'I am sorry/regret that my child stole a book today.' (uttered by the parents)

Under the context provided in (9), only (9a) instead of (9b) is a possible response to the utterance made by the teacher. (9a) instantiates the referential CPs while (9b) exemplifies the non-referential CPs. Since 'your child stole a book' is already accepted in the existing discourse, which makes the complements introduced by verbs such as *juéde* 'feel' or *yǐwéi* 'believe' that comes with a speech act introducing a proposition that is not accepted in the existing discourse inappropriate. There is no need to introduce a proposition that is already established in the context. This constitutes as the first piece evidence of a referentiality distinction in Mandarin Chinese.

The second diagnostic de Cuba & MacDonald uses to distinguish referential CPs from non-referential CPs is related to clausal anaphora: non-referential CPs allow both do-so placement and it-placement, the latter of which targets referential arguments only (10a); referential CPs by contrast only allows it-placement (10b). Similar examples can be identified in Mandarin Chinese (11) where *zhèmeyang* 'so' is available only under the non-factive/non-referential predicates (11a) but not factive/referential predicates (11b). This contrast between (11a) and (11b) similarly points to a referentiality distinction of clausal complements in Mandarin.

- (10) a. John supposed [that Bill had done it], and Mary supposed [it/so] too.
 b. John regretted [that Bill had done it], and Mary regretted [it/*so] too.
- (11) a. 我認為/以為他很聰明；他們也*(這樣)認為/以為
 Wo rènwéi/yǐwéi tā hěn cōngming; tāmen yě (zhèmeyang) rènwéi/yǐwéi
 I think/think he very smart they also so think think
 'I thought that he was smart; they also thought *(so).' (A. Li, 2013: 214)
- b. 我很後悔/討厭他做了這樣的事；
 Wǒ hěn hòuhuǐ/tǎoyàn tā zuò-le zhèyàng-de shì;
 I very regret/resent he do-PERF this-kind-de thing

他們 也 (*這麼樣) 後悔/ 討厭
 Tāmen yě (*zhèmeyang) hòuhuǐ/tǎoyàn
 they also so regret/ resent
 ‘I regret/resent that he did that kind of thing; they also regret/resent (*so).

Finally, the preposition *duì* ‘on’ in Mandarin provides language-internal evidence for the referentiality distinction of clausal complements. Huang, Li, Li (2009) treats *duì* ‘on’ as a diagnostic to distinguish adjectives from verbs: Many adjectives in Mandarin require prepositional phrases headed by *duì* ‘on’ followed by the preposed object when they are used transitively (12).

- (12) a. 這-個 工作 對 妳 很 合適
 Zhè-ge gōngzuò duì nǐ hěn héshì
 this-CL job on you very suitable
 ‘This job is suitable for you.’
 b. *這-個 工作 很 合適 你
 *Zhè-ge gōngzuò hěn héshì nǐ
 this-CL job very suitable you (Huang, Li, Li 2009: 22)

Interestingly, parallel to the two-way distinction based on referentiality of CPs shown in (9) and (11), only the referential predicates are allowed to co-occur with PPs headed by *duì* (13a); non-referential predicates are incompatible with PPs headed by *duì* (13b).

- (13) a. 張三 對 她 做-了 這樣-的 事 很 反感
 Zhāngsān duì tā zuò-le zhèyàng-de shì hěn fǎngǎn
 Zhāngsān on she do-PERF this.kind-MOD thing very resent
 ‘ZS resents that he had done this kind of thing.’
 b. *張三 對 李四 去-了 郵局 以為
 *Zhāngsān duì Lìsì qù le yóujú yǐwéi
 Zhāngsān on Lìsì go-PERF post.office believe

Combining the three diagnostics, clausal complements in Mandarin Chinese clearly exhibit a referentiality distinction. Thus, adopting this updated view of clausal complements, I restate the generalization reached in Section 3 using the referential/non-referential distinction in the CP domains: *shuō* only occurs in non-referential CPs but not referential CPs. To put it differently, *shuō* heads a subtype of clauses, namely non-referential CPs, which then explains the unacceptability of *shuō* that occurs in referential CPs (introduced by verbs such as emotive factive verbs like *gāoxìng* ‘be happy’/*shāngxīn* ‘be sad’ and *hòuhuǐ* ‘regret’/*fǎngǎn* ‘resent’).

5. Non-referential CP with or without *shuō*. It has been established in the literature that *shuō* is optional (Chappell 2008). The clausal complements with *shuō* and the one without *shuō* are argued to be interchangeable. A further scrutiny of the data pattern suggests that *shuō* is not optional, at least for complements introduced by non-speech act verbs (such as *dānxīn* ‘be worried’ and *dǎsuàn* ‘plan’, which make the *shuō* very unlikely to be interpreted in its quotative usage).

de Cuba & MacDonald (2013) report that the apparent optional complementizer *que* with *preguntar* in Spanish is in fact not optional at all. The presence of *que* suggests a presence of a speech-act operator that introduces a proposition into the common ground (which is not shared by the speakers). The minor yet significant difference between the one with the *que* and the one without is that the latter clause is an initial attempt on the part of the speakers to introduce the proposition to the common ground while the former clause instead is a non-initial attempt to

introduce the following proposition into the common ground. Put it differently, there are two subtypes of non-referential CPs that both introduce unsettled propositions to the common ground but in slightly different ways: the non-referential CPs without *que* is used in a context where the proposition is introduced by the speaker to the common ground for the first time; By contrast, the non-referential CPs with *que* can be uttered when the proposition introduced by the *que* should already be shared and accepted in the context from the point of view of the speaker but somehow has not yet been totally accepted in the common ground. Under a well-defined context, the difference between the two can be distinguished. (14) shows a similar case from *shuō* in Mandarin Chinese, with the verb *dǎsuàn* ‘plan’ as an illustration.

(14) Context: Lǐsì is sharing a travel plan with his friends, who thought Lǐsì will have to stay in town because he needs to prepare for an important exam and plus there will still be many travel restrictions due to the pandemic.

- a. 李四 打算 說 年底 到 歐洲 度假
 Lǐsì dǎsuàn shuō niándǐ dào Ōuzhōu dùjià
 Lǐsì plan say year.end to Europe go.on.vacation
 ‘Lisi plans to take a vacation in Europe at the end of the year.’
- b. 李四 打算 年底 到 歐洲 度假
 Lǐsì dǎsuàn niándǐ dào Ōuzhōu dùjià
 Lǐsì plan year.end to Europe go.on.vacation
 ‘Lisi plans to take a vacation in Europe at the end of the year.’

While (14a) and (14b) are both possible independently, under the context described in (14), only the instance without the *shuō* is appropriate. This is expected if the variant with *shuō* in (14a) reflects a ‘non-initial’ attempt on the part of the speaker to (re)introduce the following proposition (which should already be accepted in the common ground from the point of view of the speaker), which is incompatible with the context in (14) where the following proposition ‘going on vacation in Europe’ has never part of the common ground (in fact, the context is completely opposite to what is accepted in the common ground (that Lǐsì is staying in town)). This suggests that the presence or absence of *shuō* conveys two different interpretations, contra previous analyses that analyze *shuō* as a completely optional element in introducing clausal complements.

6. Syntactic positions of *shuō*. Having established that *shuō* heads a subtype of non-referential CPs, which is not entirely interchangeable with the clausal complement without *shuō*, this section provides an analysis of the syntactic positions of *shuō*.

There are at least two major approaches for referential CPs (which subsume factive CPs in the traditional sense). The first one comes from de Cuba’s own truncation approach in which referential CPs are those with an impoverished left periphery (15b) while non-referential CPs with a full periphery (15a) (see Haegeman 2006 also).

- (15) a. V [cP [CP]]
 b. V [CP] (de Cuba & Ürögdi 2009: 37)

Haegeman & Ürögdi (2010) instead argue that referential clauses are the result of an operator movement (which derives the referentiality in the clausal domain) (16) in the left periphery compared to the lack of it in non-referential clauses.

- (16) [CP OP_i C...]_{FP} t_i [TP...]] (Haegeman & Ürögdi, 2010: 115)

Given the unsettled status of referential clauses, I remain non-committal about the final implementation of these clauses⁵. Regardless, both lines of approach agree that there is a referentiality distinction in the CP domains, which I argue is the key to the understanding some of the unexpected patterns of *shuō* in Mandarin as a complementizer. Specifically, following Melvold's (1991) proposal that referential clauses are encoded with [+definite] features whereas non-referential CPs with [-definite] feature⁶. I argue that *shuō* heads a specific type of non-referential CPs with a [-definite] feature (17a), whereas the referential CPs are headed by an obligatorily null C with a [+definite] feature (17b), hence the unacceptability of *shuō* in referential clauses.

- (17) a. [V...[C_[-DEF] *shuō*...]]⁷
 b. [V...[C_[+DEF] \emptyset ...]]

With *shuō* heading a [-def] clausal complement while the referential CP headed by an obligatorily null C with a [+def] feature, some patterns atypical of *shuō* as a complementizer fall out naturally. Recall that Paul (2014) presents examples that (i) embedded clauses headed by *shuō* cannot be fronted in afterthought constructions (3a) (ii) *shuō* cannot appear in clausal subjects (3b), unlike typical complementizers such as English *that*. More examples are shown in (18a) and (18b), respectively.

- (18) a. (*說) 李四 會 來, 我 以為
 (Shuō) Lǐsì huì lái, Wǒ yǐwéi
 say Lǐsì will come I believe
 Intended Reading: 'That Lǐsì will come, I believe.'
 b. (*說) 李四 退學 讓 家裡人 很 擔心
 (Shuō) Lǐsì tuìxué ràng jiālǐrén hěn dānxīn
 say Lǐsì drop.out make home.people very worried
 '(*That) Lǐsì dropped out worries his family a lot.'

The impossibility of *shuō* occurs in (18a) and (18b) is accounted for in the current proposal by the [-def] feature in the sentence-initial position, because Mandarin generally disallows

⁵ The fact that referential clauses introduced by factive verbs such as *hòuhuī* 'regret' permits Main Clause Phenomenon (e.g. inner topics) in Mandarin (i) gives a plus to the operator movement approach and a minus to the truncation approach. It is unclear how a truncation approach accounts for the acceptability of inner topics within the factive clauses if the left periphery are truncated.

(i) 我很遺憾那本書你沒讀過
 Wǒ hěn yíhàn nà-běn shū nǐ méi dú-guo
 I very regret that-CL book you NEG read-ASP
 'I regretted that you haven't read the book.' (Liao & Kao 2017)

⁶ Melvold's original approach uses the traditional factive/non-factive distinction. Here I change the factivity distinction to a referentiality distinction, following de Cuba (2017).

⁷ Recall in Section 5, adopting de Cuba & MacDonald's analysis of *preguntar* 'ask' with *que* in Spanish, I present evidence that the absence or presence of *shuō* in similarly not optional. While they both introduce new and unsettled propositions (therefore non-referential), but the propositions introduced are slightly different, namely, they introduce proposition as either the first-attempt or the non-initial attempt from the perspective of the speaker, respectively. The difference between the two can be captured along the line of cartographic approach by which *shuō*, similar to *que* in Spanish, is analyzed as the head of an extended projection in the CP licensed by a speech-act operator (de Cuba & MacDonald 2013: 137), which is absent in the non-referential CPs without *shuō*.

indefinite subjects in clause-initial positions, which is a well-discussed fact in the syntax of Chinese (Huang, Li & Li, 2009), as shown in (19).

- (19) *三-個 學生 吃-了 蛋糕
 Sān-ge xuéshēng chī-le dàngāo
 three-CL student eat-PERF cake
 Intended: ‘Three students ate the cake.’ (Huang, Li, Li: 2009: 288)

It can be predicted that (18a) and (18b) without the sentential-initial *shuō* are acceptable with a [+def] reading, which cannot be used without discourse context. This is borne out in (20) in which a context where the participant of the embedded clause has not been established in the discourse. Only the clausal complements in its postverbal position (20b) can be used in this context, since the embedded proposition needs to be introduced to the common ground for the first time by the speakers. (20a) with the fronted clause requires a context where Lisi is already shared in the common ground, which is incompatible with the out-of-the-blue context in (20).

- (20) Context: Assuming Lisi has never been brought up in the discourse context (no one knows who Lisi is).
 a. #李四 會 來， 我 以為
 Lǐsì huì lái, Wǒ yǐwéi
 Lǐsì will come I believe
 ‘That Lǐsì will come, I believe.’
 b. 我 以為 李四 會 來
 Wǒ yǐwéi Lǐsì huì lái
 I believe Lǐsì will come
 ‘I believe Lǐsì will come.’

This proposal then unifies the two variants of *shuō* that occur both in finite and nonfinite clauses, as shown in (1b) and (1c), repeated in (21), because they both introduce non-referential CPs (introduced by non-factive verbs). Recent discussions on control (e.g. Landau, 2015) propose that there also exists a left periphery that represents the context of speech for attitude complements (e.g. *dǎsuàn* ‘plan’) in C, or a ‘logophoric center’ in the sense of Bianchi (2003), which are available for both finite and attitude nonfinite contexts (e.g. external logophoric center for finite clauses while internal logophoric center for nonfinite clauses). I argue that *shuō* exemplifies this C head for both for (21a) and (21b).

- (21) a. 我 總是 覺得 說 生活-裡 缺-了 點-兒 什麼
 Wǒ zǒngshì juéde shuō shēnghuó-lǐ quē-le diǎn-er shénme
 I always feel say life-in lack-PERF little something
 ‘I’ve always felt that there is something a little lacking in my life.’ (Fāng, 2006: 109)
 b. 我 打算 說 到 英國 留學
 Wǒ dǎsuàn shuō dào Yīngguó liúxué
 I plan say arrive UK study.abroad
 ‘I planned to study broad in the UK.’

7. Predictions for other Sinitic languages. Recall this current analysis proposes that the fact that *shuō* heads a non-referential CP accounts for the atypical patterns of *shuō* being unable to occur in sentential initial position, due to an independent restriction in syntax of Chinese where indefinite subjects are disallowed in sentence-initial position. Given that *shuō* as a

complementizer is a result of grammaticalization, there is no doubt that different varieties (that have similar type of processes of grammaticalization from verbs of saying to complementizers) exhibit different degrees of compatibility with different subtypes of verbs. Crucially, it can be predicted that if there exists a variety whose grammaticalized complementizers can occur in the sentence-initial position (a stage that Mandarin *shuō* has not yet reached), this very C should be able to introduce not only non-referential clauses, but crucially referential clauses (including factive contexts). This prediction is borne out by the counterpart of *shuō* in Taiwanese Southern Min, *koy⁵¹*, which is documented to be the most grammaticalized complementizer (compared to other Sinitic languages, see footnote 1)⁸. Sentence-initial *koy⁵¹* that introduces clausal subjects or main clauses is attested in Taiwanese Southern Min (22a) and (22b)^{9,10}. Interestingly, this *koy⁵¹* is reported to be compatible with referential clauses, as exemplified by the emotive factives *phai2-sek3* (歹勢) ‘feel embarrassed’, as shown in (22c).

- (22) a. 講 彼-個 流 目屎 落來-的 土-面
koy⁵¹ hit4-e5 lau5 bak8-sai2 loh8lai5-e5 thoo5-bin7
 say that-CL flow tear fall.come-de mud-face
 看 講 有 筍仔 浮 出來
khunann3 koy⁵¹ u7 sun2a2p phu5 tshut4-lai5
 see say have bamboo shoot come-out
 ‘That the ground is dropped with tears germinates bamboo shoots. (Tseng 2008: 66)
 (Note that there are two *koy⁵¹* in this sentence, the first *koy⁵¹* introduces a clausal subject (which is disallowed for *shuō*) while the second *koy⁵¹* introduces a complement clause).
- b. 講 有 職-款 人 啦!
koy⁵¹ u7 tsit4-khuan2 lang5 lah8
 say have this-kind people SFP
 ‘How come there is such kind of people!’ (Hsieh & Sybesma 2008: 366)
- c. 啊 伊 亦 歹勢 講
A2 i1 ia7 phai-se3 koy⁵¹
 PT 3SG also feel.embarrassed. say

⁸ Another candidate from Sinitic languages could be *da* (ㄉㄚ) in Jieyang dialect, another Southern Min language, which has been argued to be even more grammaticalized than *koy⁵¹* in Taiwanese Southern Min (Huáng 2016: 689) in that it has become obligatory for introducing complement clauses for a wide range of verbs. Also, it was reported that *da* (ㄉㄚ) is compatible with factive verbs such as ‘realize’. If *da* (ㄉㄚ) is more grammaticalized than *koy⁵¹*, it can be predicted that *da* (ㄉㄚ) is allowed to occur in sentence-initial positions. More data are needed to confirm this prediction. (See Xu and Matthews (2007) for discussions of *da* (ㄉㄚ) in Chaozhao dialect, yet another variety of Southern Min).

⁹ The sentence-initial *koy⁵¹* usage has been reported in different sources (e.g. Cheng 1997; Hsieh & Sybesma 2008; Tseng, 2008; Lien 2020, and among many others), which has received different analyses. Hsieh & Sybesma (2008) and Tseng (2008) suggest that *koy⁵¹* is a complementizer that introduces main clauses and clausal subjects, while Cheng (1997) analyzes it as an adverb with a hearsay interpretation, and Lien (2020) categorizes *koy⁵¹* as a topic marker. While more investigations are certainly needed, based on the interpretation given in (22a) with arguably no ‘hearsay’ interpretation, I take it as supporting the complementizer view, which can introduce a clausal subject (contra *shuō* in Beijing Mandarin), as shown in (22a), or even a main clause, as shown in (22b).

¹⁰ Note that *koy⁵¹* can also occur in the sentence-final position, which makes it similar to many of the sentence-final particles in Chinese languages. It expresses the subjectivity on the part of the speaker. It can also be used in imperative (Lien 2020). (See Simpson & Wu (2002) for a formal analysis of the sentence-final *koy⁵¹* as a head-initial complementizer.)

- d. 安呢 共 趕 伊 走 安-呢 啦
 an2-ni1 ka7 kuann2 i1 tsau2 an2-ni1 lah8
 so OBL expel 3SG go so SFP
 ‘He is also embarrassed that he expels him.’ (Tseng 2008: 99)

8. Conclusions. In this paper I have claimed that *shuō* is a semi-complementizer that introduces non-referential CPs based on a revisit of the distributional patterns between *shuō* and different clausal complements. The atypical patterns observed in the previous literature can be explained by the non-referential/indefinite feature of the clauses headed by *shuō*. This paper also provides predictions for the analysis of other semi-complementizers that are grammaticalized from verbs of saying in other Sinitic languages. It remains to be seen how such an account can be extended to the rich cross-linguistic patterns of (semi)-complementizers of the similar kinds. This issue is however outside the scope of this paper and must be left to further investigation.

References

- Bianchi, Valentina. 2003. On finiteness as logophoric anchoring. In Jacqueline Guéron & Liliane Tasmowski (eds.), *Temps et point de vue* [Tense and point of view], 213–246. Paris: Université Paris X.
- Bošković, Željko & Howard Lasnik. 2003. On the distribution of null complementizers. *Linguistic Inquiry* 34(4). 527–546. <https://doi.org/10.1162/002438903322520142>.
- Burukina, Irina. 2021. Arguments introduced by C heads. Paper presented at the 2021 Annual Meeting of the Linguistic Society of America.
- Chappell, Hilary. 2008. Variation in the grammaticalization of complementizers from verba dicendi in Sinitic languages. *Linguistic Typology* 12(1). 45–98. <https://doi.org/10.1515/LITY.2008.032>.
- Chappell, Hilary. 2017. Say-complementizers: *shuō* 說, *waa3* 話, *kong1* 講. In Rint Sybesma (ed.), *Encyclopedia of Chinese language and linguistics*. London: Brill. https://doi.org/10.1163/2210-7363_ecll_COM_00000098.
- Cheng, Robert. 1997. The constructional markers *shuo* ‘say’ and *kan* ‘see’ in Taiwanese and Taiwan Mandarin. In *Taiwanese and Mandarin structures and their developmental trends in Taiwan II: Contacts between Taiwanese and Mandarin and restructuring of their synonyms*. 105–131. Taipei: Yuan-Liou Publishing.
- de Cuba, Carlos & Barbara Ürögdi. 2009. Eliminating factivity from syntax: Sentential complements in Hungarian. In Marcel den Dikken & Robert M. Vago (eds.), *Approaches to Hungarian* (Volume 11: Papers from the 2007 New York Conference), 29–63. Amsterdam: John Benjamins. <https://doi.org/10.1075/atoh.11>.
- de Cuba, Carlos & MacDonald, Jonathan. 2013. Referentiality in Spanish CPs. In Victoria Camacho-Taboada, Ángel L. Jiménez-Fernández, Javier Martín-González & Mariano Reyes-Tejedor (eds.), *Information Structure and Agreement*, 117–140. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.197>.
- de Cuba, Carlos. 2017. Noun complement clauses as referential modifiers. *Glossa: A Journal of General Linguistics* 2(1). 3. <https://doi.org/10.5334/gjgl.53>.

- Fāng, Méi. 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.
- Haegeman, Liliane & Barbara Ürögdi. 2010. Referential CPs and DPs: An operator movement account. *Theoretical Linguistics* 36(2-3). 111–152. <https://doi.org/10.1515/thli.2010.008>.
- Haegeman, Liliane. 2006. Conditionals, factives and the left periphery. *Lingua* 116(10). 1651–1669. <https://doi.org/10.1016/j.lingua.2005.03.014>.
- Bernd, Heine & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Hock, Hans Heinrich. 1982. The Sanskrit quotative: A historical and comparative study. *Studies in the Linguistic Sciences* 12(2). 39–85.
- Heish, Feng-fan & Rint Sybesma. 2009. 生成语法理论与汉语语气词研究 (Shēngchéng yǔfǎ lǐlùn hé Hànyǔ yǔqì yánjiūcí yánjiū). In Yáng Shěn & Shènglì Féng (eds.), 当代语言学理论和汉语研究 Dāngdài yǔyánxué lǐlùn hé hànyǔ yánjiū [Contemporary linguistic theories and related studies on Chinese], 364–374. Peking: Shāngwù.
- Huang, Nick. 2018. Control complements in Mandarin Chinese: Implications for restructuring and the Chinese finiteness debate. *Journal of East Asian Linguistics* 27(4). 347–376. <https://doi.org/10.1007/s10831-018-9185-1>.
- Huang, Cheng-The James & Yen-hui Audrey Li & Yafei Li. 2009. *The syntax of Chinese*. Cambridge: Cambridge University.
- Huang, Yanxuan. 2016. 揭陽方言言說動詞“咁”的語法化 (Jiēyáng fāngyán yánshuō dòngcí da de yǔfǎhuà) [Grammaticalization of the SAY verb *da* (咁) in the Jieyang dialect]. *中国語文 Zhōngguō Yǔwén* 6. 686–694.
- Li, Yen-hui Audrey. 2013. Case, tense and clauses. In Guangshun Cao, Hilary Chappell, Redouane Djamouri & Theklid Weilbusch (eds.), *Breaking down the barriers: Interdisciplinary studies in Chinese linguistics and beyond*, 205–325. Taipei: Institute of Linguistics, Academia Sinica.
- Liao, Wei-wen Roger & Kao, Grant. 2017. Argumenthood, MCP, and Chinese topicalization. Presentation at International Association of Chinese Linguistics 25. Hungarian Academy of Sciences and Eötvös Loránd University.
- Lien, Chinfa. 2020. Kong2 as a verb for saying ‘on the move’ in Taiwanese Southern Min. In Chinfa Lin & Alain Peyraube (eds.), *Diachronic [erspectives and synchronic variation in Southern Min*, 197–216. London: Routledge. <https://doi.org/10.4324/9780429316517>.
- Lin, Huáyǒng & Má, Zhé. 2007. 廉江方言言說義動詞“講”的語法化 (Liánjiāng fāngyán yánshuōyì dòngcí jiǎng de yǔfǎhuà) [Grammaticalization of the speech verb (講) in the Liánjiāng dialect of Guangdong Province]. *中国語文 Zhōngguō Yǔwén* 7. 151–158.
- Melvold, Janie. 1991. Factivity and definiteness. *MIT Working Papers in Linguistics* 15. 97–117.
- 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. <https://doi.org/10.1111/stul.12020>.
- Simpson, Andrew & Zoe Wu. 2002. IP-raising, tone sandhi and the creation of S-final particles: Evidence for Cyclic Spell-Out. *Journal of East Asian Linguistics* 11. 67–99. <https://doi.org/10.1023/A:1013710111615>.

- Xu, Huiling & Stephen Matthews. 2007. 從動詞到子句標記：潮州方言和台灣閩南語動詞“說”和“看”的虛化過程 (Cóng dòngcí dào zǐjù biāoji: Cháozhōu fāngyán he Táiwān Mǐnnányǔ ‘shuō’ hé ‘kàn’ de xūhuà guòchéng) [From verb to complementizer: the grammaticalization process for verbs ‘say’ and ‘see’ in the Chaozhou dialect and Taiwanese Southern Min]. 中國語文研究 *Zhōngguō Yǔwén Yánjiū* 23. 61–72.
- Yeung, Ka-Wai. 2006. On the status of the complementizer waa6 in Cantonese. *Taiwan Journal of Linguistics* 4(1). 1-48.