



**HAL**  
open science

# Variation in the grammaticalization of complementizers from verba dicendi in Sinitic languages

Hilary Chappell

## ► To cite this version:

Hilary Chappell. Variation in the grammaticalization of complementizers from verba dicendi in Sinitic languages. *Linguistic Typology*, 2008, 12 (1), <10.1515/lity.2008.032>. <hal-03929503>

**HAL Id: hal-03929503**

**<https://hal.science/hal-03929503v1>**

Submitted on 8 Jan 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Variation in the grammaticalization of complementizers from *verba dicendi* in Sinitic languages

Hilary Chappell

## *Abstract*

This article presents the first study of crosslinguistic variation in grammaticalization processes affecting *verba dicendi* in Sinitic languages, making use of discourse data in the comparison. Its principal objective is to undertake a quantitative analysis of SAY verbs used as complementizers in Southern Min and Cantonese in order to pinpoint the stages reached in grammaticalization. In particular, it will be argued that Southern Min is more advanced in its development of a complementizer than is Cantonese.

A second and more general objective is to provide an overview of grammaticalized functions of SAY verbs in an enlarged sample of 10 Sinitic languages for which discourse data is also available, including Hakka, Xiang, Wu, Jin, Gan and 3 varieties of Mandarin.

*Keywords:* syntax, complementizer, SAY verbs, *verba dicendi*, grammaticalization, Sinitic, Chinese, Taiwanese Southern Min, Cantonese Yue, Mandarin, Hakka, quantified discourse analysis, dialectal variation

## 1. Introduction

### 1.1. The main issues

The grammaticalization of SAY verbs, or *verba dicendi*, into complementizers, subordinating conjunctions and other grammatical functions has been documented in detail for many languages in the African, South and Southeast Asian regions and also, interestingly, for creoles (see Güldemann 2001, Hock 1982, Lord 1993, Plag 1992, Saxena 1988 *inter alia*).<sup>1</sup> However, this category is not at all well-attested in the study of the Sinitic or Chinese languages. In fact, for two of the more extensively analysed ones, Standard Mandarin (*pǔtōnghuà*) and Cantonese, it has generally been held that there are no subordinating conjunctions similar to *that* in English, *que* in French, or *ti* in Shona which serve to introduce indirect speech or an embedded clause.<sup>2</sup>

The present obstacle to any ultimate verification of such claims for Sinitic languages is that there has been little research to date specifically targeting the grammatical category of the complementizer in the field of Chinese linguistics, with the consequence that descriptions are difficult to come by, if at all. It is only recently that this phenomenon has been observed and described for Beijing Mandarin (Fang 2006, §5.2), while the fact that analyses are available for Taiwanese Mandarin is due to the diligence of researchers into Southern Min as spoken in Taiwan, who noted the striking structural parallels between the use of SAY verbs in the two languages spoken on this island (Leok 1982, Cheng 1991).

Adding to the problem is the fact that the category of the complementizer is less likely to be described in grammars that are based purely on elicitation, or in ones that set out to codify a more formal register of the language (see also Noonan 1985: 137-138). As in the case of many other fascinating kinds of linguistic data, sufficient quantities of spoken discourse materials are needed to discover newly emerging grammatical phenomena such as the complementizer. Hence, the development of these special markers of complementation in several Sinitic languages presents an intriguing case.

For Southern Min and Cantonese, a quantitative analysis is presented on the complementizer usage, based on texts collected in Taipei and Hong Kong respectively. The data suggest that Southern Min is further along the pathway of grammaticalization for this function than is Cantonese, according to the series of hypothetical stages set up and supported in terms of other crosslinguistic studies.

In addition to this, a brief comparison is given of variation in the grammaticalization of SAY verbs across a further six branches of Sinitic languages for which discourse data is available, either from my own database of transcriptions or from published reference materials. Consequently, the analysis in this part of the study is expanded to regional varieties of Mandarin, Jin, Hakka, Xiang, Wu and Gan. It is contended that both Taiwanese Mandarin and Beijing Mandarin appear to have reached a relatively high degree of grammaticalization for their respective complementizers, similar to Southern Min, whereas Sixian Hakka is only at an incipient stage in developing a semi-complementizer, used mainly with other speech act verbs.

---

<sup>1</sup> A note on terminology: ‘SAY’ stands for the semantic field of verbs of saying, crosslinguistically viewed, regardless of the actual phonetic form or position in a lexical field, while ‘say’ is used as the gloss for particular instances of this type of verb.

<sup>2</sup> As proponents of this view, see Ramsey (1987: 86), Paris (1996) and Cheng (1991) on Standard Mandarin, Matthews and Yip (1994: 308) on Cantonese.

Discourse data on Jin, Xiang, Gan and Wu do not provide any evidence of complementizer usage. In this section, it is also argued that the use of a complementizer in the formal variety of language known as Standard Mandarin or *pǔtōnghuà* is clearly disfavoured in prescriptive grammars for reasons of register and style.

The exact layout of the presentation takes the following form: after this introduction in §1.1, the background to crosslinguistic studies on complementizers and their lexical sources is given in §1.2. §2 discusses the category of the complementizer from both syntactic and semantic points of view, while §3 provides a description of hypothesized stages for the pathway of grammaticalization into a complementizer from the perspectives of both general typological studies and Sinitic languages.

The main quantitative analysis regarding the comparison of complementizer development in Taiwanese Southern Min for *kóng* 講 and Hong Kong Cantonese Yue *wa*<sup>6</sup> 話 ‘say’ is presented in §4. In §5, a brief comparison is made of a further eight languages or varieties belonging to six different branches of Sinitic: the case for Standard Mandarin is compared with two regional non-standard varieties of Northern Chinese, namely Beijing Mandarin and Taiwanese Mandarin. Finally, the complementation issue for Sixian Hakka, Changsha Xiang, Nanchang Gan, Shanghainese Wu and Huojia Jin dialects is briefly addressed.

## 1.2. *Verba dicendi* and the development of complementizers

According to cross-linguistic and areal studies, at least five main sources can be identified for different kinds of complementizers:<sup>3</sup>

- (i) nouns such as ‘thing’, ‘fact’ or ‘place’, e.g. Korean *kɛs* ‘thing’; Japanese *koto* ‘thing’
- (ii) demonstrative, interrogative and relative pronouns, e.g. Faroese *tadh*; German *daß*, Georgian *raytamca*
- (iii) dative, allative and locative case markers or prepositions, e.g. Maori *ki* LOC/DAT; English ‘to’
- (iv) SAY verbs, e.g.; Chantyal *bfi*, Nepali *bhan*, Yoruba *kpé*, Vietnamese *rǎng*
- (v) simulative verbs meaning ‘resemble’ or ‘be like’; simulative manner adverbials and deictics, e.g. Idoma *bɛ* ‘resemble’; *ítí* ‘thus’ in Sanskrit; *tí* in Shona ‘be/do thus’

---

<sup>3</sup> This list has been established on the lexical sources and data presented in Lord (1976, 1993), Güldemann (2001), Heine et al (1991: 216-217; 246-247), Heine and Kuteva (2002: 37; 211-212; 254; 257-258; 261-269; 295 ; 2007: chapter 5 & §6.4) , Hock (1982), Hopper and Traugott (1993: 180-184), Noonan (1985: 47-48) and Ransom (1988). It also includes the following references on specific languages in the Asian and African regions regarding examples of complementizers < SAY: Amberber (1995), Bisang (1992), Ebert (1986, 1991), Frajzyngier (1996), Li (1988), Matisoff (1991), Saxena (1988, 1995). For Austronesian languages, Klammer (2000), François (2002) and Brill (2002) provide details of similar developments while, interestingly, this pathway can also be found in Chemehuevi (Uto-Aztecan) (Munro 1978).

Note that some of the sources could turn out to be related by one and the same grammaticalization pathway, for example, the case of nouns meaning ‘thing’, ‘matter’ and demonstrative and indefinite pronouns. Manner deictics could perhaps be classed in with demonstratives. There are also several verb classes associated particularly with mimesis including inchoatives and copulas, verbs of motion such as ‘go’ and DO verbs which are treated in Güldemann (2001:§4.2.2).

In the case of SAY verbs, the topic of this analysis, widespread occurrence of this grammaticalization pathway has been documented for language families located particularly in Africa, South and Southeast Asia:

SOUTHEAST AND EAST ASIA: Tibeto-Burman; Tai-Kadai; Hmong-Mien; Austroasiatic;  
Mongolian; Austronesian

SOUTH ASIA: Indo-Iranian; Dravidian

AFRICA: Amharic (Semitic, Afroasiatic); Chadic (Afroasiatic); Kwa (Niger-Congo)

Viewed therefore from an areal perspective, Sinitic languages appear to neatly fit in with the surrounding language families of South, Southeast and Northern Asia, namely, for those which show the emergence of newly grammaticalized uses of their *verba dicendi*, including the complementizer usage and different kinds of subordinating conjunctions. Note, however, for the South Asian and African areas, the source is often a manner adverbial ‘thus’ or even a verb ‘to be like, resemble’, described in detail in Hock (1982) and Güldemann (2001, 2002).

The SAY grammaticalization pathway for complementizers is thus spread across several continents and many language families, similarly being the source in certain Caucasian languages (Frajzyngier 1991: 226), but also importantly in many creoles with different lexifier languages (see Crowley 1989, Plag 1992).

Here is an initial example from Southern Min to set the scene, illustrating both the lexical verb use of *kóng* ‘to say’ in the first intonation unit and then the grammaticalized complementizer use in the second:

- |     |                     |              |                 |                     |
|-----|---------------------|--------------|-----------------|---------------------|
| (1) | 彼                   | 講            | 啊 , -           |                     |
|     | <i>he</i>           | <i>kóng</i>  | <i>a</i>        |                     |
|     | that <sub>3SG</sub> | say          | PRT             |                     |
| <Q  | 這                   | 事            | 感覺              | 講                   |
| <Q  | <i>che</i>          | <i>sū</i>    | <i>kám-kak</i>  | <i>óng /kóng/</i>   |
|     | this                | matter       | feel            | SAY <sub>THAT</sub> |
|     | 這                   | 足            | 危險              | Q>                  |
|     | <i>che</i>          | <i>chiok</i> | <i>gúi-hiàm</i> | Q>                  |
|     | this                | very         | dangerous       |                     |

‘He (Toyotomi) said, “I **feel that** the situation is very dangerous”.’  
(*Japanese tales* 1081-1083).

With regard to other grammaticalized uses of SAY verbs, the list provided below shows evidence of the high functional load that grammaticalized SAY verbs may carry:<sup>4</sup>

---

<sup>4</sup> This list is synthesized from the discussions of grammaticalization pathways in Lord (1976: 179), Hock (1982), Saxena (1988), Heine and Kuteva (2005) and my own research on Sinitic languages.

1. Quotative marker or complementizer; marker of embedded questions
2. Conditional conjunction
3. Reason or purpose conjunction
4. Causal conjunction
5. Hearsay marker of evidentiality
6. Marker used with onomatopoeic words
7. Comparative marker
8. Mirative marker
9. Listing constructions
10. Topic marker
11. Clause-final discourse particle expressing self-evident assertions, warnings and echo questions (different construction types)
12. Clause-initial discourse marker for exclamations

The first five of these types of reanalysis are possible in Taiwanese Southern Min and other Sinitic languages, including Cantonese and certain colloquial varieties of Mandarin. However, the hearsay evidential usage cannot yet be described as fully grammaticalized in Sinitic languages.<sup>5</sup>

The topic marker use and discourse marker functions in clause-initial and clause-final positions, the last three in the list above, are also found widely employed in Sinitic languages, and are functions that have not received adequate attention to date. The remaining uses are, however, not revealed in any of the discourse data examined: these are the onomatopoeic, comparative, mirative and listing uses, although there appears to be a negative comparative usage in Changsha Xiang of ‘to be not like’.<sup>6</sup>

In this analysis, we are going to concentrate on just the complementizer pathway, whose stages of grammaticalization are first summarized below:

- SAY
- > quotative function as main verb of principal clause introducing direct or indirect speech
  - > quotative function as V<sub>2</sub> in a serial verb construction where V<sub>1</sub> = speech act verb
  - > complementizer upon extension of V<sub>1</sub> to other verb classes including cognition, perception, emotion and modal

Figure 1: Pathway for the grammaticalization of SAY verbs into complementizers in Sinitic languages

<sup>5</sup> For example, the hearsay evidential use in the Southern Min phrase *kóng sì* say-be ‘it is said that’, ‘people say’ is also found in the conversational data, but, as foreshadowed, I treat it as one of the lexical interpretations of *kóng* ‘say’.

<sup>6</sup> SAY thus presents a case of polygrammaticalization in Sinitic showing several pathways of grammaticalization, defined by the structural position of this verb in the clause and other combinatorial factors. For example, to merely give an informal description: when subjectless in the initial position of turn-construction units of conversation, SAY may develop into a topic marker (SAY X > ‘talking about X, ...’), a pan-Sinitic use. The segment, ‘2SG SAY’, in clause-initial position may come to serve a rhetorical function over time, forming the basis of a new grammatical construction type coding exclamations (2SG SAY + CLAUSE > ‘Would you believe + CLAUSE’; Sin’ on Hakka, Nanchang Gan) while at the end of a clause, SAY assumes other kinds of irrealis modal usages, such as warnings in Southern Min (CLAUSE + SAY > ‘I’m telling you, you just try and do VERB’). In yet other developments, SAY attaches to an array of conjunctions to form new compound markers of the conditional, the purposive, concessive and the causal (Southern Min, Cantonese, Mandarin, Shanghainese, Jin). These are dealt with in Chappell (in prep., chapter 7).

## 2. Syntax of complementation

The term ‘complementizer’ refers to a grammatical category, including the type known as the subordinate conjunction, which introduces a second, often, dependent clause in the formation of a complex sentence. I use the term ‘complementizer’ in this analysis for the sake of neutrality, since, in many languages, these markers are not necessarily subordinating conjunctions.<sup>7</sup> A large array of constructional types involved in complementation is detailed in Noonan (1985) and Güldemann (2001).

### 2.1. Semantic and structural characteristics of complementizers and complementation

A hallmark in the use of complementizers is that in general they appear to be semantically governed by the verb in the principal clause in many languages. In studies of this complex construction type in functional grammar, co-occurrence possibilities have been well-described, in the case of English, by Bolinger (1972), Wierzbicka (1988) and Dixon (1991) for *that, for...to*, interrogative pronouns and *POSS...ing* complementation, and crosslinguistically by Noonan (1985) and Horie (2001). Still earlier, Bresnan (1970) made similar observations on dependency, arguing that complementizers in English were not semantically empty, with the consequence that verbs would need to be subcategorized for their complementizer usage.

Complementizers also play a role in the modality expressed by the embedded clause they introduce, for example, the irrealis *de dicto* domain (Frajzyngier 1991, 1995, 1996) and different types of truth values (Ransom 1988). Their specific modality can be viewed as a semantic trace of the source for each particular complementizer, in other words, a case of persistence (cf. Hopper and Traugott 1993). For example, the complement structure may be chosen to express less commitment to the truth value of the assertion it contains (see §4.2.1 for the case of Taiwanese Southern Min).

As a starting point in analysing Sinitic languages, structural features that have been identified crosslinguistically for complementation are presented in the following list, synthesizing here relevant sections in Bresnan (1970, 1979), Dixon (1991), Horie (2001) and Noonan (1985). English examples are mainly used in the illustration of each point for reasons of convenience:

#### *Structural features of complementizers:*

- (i) Head-initial complementizers in complement clauses generally harmonize with VO word order and head-final with OV. English exemplifies the former case: *She thinks [that he’s intelligent]*.
- (ii) Matrix verbs are subcategorized for the semantically appropriate complementizer out of a range of possible markers. In the case of English, this includes *that, for...to*, interrogative pronouns and *POSS...ing*.  
*She wonders whether he’s coming* (\**She wonders [for him to come]*).

---

<sup>7</sup> The term ‘complementizer’ was coined by transformational grammarians in the 1960s, in preference to the more traditional term ‘subordinating conjunction’ (Bresnan 1970, Munro 1982). Note that this section on the syntax of complementation has the aim of defining the structural characteristics without purporting to give an exhaustive or representative overview of theoretical developments in its analysis.

- (iii) The particular complementizer may determine whether the complement is finite or not. For example, English *whether* and *that* govern finite clauses while *for...to* takes an infinitive complement.
- (iv) Typically, only one complementizer is possible per complex sentence: \**She thinks that whether he's intelligent*.
- (v) The complementizer may be facultative, as is the case for *that* in English: *She thinks [\_\_\_ he's intelligent]* but not for the *to* in (*for*) ... *to: I really want him \*(to) believe me*.
- (vi) Complementizers like English *that* are, however, non-deletable in subject position: ***That he is leaving makes me sad*** versus \**He is leaving makes me sad*.
- (vii) Complementizers are not permitted in non-embedded clauses, that is, in the matrix clause: ***That he is leaving*** cannot normally stand alone as an independent utterance.

The following list of structural features summarizes the grammar of complementation in the Sinitic languages investigated in the present analysis, namely, for those found to possess this category: Southern Min, Taiwanese Mandarin, Beijing Mandarin, Hong Kong Cantonese and Sixian Hakka.

*Structural features of complementizers in Sinitic languages:*

- (i) Complementizers precede the complement clause in all the Sinitic languages studied. Although Sinitic languages, and Mandarin Chinese in particular, show a mix of head-final and head-initial features, 'head-initial' complementizers appear to be in harmony with their basic SVO word orders in this case.<sup>8</sup>
- (ii) Main verbs in matrix clauses in the different Sinitic languages, including speech act, cognition, perception, emotion and modal verbs, are subcategorized for their ability to take SAY complementizers for introducing complements. The number of different verb classes allowed to co-occur with the complementizer depends on the degree of grammaticalization of the SAY verb. Hence, often only a subset of these verb classes will take the available complementizer in any given language (§3.1).
- (iii) Complementizers take syntactically complete clauses in Sinitic languages (§3.1), equivalent to finite clauses in languages which make this morphological distinction.
- (iv) Apart from Southern Min, there is typically just one complementizer available for use in the relevant Sinitic languages, and these are object-clause complementizers.

In Southern Min, the verb 'to see' *k'ua*<sup>n3</sup> 看 has also undergone grammaticalization into a complementizer with interrogative complements (Cheng 1991), similar to English 'whether'. The set of main verbs taking this complementizer is different from those taking 'say<sub>COMP</sub>'.<sup>9</sup> Regardless of the number of complementizers available, doubling up of complementizers is not possible in the head position of the complement clause, similarly to the case for English.

---

<sup>8</sup> See Chappell, Li and Peyraube (2007) for a list of head-final and head-initial features. One of the anonymous reviewers points out that this seemingly random mix explains why the principle of harmony does not work very well in the case of Chinese, referring to research on processing typology by Hawkins (1994). The reviewer further adds that Hawkins (1994) argues, on this basis, that the implication VO -> [COMP S] is actually exceptionless in the language samples analyzed to date.

<sup>9</sup> A complementizer derived from a SEE verb is also attested for Chaozhou, another language which belongs to the Southern Min subgroup (see Xu and Matthews 2007).

- (v) The complementizer is *not* obligatory in any of the Sinitic languages investigated which possess this category. – Clause combining can be achieved by simple juxtaposition.
- (vi) Unlike English, complements introduced by SAY complementizers are not found as an ‘expansion’ of the subject position of the main clause in Sinitic languages.<sup>10</sup> Complementation involves the integration of two independent sentences into one conflated one (Heine & Kuteva 2007: 224-261).
- (vii) Similarly to English, the complement structure including the SAY complementizer cannot stand alone as a syntactically finished structure.

The issue of finiteness, raised in (iii), is a controversial one for analytic verb-serializing languages, and so deserves a mention here, albeit cursory. The syntactic configuration for complementation in Sinitic languages, given in Figure 5 below, mirrors the claim made in my analysis that complement clauses of SAY verbs are syntactically complete (or ‘finished’), equivalent to a finite clause in languages which make such a morphological distinction. This is for the simple reason that they could all stand alone as independent clauses, can take lexical subjects, or contain modal verbs, while their main verbs may be modified by the regular gamut of aspect and polarity markers. In fact, in generative treatments of finiteness for Chinese languages, this is noted as a special feature of the syntax of SAY type verbs including *shuō* ‘say’, *gàosu* ‘tell’, and *xiāngxìn* ‘believe’ for at least standard Mandarin (see Huang 1998 and Li 1990 as proponents of this view while *contra* any such finite–non-finite distinction, see Hu, Pan and Xu 2001; Bisang 2001 provides an overview of the issues).

Hence, similar criteria to those applicable to English [COMP S] assist in defining the structural features of complementizers in Sinitic languages but are also useful in highlighting the differences, as for feature (vi). Many other issues concerning complementation and quotative constructions, including the debatable transitivity of SAY verbs and the hypotaxis versus parataxis question, cannot be dealt with here in any detail, for which see Munro (1982), Klamer (2000) and Güldemann (2001) *inter alia*. The issue regarding the placement of the complementizer within the matrix or complement clause is briefly treated in §3.1 which follows.

### 3. Grammaticalization from quotatives to complementizers: a crosslinguistic perspective on Sinitic languages

#### 3.1. General overview

In this section, I refer to five main interacting parameters which can be used to identify grammaticalization processes, based on Heine & Kuteva (2007: 33-46 & chapter 5). These are extension, desemantization, decategorialization, erosion and syntactic reanalysis.

##### 3.1.1. Extension and desemantization

Extension and desemantization are pragmatic and semantic parameters which essentially involve change in the contexts of use, and bleaching of lexical meaning respectively. On the application of a form to a novel context, not only is the paradigm of usage extended, in this case,

---

<sup>10</sup> They can however occur in the first position of the first part of a complex sentence but with different functions, such as topic marker or conditional marker.

to an increasingly larger number of matrix verb classes, but also a new grammatical meaning is evoked, that of the complementizer, a grammatical form which introduces a second clause, as a consequence of desemantization of its lexical meaning of ‘say’.

For complementation constructions based on quotatives, studies of both unrelated languages and linguistic areas attest to the fact that the number of permissible verb classes in the matrix clause expands according to an apparently well-trodden path creating the following implicational ordering. In fact, five main stages in the development of full-fledged complementizers from quotatives can be discerned on a crosslinguistic basis from recent studies on this topic in combination with the present research on Sinitic.<sup>11</sup>

- (1) factive verbs in general  $\supset$  (2) modal verbs  $\supset$  (3) stative and emotion verbs  $\supset$   
 (4) cognition and perception verbs  $\supset$  (5) speech act verbs

Figure 2: Implicational hierarchy of verb classes co-occurring with quotative complementizers

That is, if a language Y uses its complementizer, *C*, with modal verbs in the matrix clause, then it can be expected to also use it with cognition verbs. Such is the case with the Southern Min *C*, *kóng* < ‘say’, while Beijing Mandarin uses its *C*, *shuō* < ‘say’, as a complementizer up to the stative and emotion verb class. From this it can be correctly predicted that it also co-occurs with the cognition and speech act categories, but not with modals. Cantonese Yue uses its complementizer *wá*<sup>6</sup> with a limited number of cognition verbs, from which its co-occurrence with the basic group of speech act verbs can similarly be deduced.

This hierarchy does not necessarily represent a set of discrete stages, nor does it rule out the possibility that languages may add on two verb classes simultaneously. However, in this case, the verb classes would need to be adjacent. In other words, if a language expands its complementizer usage to two verb classes simultaneously, these could be emotion and modal verbs but not non-adjacent verb classes such as speech act and modal verbs.<sup>12</sup>

### Decategorialization

While expansion of verb classes is taking place and thus generalization of the complementation mechanism, each stage in the grammaticalization process may be accompanied by morphosyntactic changes which affect the quotative verb-cum-complementizer as it heads towards its eventual destination of invariable particle, with the typical characteristics associated with an independently-functioning full verb being gradually lost. Depending on the language

<sup>11</sup> This continuum, initially established on the basis of a smaller corpus of Sinitic languages in earlier work (Chappell 1997, 2001b), is strikingly similar to those proposed independently for other language taxa, and thus points to the possibility of universal features for this pathway. See, for example, Hoek (1982) and Saxena (1995) on Tibeto-Burman, Indo-Iranian and Dravidian languages; Klamer (2000) on Malayo-Polynesian languages (Austronesian); Lord (1976) for Ewe (Kwa, Niger-Congo), Frajzyngier (1991) for Mupun (Chadic); while for creoles, the reader is referred to Crowley (1989) on Bislama (English/French lexifier) and Plag (1992) on Sranan (English/Dutch lexifier); also see the binding hierarchy presented in Givón (1980) and a similar set of stages in the evolution from ‘say’ to complementizer discussed in Heine & Kuteva (2007: Table 6.5, 292, §5.3.2.3). Note that attested quotative complementizers appear to be restricted to particular geographical areas, as outlined in §1.2.

<sup>12</sup> Heine (2002: 95-96) raises the problem of potential gaps in implicational scales, observing that it would be wiser to view the latter in terms of probabilities.

type, this decategorialization includes the atrophying of subject concord, as well as tense and aspect marking, not to mention the potential of the SAY to be negated (Heine & Kuteva 2007: 40-42). In languages with little or no inflectional morphology, this development may also take the form of an impersonal interpretation of the quotative verb as ‘It is said’ following on from conventionalization of subject ellipsis (Klamer 2000).

The parameter of decategorialization ties in with the obligatory use of a complementizer which can be located at different points on the grammaticalization hierarchy given above.<sup>13</sup> For example, the Bislama complementizer *se* is obligatory with all verb classes up to modals (Crowley 1989), as too for Ewe *bé* (Saxena 1995, Heine et al 1991). At the endpoint of this process of grammaticalization, the original quotative verb may cease altogether to function grammatically as a verb, and become ‘fossilized’, as is the case for *fēn* in Buru (Klamer 2000), *bhan* in Nepali (Ebert 1986) and *se* in Bislama, all defunct as transitive verbs of saying. In Sinitic languages, however, complementizers remain optional in use, suggesting that the grammaticalization process is less entrenched and that the complementizers are in general ‘younger’ than in the aforementioned examples.

Once the stage is attained where the quotative verb has become an indeclinable particle and one that is completely semantically bleached, lexical renovation may come into play, with a new quotative verb replacing the older one. Such is the case in Vietnamese: its two complementizers, *là* and *rǎng* have both reached the modal verb stage on the hierarchy. However, *rǎng* < ‘say’ is more appropriate for use as a complementizer in the written register, while two other verbs are taking over its functions, particularly in the spoken register. Obsolescent as a lexical verb of saying, *rǎng* is being replaced by the speech act verb *nói*; and in its complementizer use by *là* < ‘be’ (see Tru’o’ng 1970: 175-177).<sup>14</sup> This recalls the celebrated case of Yoruba, with its three complementizers based on quotative verbs, each of which is at a different stage of evolution (Lord 1976).

### 3.1.3. Erosion

The fourth parameter of phonetic erosion may be involved in grammaticalization, as in loss of phonetic or syllabic segments of the complementizer, including constituent morphemes. One such example is the English subordinating conjunction *while*, a reduction of the phrase in Old English *ðe hwile ðe* ‘the while that’ (Harris & Campbell 1995: 288); another is the quotative form *mî* in Khoekhoe (Khoisan) which appears to be a contracted form of *mîl* ‘say’ (Güldemann 2001: 231, 465). In the spoken discourse data on Sinitic languages, such as for Southern Min, the complementizer *kóng* /kɔŋ<sup>51-55</sup>/ sometimes manifests itself as the reduced form [ɔŋ<sup>55</sup>], suggestively losing its initial consonant in fluent speech, in this function of bound element to the preceding matrix verb (see example (1)). We do not have sufficient data however to claim that this is a consequence of the grammaticalization process and not simply a general feature of fluent speech which may affect its other lexical uses. A reduction of tonal possibilities may also be underway to just high level /kɔŋ<sup>55</sup>/, briefly discussed in §4.2.

<sup>13</sup> This process is known as ‘obligatorification’ or ‘syntacticization’ in certain works on grammaticalization.

<sup>14</sup> I thank Danh Thanh Do-Hurinville for checking these data, including testing for co-occurring verb classes.

### 3.2. Syntax: reanalysis of constituency

Syntactic change globally affects the two parts of the complex sentence – the matrix clause along with its complement. That is, while the erstwhile quotative verb is undergoing the various semantic, pragmatic and morphosyntactic changes described in §3.1, its role as the new linking mechanism for the entire construction may be subject to reanalysis regarding which constituent it belongs to.

Some claim this involves a repositioning of the complementizer as it ‘shifts down’ from the matrix clause to become the head of the complement or embedded subordinate clause, for example, in Chamling (Tibet-Burman, Nepal, see Ebert 1986). Complementation is thus accounted for in terms of a boundary shift (Langacker 1977) or reinterpretation of constituency (Harris & Campbell §7.3, chapter 10). Figure 3 is reproduced from Heine et al (1991: 215-216) to illustrate this redefining of constituent boundaries:

(A, B)	C	>	A	(B, C)
<i>I say that:</i>	<i>he comes.</i>	>	<i>I say</i>	<i>that he comes.</i>

Figure 3: Constituent-internal reanalysis (Heine et al 1991: 215-216)

The trigger for integration of the two clauses into a complex sentence is set off when the quotative verb, extended to a novel context where it co-occurs with a cognitive class matrix verb, completely loses its ability to introduce a quotation (see also Plag 1992, Ebert 1991, Güldemann 2001: 166). This yields the following reanalysis:

NP V<sub>[quotative]</sub> (*intonational pause*) : s[“Direct Speech”]

> NP (V<sub>1[*speech act*]</sub>) V<sub>2[*quotative*]</sub> s[*Indirect speech*] (+changes in prosody and pronominal deixis)

> NP V<sub>1[*cognitive*]</sub> — [COMP Complement clause].

Figure 4: Fusion of the matrix construction with its linked proposition

Indeed, in recent generative grammar analyses, a similar analysis is proposed for this final grammaticalized stage of a CP or complement phrase embedded under the matrix clause by means of a complementizer, C (Wu 2004, Yeung 2006):

> NP VP<sub>CP</sub> C<sub>IP</sub> [Complement clause]

Nonetheless, I do not subscribe to this view of constituency in the case of Sinitic languages, since a different kind of syntactic reanalysis appears to have taken place: this is seen first of all in the placement of the intonation break *following* the complementizer which is recorded far more frequently than not in my Taiwanese Southern Min and Cantonese Yue spoken discourse data. Secondly, in losing its verbal characteristics, the complementizer forms a new complex unit attached to its preceding speech act or cognitive verb, thus forging a looser syntactic relationship of clause linkage with the complement clause which follows. Hence, for spoken data in the relevant Chinese languages, we specifically find the following configuration:

NP V<sub>[quotative]</sub> (*intonational pause*) : s[“Direct Speech”]

> NP (V<sub>1[*speech act*]</sub>) V<sub>2[*quotative*]</sub> s[*Indirect speech*] (+changes in prosody and pronominal deixis)

> NP V<sub>1[*cognitive*]</sub> - COMP (*intonational pause*) — [Complement clause<sub>[*finite*]</sub>]

Figure 5: Juxtaposition of the matrix construction with its linked proposition for Sinitic languages

This feature has been characterized by Matisoff (1991: 398-400) for Southeast Asian languages in the sinospheric zone as the general tendency of the verb category to develop into verb particles when in V<sub>2</sub> position. Two such representative languages in his study, Thai and Khmer, both use grammaticalized SAY verbs at the end of a non-final clause to introduce the complement clause. Klamer (2000) also describes the intonation break as occurring in the position following the complementizer in Buru (Malayo-Polynesian, Austronesian) while Güldemann (2001: 164-169) discusses pause and segmental sandhi phenomena with respect to several African languages. He goes on to observe (2001: 179) that such a repositioning of the complementizer in a ‘lower’ embedded clause may not be discernible in all languages. Hence, complementation based on quotative syntax does not necessarily involve the creation of hypotaxis in the form of a clause-initial (or -final) complementizer in the lower embedded clause, as earlier remarked upon in §2.

### 3.3. Stages in the grammaticalization chain for SAY verbs in Sinitic languages

Sinitic languages for which the categories of either complementizer or semi-complementizer have been identified at present are Southern Min, Cantonese Yue, Beijing Mandarin, Taiwanese Mandarin and Sixian Hakka. I did not find any evidence of this grammaticalized category in the discourse data for Shanghainese Wu, Changsha Xiang, Nanchang Gan and Huojia Jin, nor in descriptions of Standard Mandarin.

The findings for the 10 Sinitic languages in this study, which belong to 8 of the major branches, are summarized in Table 1 found directly below. Note that, of the ten recognized Sinitic languages, only representatives for the Pinghua and Huizhou dialect groups are missing from this table, merely due to the unavailability of discourse materials at present.<sup>15</sup>

---

<sup>15</sup> Tables with the more specific findings and classifications of the functions of SAY verbs for each of the additional Sinitic languages are reproduced in Appendix 2, where not presented in the main body of the analysis. Furthermore, note also that the findings for Huojia Jin, Nanchang Gan and Sixian Hakka must be treated as merely showing tendencies at this point in research, in view of both the small amount of data available for the Jin and Gan dialects and the fact that the only ample corpus of discourse data for Sixian Hakka dates from the 1950s. Nonetheless, given that the semi-complementizer stage can already be detected in the Hakka texts, one can surmise that the degree of grammaticalization should, theoretically speaking, be more advanced today.

Table 1: SAY verbs in 8 branches of Sinitic

LANGUAGE	VARIETY / DIALECT	SAY VERB	GRAMMATICALIZED FUNCTION	CORPUS / SOURCE
1. Southern Min 闽南话	Taiwanese	<i>kóng</i> 講	Stage V complementizer	Discourse data See §4.2 and <i>Appendix 1</i> TSL
2. Cantonese Yue 粵	Hong Kong	<i>wa<sup>6</sup></i> 話	Stage II-III semi-complementizer	Discourse data See §4.3 and <i>Appendix 1</i> TSL
3. Mandarin 官话	Standard ( <i>pǔtōnghuà</i> )	<i>shuō</i> 說	—	Meng (1982) Y. Liu (1986).
	Beijing	<i>shuō</i> 說	Stage IV Complementizer	Fang 2006
	Taiwanese	<i>shuō</i> 說	Stage IV-V Complementizer	Cheng 1991 Liu 1996 Huang 2003
4. Hakka (Kejia) 客家	Sixian, Taoyuan, Taiwan	<i>koy<sup>31</sup></i> 講	Stage II semi-complementizer	Yang 1957 <i>See Appendix 2</i>
5. Xiang 湘	Changsha	<i>kan<sup>41</sup></i> 講	—	Discourse data TSL <i>See Appendix 2</i>
6. Gan 贛	Nanchang	<i>wa<sup>6</sup></i> 話	—	Discourse data TSL <i>See Appendix 2</i>
7. Wu 吳	Shanghai	<i>kā<sup>34</sup></i> 講	—	Discourse data TSL <i>See Appendix 2</i> Xu and Tang 1988
8. Jin 晉	Huojia	<i>suə<sup>73</sup></i> 說	—	He 1989 <i>See Appendix 2</i>

Legend: ‘—’ = not found in the sample of discourse data or in the relevant publications.

TSL = Data belongs to the corpus of discourse materials set up for the project: *Typology of Sinitic Languages*, CRLAO, Paris

An overview of stages relevant for Sinitic languages is next presented.

The very first precondition for any complementizer development based on quotatives is evidently some kind of quotative verb, specifically a speech act verb in Sinitic. This source of a future complementizer may have either, if not both, transitive or intransitive frames and be polysemous between ‘say’, ‘talk’, ‘speak’, ‘tell’ with further semantic specialization to pejorative speech acts such as ‘criticize’, ‘talk about’ and ‘blame’. The SAY verbs in question are typically high frequency and semantically general verbs, possessing extended cognitive senses including ‘mean’, ‘think’, ‘realise’ and ‘know’ which prepare the ground well for its future complementizer usage with cognitive verbs. In Chinese languages, SAY verbs are low in transitivity, since they regularly take cognate objects to do with speech or language (Chappell 2002, §9.1), for example, *kòng-oē* say-words = ‘speak’ in Southern Min. The exception is the transitive frame for ‘criticize’, ‘talk about’ and ‘blame’ which typically takes a human NP as its object and is, by way of contrast, highly transitive. Next, I present the putative stages for grammaticalization of Sinitic SAY verbs in more detail, using the context-based framework proposed in Heine (2002), comprising four stages on a continuum of grammaticalization that can account for the rise of new grammatical meaning: initial stage, bridging context, switch context and conventionalization. This complements the broader discussion of grammaticalization processes

given in §3.1 and §3.2, particularly with respect to the parameters of extension and desemanticization.

Stage I : *Quotative construction (Initial stage)*

*Syntactic configuration:* NP<sub>SUBJECT</sub> (PP<sub>ADDRESSEE</sub>) V<sub>QUOTATIVE</sub> : [QUOTATION]

In general, quotative constructions serve to introduce direct speech and indirect speech, although in some language areas, there may be a restriction to just direct speech, as in the South Asian linguistic area (Noonan 2006 on Chantyal [Tibeto-Burman], Ebert 1991 on South Asia in general). In the Sinitic taxon, the three main SAY verbs which take on this function of complementizer are JIANG 講, SHUO 說 and HUA 話, depending on the language. Note that there is no restriction to direct speech for Sinitic. A list with the phonetic forms of SAY verbs found in the Sinitic languages in this survey is given in Table 1.

As already observed, in this initial preparatory stage preceding grammaticalization, the source verb for the quotative is a full verb which can be negated, can take aspect marking or be governed by a modal verb. Importantly, many of these morphosyntactic properties transfer to the bridging context when it takes part in a quotative construction introducing reported speech.

The SAY verbs in the Sinitic languages, Changsha Xiang, Nanchang Gan, Shanghainese Wu, Huojia Jin and Standard Mandarin, exemplify this first stage for languages that have not yet embarked upon any grammaticalization process (§5.4).

STAGE II : *Semi-complementizer in V<sub>2</sub> position of a serial verb construction with quotative function (Bridging context)*

*Syntactic configuration:* NP<sub>SUBJECT</sub> V<sub>1</sub> (X) V<sub>2</sub>[SEMI-COMPLEMENTIZER/QUOTATIVE] : [QUOTATION]  
V<sub>1</sub> = speech act

X = { DO  
discourse particle  
aspect marker  
PP  
pause  
adverb

To reinforce a generic speech act verb serving in quotative function, a variation on the Stage I construction comes into being whereby a more specific speech act or communication verb (COMM) combines with this quotative as a semantic bolster or reinforcement: V<sub>1</sub>[SPECIFIC COMM] V<sub>2</sub>[QUOTATIVE].

Verb complexes made up of a combination of specific and generic verbs belonging to the same semantic field are not uncommon in verb serializing languages, as is the case both synchronically and diachronically in Chinese languages. For example, the use of two verbs of giving in Late Medieval Chinese (7<sup>th</sup> – 9<sup>th</sup> centuries) in just such a V<sub>1</sub>V<sub>2</sub> complex led to the

reanalysis of the semantically generic  $V_2$  as a dative preposition introducing a following indirect object (Peyraube 1996). The new construction then extended its use to verb classes other than ‘give’ which could code the overall sense of transferral of an object. A similar phenomenon found in other language families is that of verb copying, discussed in Sranan (Plag 1992), or bisected constructions (see Güldemann 2001: 113). Hence, the Sinitic case is not an isolated development, given that similar kinds of ‘verb doubling’ or reinforcement turn up in different parts of the world.

For this bridging stage as a quotative serial verb construction, the  $V_1$  slot continues to be filled in a semantically appropriate way by speech act and communication verbs such as *ask*, *nag*, *tell*, *mock* and *advise*. However, a new inference concerning grammatical meaning becomes available for the semantically generic quotative  $V_2$  ‘say’, enabling a simultaneous interpretation as a linking mechanism which introduces the following clause, while not cancelling out its lexical sense. In terms of morphosyntax, apart from modification of  $V_1$  by aspect and negation markers, direct objects, prepositional phrases, and other linguistic material such as particles may intervene between  $V_1$  and  $V_2$ . Nonetheless, although sharing these kinds of modification,  $V_1$  and  $V_2$  do not yet function as a single verb unit, evident in the fact that they can be separated by the constituents just listed. Furthermore, the semantics of  $V_1$  and  $V_2$  remain clearly compositional, even though the lexical load of  $V_2$  is relatively light: *He teased me saying pianists should only play noteworthy music.*

Sixian Hakka and Cantonese Yue have both clearly reached this stage, with Cantonese about to embark upon Stage III.

STAGE III: *Complementizer stage with cognition verbs (Switch context)*

*Syntactic configuration:* NP<sub>SUBJECT</sub> ...  $V_1$ - COMPLEMENTIZER [CLAUSE]

$V_1$  = speech act; cognition and perception

By stage III, the erstwhile quotative verb has fully developed into a complementizer. This is evident from at least two main angles: firstly, in Sinitic languages such as Southern Min, the quotative  $V_2$  has become an unmodifiable verb, resembling an invariant particle; secondly, in that it now combines with cognitive verbs, such as *think*, *reckon* and *plan*, the construction can no longer be used to introduce reported speech, but rather, different kinds of propositions. This is also reflected in the fact that correspondingly the combined meaning of  $V_1$  and  $V_2$  is no longer compositional as it was at Stage II, e.g. *I thought (\*saying) that she would win at chess.* The original lexical meaning of  $V_2$  is incompatible with its new grammatical meaning, the only one available for the construction. This turning point in grammaticalization or switch context, as defined in Heine (2002), has effectively been reached.

In the Sinitic languages under consideration, no intervening material is allowed between cognitive verbs as  $V_1$  and the fully developed complementizer, the former  $V_2$ . Together they create an even more tightly-bonded unit, since elements such as direct objects and discourse particles cannot be inserted between  $V_1$  and the former  $V_2$ , now a virtually indivisible unit. In Taiwanese Southern Min, for example, all the instances of  $V_1$ - *kóng* (‘say’ > COMP) are transcribed as a complex contained in the one intonation unit in the spoken data examined. That is, none of the  $V_1$ - COMPLEMENTIZER is separated by segmental material, an intonation break or pause of any kind.

In unrelated languages such as Sranan, a similar situation obtains: the high degree of fusion between  $V_1$  and the newly forged complementizer (that is, the ex- $V_2$ ): means that it is not possible to insert pauses between  $V_1$  and *taki* ('say' > COMP) (Plag 1992). This bonding of  $V_1$ - $V_2$  is also reflected in the discourse data in the case of self-initiated repairs: if the  $V_1$ -COMPLEMENTIZER is subject to some kind of interruption or disfluency, the whole unit will be repeated by the speaker.

After this crucial turning point in the grammaticalization of quotative verbs where they are linked with cognitive verbs, the scope of the semantic field for the matrix verb is open to further expansion. The first that should be mentioned is the perception class. These verbs including 'see', 'hear' may be added on at the same time as the cognition class, being closely allied through common metaphorical extension to 'understand' and 'know'.

Hence, arrived at this stage in the grammaticalization process, the erstwhile quotative marker no longer introduces reported speech of any kind but rather propositions containing the thoughts, reflections and feelings of the subject coded in the matrix clause (cf. also Plag 1992). The construction is associated first and foremost with the representation of knowledge and understanding, allying it closely with the functions of *that* in English (Bolinger 1972, Wierzbicka 1988).

As mentioned above, Cantonese Yue is just in the process of adding on cognition and perception verbs, and could be described as straddling the two stages, II and III.

STAGE IV and onwards: *Broadening in scope of verbs taking the complementizer*  
 $V_1$  = speech act; cognition and perception, emotion and stative verbs

As the next step along this continuum, emotion and other kinds of stative verbs depicting feelings may be added to the list of verbs 'subcategorized' for this type of complementizer: *I feel (\*saying) that he's become more wary*. Again, they are semantically close to perception verbs: consider English 'feel' which belongs to both perception and psychological categories (Sweetser 1990). This may not necessarily be a discrete stage in all languages, for example, if these verb classes are added on simultaneously with those of Stage III. However, it is distinguishable for the relevant Sinitic languages due to a greater syntactic integration with the matrix verb.

Beijing Mandarin has reached this point on the continuum.

STAGE V: *Onset of conventionalization of the complementizer usage*  
 $V_1$  = speech act; cognition and perception, emotion and stative verbs, modal verbs

One of the final stages for complementizers based on quotatives is expansion to the modal verb class. For Sinitic, this is possible in at least Taiwanese Southern Min, Chaozhou Southern Min (Xu and Matthews 2007) and Taiwanese Mandarin, but not in Beijing Mandarin, as far as the data reveal. This stage also witnesses the conventionalization of the complementizer, for which Southern Min shows clear signs of the onset. The meaning of 'say' is completely at semantic odds with the contexts in which the newly grammaticalized complementizer now comes to be used: *It must be (\*saying) that they've forgotten*.

For many languages which have achieved this degree of generalization early on in the use of their ex-quotative complementizer, the implication may be that the complementizer is no

longer synchronically used as a verb, as in Bislama, where *se* is highly grammaticalized, being obligatory with all verb types up to and including *want* (Crowley 1989).

Sinitic languages have clearly not all reached the same stage as Bislama, since complementizers are optional. Moreover, SAY verbs present a case of divergence (Hopper 1991), retaining all their lexical usages, since they still function as quotative and lexical verbs in syntactic contexts other than complementation. Nonetheless, that the degree of grammaticalization is greater in some languages than in others is evident in the fact that in Taiwanese Southern Min the complementizer can be directly attached to its related quotative verb, as *kóng kóng* ‘say that’. This suggests a reasonably high degree of conventionalization for the new complementizer use in that source and target meanings can co-occur in the same clause and in this case adjacent to one another. Compared with this, Beijing Mandarin and Hong Cantonese show a lower degree of conventionalization of the complementizer usage, since there is a requirement for at least a pause or particle to intervene between  $V_1$  and the complementizer: *shuō... shuō* ‘say that’ and *waa<sup>6</sup> ... waa<sup>6</sup>* ‘say that’, treated in §5.2 and §4.3 below respectively.<sup>16</sup>

Upon reaching this level, expansion of verb classes taking the complementizer to other kinds of factive verbs may take place, with the potential of broadening the scope to further verb classes in an unrestricted manner, if the complementizer becomes generalized in function as a clause linker.

Figure 6 summarizes the findings for the 10 Sinitic languages in this study with respect to presence or absence of a complementizer and the degree of grammaticalization of their SAY verbs.

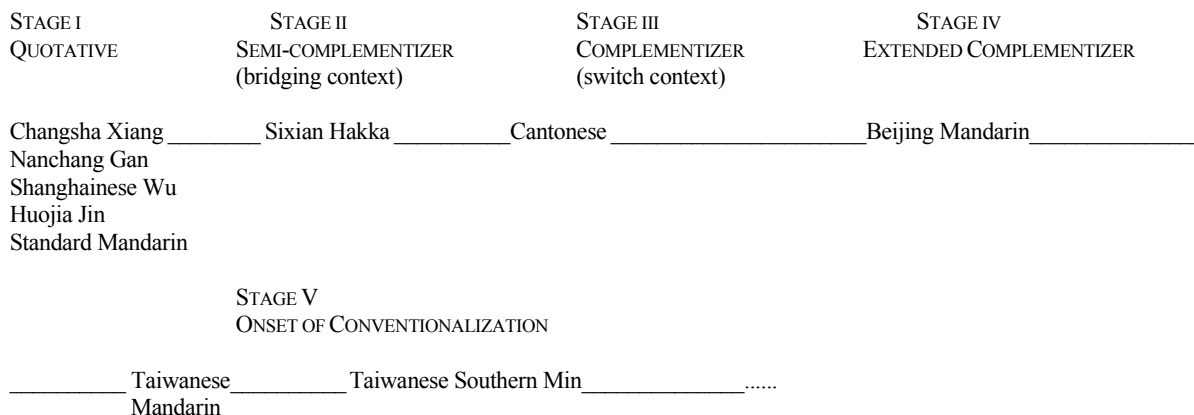


Figure 6: Stages in the grammaticalization chain for complementizers in 10 Sinitic languages

In the next section, apart from treating the data in terms of the stages of grammaticalization just presented, the quantitative analysis of Taiwanese Southern Min and Hong Kong Cantonese Yue is presented, using a database of colloquial texts collected in Taipei, Taiwan and in Hong Kong. These have been transcribed using the format and practices developed for the project on

<sup>16</sup> For a more detailed description of conventionalization, see Heine (2002: 85).

Spoken American English and other languages at the University of California at Santa Barbara (see J. Dubois et al 1993).<sup>17</sup>

#### 4. Quantitative analysis of Taiwanese Southern Min and Cantonese Yue

Min constitutes one of the smallest language groups in mainland China comprising approximately 4% of speakers of Sinitic languages (estimates vary between 40-50 million speakers). It is concentrated in Fujian province on the southeastern seaboard of China. In Taiwan, a variety of Southern Min is also spoken as the first language of the majority of the population (67%, or approximately 15/22 million; Huang 1993). Mandarin is the official language, as on the mainland.

The Yue languages are located principally in Guangdong province and the eastern parts of neighbouring Guangxi autonomous region, making up approximately 5% of the Han Chinese population in China (50-60 million speakers). Best-known in this dialect group is Cantonese, represented by the closely related varieties spoken in Hong Kong and Guangzhou.

##### 4.1. The quotative verb *kóng* 講 in Taiwanese Southern *Min*

The verb *kóng* in Taiwanese may simply be used as a transitive verb ‘say, talk, tell, explain’ as in (2) below, and also in (1) above, with regard to the first instance only. In this function, it forms two main construction types.

For such monotransitive clauses with *kóng* in Taiwanese, there is an optional preverbal slot for a prepositional phrase to code the addressee: This basic lexical use of ‘say’ with a direct object accounted for 81/263 instances in the texts.

I.	SUBJECT <sub>AGENTIVE</sub>	pp[PREP	NOUN <sub>ADDRESSEE</sub> ]	SAY	NOUN <sub>PATIENT</sub>			
(2)	啊 我	共	厝裏	著	無	講	啥物	話
	<i>a goá</i>	<i>kah</i>	<i>chhu-li</i>	<i>toh/tioh/</i>	<i>bô</i>	<i>kóng</i>	<i>sím-mìh</i>	<i>oē</i>
	PRT 1SG	COM	family	then	NEG	say	what	word

‘So I didn’t talk a lot with the family.’ (*Japanese* 46)

The verb *kóng* can also introduce the locutionary topic as its object, as in the following example where it has the meaning ‘to talk about’:

(3)	啊	伊	講	汝	按怎?
	<i>a</i>	<i>i</i>	<i>kóng</i>	<i>lì</i>	<i>àn-noá<sup>n</sup>?</i>
	PRT	3SG	say	2SG	how

‘So what did he say about you?’ [Context: Youngest Uncle practices divination to tell the fortunes of all the family members.] (*Fate* 200)

<sup>17</sup> The combined set of texts comprise 79:23 minutes of recording of narratives and conversations which have been segmented into intonation units, a unit of discourse that has prosodic, syntactic and cognitive ramifications (see Chafe 1993). The Southern Min texts total 58:17 minutes and the Cantonese texts 21:06 minutes. See Appendix 1 for more details.

The second main syntactic configuration for the verb *kóng* is one which can introduce either a direct quotation or indirect reported speech. In fact, in the recorded texts, this was the largest category of examples for *kóng*: 88/263. The example in (4) comes from a conversation recorded in Taipei and contains a direct quote.

II.	SUBJECT <sub>AGENTIVE</sub>	SAY	COMPLEMENT				
(4)	阮 彼	個	系主任		着		講,
	<i>gún he</i>	<i>ê</i>	<i>hē-chújīm</i>		<i>toh /tioh/</i>		<b><i>kóng</i></b> ,
	1PL that	CL	department-director		then		SAY
<Q	啊 汝	順續	去	錄	蜀寡	音	轉來 啊 Q>
	<i>a lí</i>	<i>sūn-sòā</i>	<i>khi</i>	<i>lòk</i>	<i>chit-koá im</i>	<i>tng-lai</i>	<i>a</i>
	PRT 2SG	convenient	go	record	one-CL <sub>PL</sub> sound	return	PRT

‘That director of mine said, “Well, you might as well do some taping while you’re there to bring back .”’ [*Japanese tales* 29-30]

The next example shows the use of *kóng* in Taiwanese where the complement is indirect or reported speech. The syntactic structure is the same as for a direct quote with the difference residing in prosody and pronominal usage. In other words, the complement clause is zero marked. Hence, as can be seen, the use of a complementizer to introduce the reported speech, similar to *that* in English, is not required by the grammar of Southern Min.

(5)	因為	伊	置	講	欲	共	<MC 恩澤 MC>
	<i>in-ui</i>	<i>i</i>	<i>tī</i>	<b><i>kóng</i></b>	<i>beh</i>	<i>kah</i>	<i>Ēn-zé</i>
	because	3SG	PROG	<b>say</b>	want	OM	NAME
	送	於	美國				
	<i>sàng</i>	<i>e</i>	<i>Bí-kok</i>				
	send	LOC	USA				

‘Because she’s been saying that (she) would send En-Tse to America to study.’  
(*Fate* 748)

When it is used like this as a fully lexical quotative verb, *kóng* can still be aspectually modified or negated. Aspect is mainly coded in Southern Min by a set of preverbal markers, and negation by preverbal negative adverbs. This is illustrated in example (5), where *kóng* is preceded by the progressive aspect marker *tī* 置.

#### 4.2. Grammaticalized uses of *kóng* 講 ‘say’ in Taiwanese Southern Min

There are at least five uses of *kóng*, all found in the database, which can be distinguished from its lexical and quotative verb uses. These belong to the irrealis and metalinguistic domains,

and show varying degrees of grammaticalization along different paths. Apart from the complementizer and topic marker usages (*kóng* + NP/CLAUSE ‘As for...’), there is a variety of subordinating functions, including the purposive and conditional. These are only partially grammaticalized, however, as *kóng* occurs in collocation with the regular grammatical marker for the purposive (*só-i kóng* so-SAY<sub>PURP</sub> ‘so that’s why ...’) and the conditional (*nā kóng* if-SAY<sub>COND</sub> ‘if ...’). There is also a clause-final usage as a modal particle in assertions and warnings. Note that these functions can be formally distinguished since each one corresponds to a different syntactic frame or collocation, as indicated in Table 2. Lexical verb use accounts for the majority of tokens across the three texts of 169/263 while the five nonlexical uses comprise 87/263; the remaining seven instances represent false starts and interruptions (7/263).

Not only is the complementizer usage the most grammaticalized of these five nonlexical functions, but as Table 2 shows, it is also the most frequent in the database, providing sufficient material for a comparison with Cantonese Yue discourse data. Discounting the two main lexical uses, the complementizer function accounts for 52/87 grammaticalized tokens. Table 2 summarizes the data in terms of descending order of frequency.<sup>18</sup>

Table 2: Tokens of *kóng* in 3 recorded texts in Taiwanese Southern Min

Lexical verb uses	FATE	JAPANESE	JESSE	TOTAL
Length in minutes	14:06	26:23	17:48	58:17
N <sup>0</sup> Intonation units	796	1216	831	2843
1. Lexical : Verb <sub>trans</sub> <i>kóng</i> ‘to say’	40	38	3	81
2. Quotative : NP <i>kóng</i> « QUOTED SPEECH »	46	27	15	88
<b>Grammatical functions</b>				
1. Complementizer : Verb- <i>kóng</i> + CLAUSE ‘that’	20	23	9	52
2. Topic : ( <i>lái</i> ) <i>kóng</i> + NP/ CLAUSE ‘As for ...’	4	7	5	16
3. Conditional : <i>nā kóng</i> + CLAUSE <sub>PROTASIS</sub> ‘if ...’, ‘supposing ...’	9	3	2	14
4. Purposive : <i>m̄ chiah kóng</i> ; <i>só-i kóng</i> ‘so that’s why ...’	0	3	0	3
5. Discourse marker of self-evident assertions or warnings: Clause + <i>kóng</i> ‘I’m telling you ...’	1	0	1	2
Indeterminate : false starts & truncations	2	1	4	7
TOTAL	<b>122</b>	<b>102</b>	<b>39</b>	<b>263</b>

<sup>18</sup> Across the three Southern Min texts, there were also 19 examples of the set phrases or lexicalizations: *kan-na kóng*; *thiong kóng*; *an-ni kóng* ‘like this’, ‘for example’, which are all used to introduce an analogy or some kind of illustration as an explanation. There were a further 14 examples of the set phrase *tōh sī kóng* 着是講 ‘that is to say’ used to make an elaboration. I have not included either group in this analysis.

The main analysis concentrates on the complementizer usage for *kóng*, the most grammaticalized usage, while the other uses, listed in Table 2, are treated in a separate study on the polygrammaticalization of SAY.<sup>19</sup>

I next examine the verb classes co-occurring with *kóng* in the discourse materials and discuss its degree of grammaticalization as a complementizer in Southern Min. This serves to exemplify the hypothetical stages proposed in §3.3.

#### 4.2.1. Complementizer: (NOUN<sub>SUBJECT</sub>) - VERB<sub>1</sub>-*kóng* + CLAUSE

In Southern Min, the complementizer *kóng* forms a verb complex with the preceding speech act, cognition, perception or modal verb, introducing an embedded clause with new information status containing an assertion, question, thought, feeling, wish or discovery. In the texts, this function constituted 52 of the 87 nonlexical uses of *kóng*. However, as mentioned earlier, its use in this position is not a requirement of Southern Min grammar.<sup>20</sup> Consider the following pair of examples. One is from the corpus and the other has been elicited for the purposes of highlighting the contrast in modality between presence and absence of the complementizer. – There were not any serendipitous minimal pairs in the corpus.

The example concerns youngest uncle's divination of one of the sisters, Yu-chin (= 'she'), to see whether she will return home to Taiwan or not. He predicts that she will, as reported in (6):

- (6) 阿叔 講 伊 明年 會 轉來。  
*á-chek kóng i mē<sup>n</sup>-nī<sup>n</sup> ē tng-lâi.*  
 uncle say 3SG next-year will return  
 'Your uncle said (*kóng*) she will return home next year.' (Fate 23b)

The elicited variation of this example shows the addition of the complementizer *kóng* :

- (7) 阿叔 講 講 伊 明年 會 轉來。  
*á-chek kóng kóng i mē<sup>n</sup>-nī<sup>n</sup> ē tng-lâi.*  
 uncle say SAY<sub>THAT</sub> 3SG next-year will return

'Your uncle said that (*kóng kóng*) she would return next year.'  
 (Elicited example)

Strikingly, the addition of the complementizer in (7) has the consequence of conveying a different nuance in its modality. This involves less commitment on the part of the speaker regarding the certainty of Yu-Chin's return home to Taiwan for good from Australia.<sup>21</sup> Example (7) is thus used in contexts where the speaker wishes to distance herself from the embedded

<sup>19</sup> This is a chapter of my book (in prep.). See also Chappell and Peyraube (2006).

<sup>20</sup> The apparent optionality of *kóng* appears to be a similar case to the English complementizer *that* which can be omitted under certain discourse conditions, particularly in less formal styles and genres. This contrasts to the case for the complementizer *que* in Standard French, which is rarely omitted.

<sup>21</sup> Random testing of examples with a native speaker showed that the complementizer *kóng* could be added in, without difficulty, after *verba dicendi* and cognition verbs, that is, when it hadn't been used by speakers, or freely omitted when it had.

proposition and the likelihood of the event taking place. A contextually unrelated example of *kóng kóng* is found in example (17) below where a different speaker, the father, expresses doubts about another one of youngest uncle's divinations: 'he said that I had done some good deeds'. The lack of commitment to this statement is clear from the context, since the father proceeds in fact to explicitly disagree with the assertion, saying that he thought he had never done any good deeds at all in his life (= example (10) below).

In my data, *kóng* was used as a complementizer with 18 different speech act, cognition, perception and modal verbs. Not surprisingly, the highest frequency combination was with the main verb of cognition, *siū<sup>n</sup>* 'think' (17/52). *Kóng* also occurred with *m̄ng* 'ask'; *chhiò* 'laugh'; *kì-chài* 'record'; *liām* 'nag, insist'; *kámkak* 'feel', *beh* 'want' and *m̄-chai* 'not know', among other verbs. It has already been noted in §3.3 that there is no constraint on *kóng* acting as a complementizer to itself, that is, to its basic lexical verb usage: *kóng kóng* 'say that' (10/52), whereas this kind of doubling up is subject to certain restrictions in Cantonese and Beijing Mandarin (§4.3 and §5.2). This possibility is illustrated by the Southern Min examples (7) and (17) in particular. Table 3 presents a classification of all the verbs co-occurring with *kóng* across the three texts:

Table 3: V<sub>1</sub> in the verb complex formed with the complementizer *kóng* across three Southern Min texts

1. <i>kóng kóng</i> 講講 'to say that'	10
2. <i>m̄ng kóng</i> 問講 'to ask that'	4
3. <i>chhiò kóng</i> 笑講 'to mock that'	3
4. <i>liām kóng</i> 唸講 'to nag that'	3
5. <i>iok-sok kóng</i> 約束講 'to agree that'	1
<i>kì-chài kóng</i> 記載講 'to record that'	1
7. <i>siū<sup>n</sup> kóng</i> 想講 'to think that'	17
8. <i>siau siū<sup>n</sup> kóng</i> 數想講 'to intend that'	1
9. <i>m̄-chai (iá<sup>n</sup>) kóng</i> 怀知(影)講 'to not know that'	2
10. <i>hoat-hiān kóng</i> 發現講 'to find that'	1
11. <i>kámkak kóng</i> 感覺講 'to feel that'	1
12. <i>khoa<sup>n</sup> kóng</i> 看講 'to be of the view that'	1
13. <i>thia<sup>n</sup> kóng</i> 聽講 'to hear that'	1
14. <i>chin gao kóng</i> 真賢講 'very clever that'	1
16. <i>beh kóng</i> 慾講 'to want that'	1
17. <i>m̄-biēn kóng</i> 怀免講 'to not need to'	1
18. <i>e-sài kóng</i> 会使講 'it is possible that'	3
<b>TOTAL</b>	<b>52</b>

In the ensuing discussion, verb classes compatible with the use of complementizer *kóng* are treated one-by-one in terms of the putative stages of grammaticalization proposed in §3.3.

In the use of a serial verb construction, where SAY verbs come to be used as quotative markers in V<sub>2</sub> position, the lexical meaning is not at first completely bleached. Two corpus

examples follow with a representative speech act verb. The verb complex *liām kóng* ‘nag-say’ in (8) could still be rendered as ‘nagged saying’, but also as ‘nag that’ or ‘insist that’.<sup>22</sup> Similarly, *chhiò kóng* ‘laugh-say’ in (9) could be interpreted as either ‘laughed (at him) saying’ or ‘mocked (him) as’:

- (8) 伊 干單 唸 講,  
*i kan-na liām kóng,*  
 3SG just nag SAY<sub>THAT</sub>
- 阿和 趁 足濟 錢 ,  
*á-hô thàn chiok-chē chí*  
 NAME earn much money  
 ‘He kept insisting that A-Ho earns a lot of money.’ (*Fate* 455-456)
- (9) 遐 個 <MC: 敵對> 個 武將 共 笑 講  
*Hia ê <MC: didui> ê búchiōng kah chhiò kóng,*  
 that CL opposing LIG general OM:3SG laugh SAY<sub>THAT</sub>
- 這 是 號作 <J: 猴面 冠者>.  
*che sì hò-chò <J: Saru-men Kanja >.*  
 this be name-as monkey-face youngster

‘Those generals who opposed him mocked him (General Toyotomi) as the so-called “monkey-face boy”.’ [*Japanese tales* 629-630]

The compositional semantics of speech act verb + *kóng* reflects the looser syntactic bonding of the unit: in the corpus, we can find alternative (non-grammaticalized) expressions such as *m̄ng lāng kóng* 問儂講 ‘ask people saying’ [*Jesse*, line 670] introducing a quotation in a more prolix fashion than *m̄ng kóng* 問講 ‘ask saying’ or ‘ask that’. Unlike speech act verbs of Stage II, in the next stage of grammaticalization proposed, the combination of cognition verb + complementizer *kóng* is closely bonded: We do not find circumlocutions such as *siū<sup>n</sup> sím-mìh tāichì kóng* \*想啥物代誌講 ‘think-what-matter-say’, even though *siū<sup>n</sup> sím-mìh tāichì* 想啥物代誌 ‘think-what-matter’ ‘think about something’, without *kóng*, is possible. Furthermore, if the speaker self-corrects, the whole unit of  $V_1 V_2 >_{\text{COMP}}$  will be repeated in such a self-initiated repair and not just one of the constituents. For example, *siū<sup>n</sup>-kóng* 想講 ‘think that’ ends the interrupted intonation unit in line 367, followed by a reprise at the beginning of line 368 of the narrative *Fate*.

Hence, as foreshadowed in the discussion in §3.3, a further abstraction of meaning can be clearly identified for the switch context of Stage III when *kóng* is used with cognition verbs

<sup>22</sup> In Hmong, a similar phenomenon has been observed by C. Li (1988, and pers. comm.) whereby the verb ‘say’ remains ambiguous between the two readings of a complementizer and a verb of saying with speech act verbs such as ‘tell’, ‘recount’, ‘state’ and ‘proclaim’.

including *siū<sup>n</sup>* 想 ‘think’, *siau siū<sup>n</sup>* 數想 ‘intend that’ or *m̄-chai (iá<sup>n</sup>)* 唔知(影) ‘not know’, all of which occur in the corpus. The constructional semantics of complementation generalizes from the literal meaning of verbs in the SAY class, which belongs to the perceptible domain of speech in the physical world, to the domain of THINK verbs, and thus the representation of knowledge. This now involves the internal world of the speaker. Hence, for the collocation of VERB<sub>THINK</sub> + *kóng*<sub>COMP</sub>, the meaning of *kóng* as ‘say’ is hardly plausible as an interpretation — ‘?I think saying’ or ‘?I don’t know saying’:

- (10) 爸:                    啊     我,  
 (Father:)                *a*     *goá*,  
                               PRT     1SG
- 想     講     我     唔     識     做     善事  
*siū<sup>n</sup>*   *kóng*   *goá*   *m̄*   *bat*   *chò*   *siān-sū*  
 think   SAY<sub>THAT</sub>   1SG   NEG   EVD   do   good-deed

Father: ‘Me, - I thought that I’d never done any good deeds.’ (*Fate*: 81-82)

In support of establishing this stage in the grammaticalization process, note that as a full-fledged complementizer, *kóng* may not take any kind of aspectual marking, adverbial modification, nor even a preceding prepositional phrase containing the addressee, that is, anything that could be inserted between it and the preceding verb: We neither find such marking in the texts nor do native speakers accept such sentences. For example, (10) above cannot be rearranged to insert a negative marker between *siū<sup>n</sup>* ‘think’ and *kóng* ‘SAY<sub>THAT</sub>’: \**siū<sup>n</sup>-m̄-kóng* ‘think-NEG- SAY<sub>THAT</sub>’ to produce ‘I don’t think that...’. The negative marker has to precede the verb complex, as in examples (11) *mbiēn kóng* NEG-use-SAY<sub>THAT</sub> ‘there’s no need to’ and also (21) for Taiwanese Mandarin. Reinforcing this state of affairs, no element which modifies the whole  $V_1V_2 >_{COMP}$  complex, such as a postverbal aspect marker or discourse particle, may intervene between  $V_1V_2 >_{COMP}$  and the embedded clause which follows. These features point to a clear loss of verbal characteristics for  $V_2$ .

A second pertinent feature relates to a tone sandhi phenomenon. In most cases in the three texts, complementizer *kóng* occurs at the end of an intonation unit containing the matrix clause. We would expect juncture or isolation tone in this position, that is, high falling 51, as it is not inside the tone sandhi domain but rather at its edge. In this changed function, however, the complementizer displays its context tone of high level 55, used inside an intonation group. I suggest that there is a discourse and syntactic reason for this: the tone sandhi is used to signal there is more to come, namely the embedded clause introduced by *kóng*. This nicely reflects the clause-linking function of *kóng* and the fact that the utterance is syntactically incomplete without the following clause. A similar phenomenon is reported for tone raising in Kisi (Atlantic, Niger-Congo) in Güldemann (2001: 165).

If this changed tone feature becomes invariant, for which the data is merely suggestive, it could be seen as an indication of the phonological reduction typically associated with grammaticalization processes (§3.1.3), realised in this case as a reduction and ‘obligatorification’

of tonal possibilities.<sup>23</sup> In other words, once the new grammatical meaning becomes conventionalized, this might be accompanied by a fixing of tonal properties to the invariant one of high level with the application of any tone sandhi rules thus becoming redundant, as, for example, is already the case for the diminutive suffix *-à* 仔 (invariably high falling tone) and for *kóng* in sentence-final position as a discourse marker (see Simpson and Wu 2002).

A further reflection of the change in the degree of grammaticalization between Stage II and Stage III regards the different syntactic behaviours of speech act versus cognition verbs. This is suggested by frequencies of use for verbs with and without a complementizer in the texts. As Table 2 shows, *kóng* occurs as a quotative verb meaning ‘say’ 88 times across the three texts in the corpus for Southern Min without a complementizer, and only 10 times with one (Table 3); *m̄ng* ‘ask’ is found 16 times without a complementizer and 4 times with *kóng*. In contrast with these two speech act verbs, the cognition verb *siū<sup>n</sup>* ‘think’ is more likely to occur with complementizer *kóng*, than without it: 18 versus 10.<sup>24</sup> Given the low numbers, we merely observe this as a tendency which would need to be verified by a larger corpus of data.

A fourth putative stage is where the complementizer increases its scope to emotion verbs. The data on *kóng* from younger generation speakers collected by Liu (1996) on Taiwanese Southern Min includes precisely this class of emotion verbs including *hoân-ló kóng* 煩惱講 ‘to be worried that’ and *hod<sup>n</sup>-hí kóng* 歡喜講 ‘to rejoice, to be happy that’.<sup>25</sup> I had just one example of this group, however, in my Southern Min corpus, namely, *chin gao kóng* 真賢講 ‘very clever SAY<sub>COMP</sub>’ in the context: ‘These historians are very clever in creating them too’ referring to myths surrounding the genesis of nations (*Japanese Tales* 330-332).

A fifth putative stage in the grammaticalization process is also attested in the corpus. This is the extension to modal verbs and hypotheticality. In the Southern Min corpus, there are just five examples where *kóng* occurs with modals, namely *beh* ‘want that’ (1), *m̄-biēn* ‘don’t need to’ (1), and *e-sài kóng* ‘it is possible that’ (3).<sup>26</sup> Here is one such example with *m̄-biēn* ‘don’t need to’:

---

<sup>23</sup> Tone sandhi operating on *kóng* was randomly checked by a native speaker. One of the anonymous reviewers pointed out that the pertinent issue here might rather be the shifting of tone sandhi domains (from the *V<sub>1</sub>-kóng* cluster to the *kóng* + complement clause) and that there is in fact no clear case of phonological reduction, since the usual tone sandhi rules would apply in any case under this interpretation. Nonetheless, such an analysis would imply that the intonation break between *V<sub>1</sub>-kóng* cluster and the complement clause, originally signalling the onset of a quotation, has become blurred, if not disappearing completely, with the complementizer now being reinterpreted as head of the lower complement clause, for which, indeed, its tonal properties would accord (that is, bearing the changed tone of high level). This stage of grammaticalization is not however reflected in the data, and hence the changed tone phenomenon remains in need of explanation since it occurs non-canonically in a juncture position.

<sup>24</sup> Excluded are 12 examples that show other meanings of *siū<sup>n</sup>* such as ‘plan’ or ‘think over’ and which only take object NPs.

<sup>25</sup> This class of verbs is not well represented in my data, and may simply be an accidental gap. As Liu (1996) does not give the romanization for her Southern Min examples, I will not reproduce them in full here. See also the notes on the use of characters and romanizations at the end of this article.

<sup>26</sup> One of the anonymous reviewers made the challenging comment that the use of emotion and modal verbs with *kóng* looks like a case of non-finite complementation and consequently that *kóng* is a complementizer which is able to take both finite and non-finite clauses in its complement, another feature distinguishing it from English *that*. However, I note that the example given in (11) contains a subjectless complex construction in the complement clause, subjectless due to ellipsis of (or zero anaphoric reference to) the preceding, understood 3SG ‘he’, and that it also contains a modal verb: *beh* ‘want’. Thus, I would treat it as finite, while remarking that the finite–non-finite distinction remains a complex and unresolved issue in Chinese linguistics (cf. Hu et al 2001). The question of same and switch subject between matrix and complement clause would need to be taken into account.

(11)	擔	有				
	<i>tān</i>	<i>ū</i>				
	if	have				
	汝	講	伊	麼	免	講
	<i>li</i>	<i>kóng</i>	<i>i</i>	<i>mā</i>	<i>mbién</i>	<i>kóng</i> ,
	2SG	say	3SG	also	NEG:need	SAY <sub>THAT</sub>
	欲	納	錢	着	煩惱	啊
	<i>beh</i>	<i>lap</i>	<i>chí<sup>n</sup></i>	<i>təh</i>	<i>hoān-ló</i>	<i>a</i>
	want <sub>FUT</sub>	pay	money	then	worry	PRT

‘If he has (earned a lot of money), you know that there’s no need for him to worry when it’s time to pay the money [for a housing loan].’ (*Fate* 476-478)

Despite the small number of examples with modal verbs used with *kóng* in the corpus, these data concur with many other crosslinguistic studies on complementation such as Frajzyngier’s on Mupun, a Chadic language of Nigeria. Frajzyngier observes that the complementizer usage of demonstrative *nə* is principally found with *verba dicendi* and verbs of cognition. Nonetheless, when *nə* occurs with other verb classes such as modals, the complex sentence so formed may take on the specific modal meanings of a wish or an obligation (1991: 229).

While the close interaction between complementation, the irrealis mood and verb classes coding speech acts, cognition, fear and desire is not surprising from the point of view of the subjunctive mood found in many European languages, this is the first time this phenomenon has been described in the case of Sinitic languages.

#### 4.3. Hong Kong Cantonese *wa<sup>6</sup>* 話

In this section, I compare the use of complementizer *wa<sup>6</sup>* 話 in Hong Kong Cantonese with that of *kóng* in Southern Min on the basis of a quantitative study of discourse materials. These data reveal in a striking way the difference in the degree of grammaticalization between Southern Min and Cantonese for this complementizer pathway. Cantonese is at an earlier stage than Southern Min, as indicated in previous sketches of this phenomenon (Chappell 1997, 2001b).<sup>27</sup>

##### 4.3.1. Comparison of complementizer usage in Taiwanese Southern Min and Hong Kong Cantonese Yue discourse

The main speech act verb in Hong Kong Cantonese that can introduce a quotation is *wa<sup>6</sup>* 話 ‘say’. In conversational and narrative texts, *wa<sup>6</sup>* 話 ‘say’ functions mainly as a quotative marker

<sup>27</sup> According to Matthews & Yip (1994: 308), Cantonese does not possess any element similar in function to the complementizer *that* in English, although they do observe that there is a redundant use of *wa<sup>6</sup>* 話 ‘say’ in a serial verb construction where it combines with *gong<sup>2</sup>* 講 ‘talk’. The *wa<sup>6</sup>* 話 I identify in this construction is the semi-complementizer, described in this section.

with speech act verbs, but also less frequently as a semi-complementizer and as an irrealis marker.

In my data from Hong Kong Cantonese, the most common nonlexical use of *wa*<sup>6</sup> as a semi-complementizer is represented by fourteen instances where it introduces direct speech in combination with speech act verbs such as *jaan*<sup>3</sup> ‘praise’ and *gong*<sup>2</sup> ‘tell’ which notably precede it, but are not adjacent to it (14/55). Table 4 gives the classification and frequency of uses of *wa*<sup>6</sup> as seen in a small corpus of two narrative texts which I recorded in Hong Kong.

Table 4: Frequency and uses of *wa*<sup>6</sup> in two Cantonese narratives

NARRATIVE	<i>Tale of the Reborn Lady at the Red Flowering Plum</i> 再世紅梅記	<i>Balcony Rendezvous</i> 樓台會	TOTALS
<b>Length in Minutes</b>	7:24	13:42	21:06
<b>N<sup>o</sup> Intonation units</b>	416	754	1,170
Lexical: Quotative verb (NP) <i>wa</i> <sup>6</sup> : “....”	18	17	35
Semi-complementizer: (NP) VERB <sub>COMM</sub> X <i>wa</i> <sup>6</sup> : “....”	6	8	14
Complementizer: VERB <sub>gong<sup>2</sup></sub> - <i>wa</i> <sup>6</sup> : “....”	1	1	2
Irrealis: ‘supposedly’	0	2	2
Indeterminate or false starts	1	1	2
<b>TOTAL</b>	<b>26</b>	<b>29</b>	<b>55</b>

In fact, in several examples, such as the one reproduced below with *hei*<sup>1</sup>*pin*<sup>3</sup> ‘deceive’, semi-complementizer *wa*<sup>6</sup> was separated from the speech act verbs either by a noun denoting the direct object in the syntactic structure: (NP<sub>SUBJ</sub>) VERB<sub>SPEECH ACT</sub> NP<sub>DO</sub> *wa*<sup>6</sup> QUOTATION, or by a phrase introducing the addressee: *jau*<sup>6</sup> *gong*<sup>2</sup> *bei*<sup>2</sup> *hui*<sup>5</sup> *teng*<sup>1</sup> *wa*<sup>6</sup> 就講畀佢聽話 then-tell-[GIVE<sub>TO</sub>-3SG-listen]-say ‘then told him’ [Reborn 202]. ‘X’ is used to represent these two possibilities in the syntactic configuration:

- (12) Semi-complementizer usage of *wa*<sup>6</sup> in Cantonese:  
(NP<sub>SUBJ</sub>) VERB<sub>SPEECH ACT</sub> X *wa*<sup>6</sup> – QUOTATION:

72. ... 你 就 = ,  
... *lei*<sup>5</sup> *jau*<sup>6</sup> = ,  
2SG then

73. ... 欺騙 我 話 你 係 一 個 = ,  
 ... hei<sup>1</sup>pin<sup>3</sup> ngoh<sup>5</sup> wa<sup>6</sup> lei<sup>5</sup> hai<sup>6</sup> yat<sup>1</sup> goh<sup>3</sup> = ,  
 deceive 1SG say 2SG be one CL
74. ... eh 已經 ,  
 ... eh yi<sup>5</sup>ging<sup>1</sup> ,  
 already
75. .. 出 - 出嫁 嘅 =  
 .. chut<sup>1</sup> - chut<sup>1</sup>ga<sup>3</sup> ge<sup>3</sup> =  
 out marry:a:man GE<sub>ASST</sub>
76. ... 婦人。  
 ... foo<sup>5</sup>yan<sup>4</sup> .  
 woman

‘You deceived me by saying that you were already a marry- married woman.’  
 (*Tale of the reborn lady at the red flowering plum* 63-76)

This corresponds to the bridging context or semi-complementizer Stage II where the morphosyntactic relationship between V<sub>1</sub> and V<sub>2</sub> is relatively unbonded. The other main verbs with which it co-occurred in this quotative structure are presented in Table 5:

Table 5: Speech act verbs followed by the semi-complementizer wa<sup>6</sup> 話 in the structure (NP<sub>SUBJ</sub>) VERB<sub>SPEECH ACT</sub> X wa<sup>6</sup> 話 in two Cantonese texts

wa <sup>6</sup> ... wa <sup>6</sup> 話 ... 話 ‘to say ... saying/that’	4
gong <sup>2</sup> ... wa <sup>6</sup> 講 ... 話 ‘to tell ... saying/that’	4
man <sup>6</sup> ... wa <sup>6</sup> 問 ... 話 ‘to ask ... saying/that’	1
hei <sup>1</sup> pin <sup>3</sup> ... wa <sup>6</sup> 欺騙 ... 話 ‘to deceive ... saying/that’	1
bo <sup>3</sup> ... wa <sup>6</sup> 報 ... 話 ‘to report ... saying/that’	1
jaan <sup>3</sup> ... wa <sup>6</sup> 讚話 ‘to praise ... saying/that’	1
aat <sup>3</sup> wa <sup>6</sup> 噏 ... 話 ‘to protest ... saying/that’	1
ying <sup>1</sup> sing <sup>4</sup> wa <sup>6</sup> 應承 ... 話 ‘to acquiesce ... saying/that’	1
<b>TOTAL</b>	<b>14</b>

The restriction of wa<sup>6</sup> to co-occurrence largely with verbs of communication or cognition is independently confirmed by two other studies on Hong Kong Cantonese, Yeung (2003) and Chui (1994: 3), both of whom make use of large corpora of telephone conversations, radio interviews, and phone-in programs. Thus, the Cantonese corpora studied by Yeung (2003, 2006) contain two main verb classes that co-occur with wa<sup>6</sup>: communication and information (20/25 different types of verbs) as well as cognition and perception (5/25) but, interestingly, no verbs of emotion as attested for Southern Min, such as *hoânló kóng* 煩惱講 ‘be anxious that’ and

*hoa<sup>n</sup>hi kóng* 歡喜講 ‘to rejoice that’.<sup>28</sup> Hence, an observation highly revealing for this crosslinguistic study of Sinitic languages is that while Yeung (2003) and Chui (1994) present an entirely similar range of verb classes to those found in my own smaller Cantonese corpus, taken together, they represent merely a subset of those attested for Taiwanese Southern Min and discussed above in detail in §4.2.

Nonetheless, there are also two instances of stative verbs in the database used by Yeung which need to be accounted for, namely, *gaan<sup>1</sup>naan<sup>4</sup> waa<sup>6</sup>* ‘to be difficult’ and *dang<sup>2</sup>jyu<sup>4</sup> waa<sup>6</sup>* ‘to be equal’.<sup>29</sup> The example with the stative verb ‘to be difficult’ is reproduced in (13). It occurs with a small clause following *waa<sup>6</sup>*, suggesting that it is really closer to an irrealis marker rather than a complementizer with a clause-linking function, if not to the meaning of ‘predict’ or ‘judge’ as in the interpretation ‘I think it is very difficult to say it will (happen)’:

- (13)
- |                       |                       |                                       |                        |                        |                        |                        |                                         |                        |                        |
|-----------------------|-----------------------|---------------------------------------|------------------------|------------------------|------------------------|------------------------|-----------------------------------------|------------------------|------------------------|
| 即                     | 個                     | 現實                                    | 我                      | 諗                      | 都                      | 好                      | 艱難                                      | 話                      | 會                      |
| <i>ze<sup>1</sup></i> | <i>go<sup>3</sup></i> | <i>jin<sup>6</sup>sat<sup>6</sup></i> | <i>ngo<sup>5</sup></i> | <i>lam<sup>2</sup></i> | <i>dou<sup>1</sup></i> | <i>hou<sup>2</sup></i> | <i>gaan<sup>1</sup>naan<sup>4</sup></i> | <i>waa<sup>6</sup></i> | <i>wui<sup>5</sup></i> |
| that:is               | CL                    | fact                                  | 1SG                    | think                  | all                    | very                   | difficult                               | say                    | will                   |
- ‘that is, I think in reality that would hardly be true.’ (HKCAC)

I next discuss syntactic features for Cantonese *wa<sup>6</sup>*. In my data, there are just two occurrences of the complementizer usage where *wa<sup>6</sup>* formed a verb complex with its preceding speech act verb (2/55, as per Table 4 above), unlike Southern Min where this usage is the rule in the discourse data collected (§4.2). The two Cantonese examples were both exemplified by the same verb complex *gong<sup>2</sup>-wa<sup>6</sup>* 講話 ‘tell-say’. Note that Cantonese *gong<sup>2</sup>* 講 is the cognate form to Southern Min *kóng*, but has not been singled out for grammaticalization into a complementizer, the ‘fate’ rather of Cantonese *wa<sup>6</sup>* 話 ‘say’:

- (14) Complementizer usage of *wa<sup>6</sup>* in Cantonese
- |                        |                        |                        |                                          |                       |   |  |  |  |
|------------------------|------------------------|------------------------|------------------------------------------|-----------------------|---|--|--|--|
| 噉                      | 個                      | 裴生                     | 呢,                                       |                       |   |  |  |  |
| <i>gam<sup>2</sup></i> | <i>goh<sup>2</sup></i> | <i>goh<sup>3</sup></i> | <i>Pooi<sup>4</sup> Sang<sup>1</sup></i> | <i>le<sup>1</sup></i> | , |  |  |  |
| so                     | that                   | CL                     | Pooi Sang                                | P <sub>TOP</sub>      |   |  |  |  |
- 
- |                        |                        |                        |                        |                                          |                                        |                         |                       |    |
|------------------------|------------------------|------------------------|------------------------|------------------------------------------|----------------------------------------|-------------------------|-----------------------|----|
| 就                      | 係                      | 哩                      | 個                      | 丞相                                       | 面前                                     | 講                       | 話                     | =, |
| <i>jau<sup>6</sup></i> | <i>hai<sup>2</sup></i> | <i>lei<sup>5</sup></i> | <i>goh<sup>3</sup></i> | <i>sing<sup>4</sup>seung<sup>3</sup></i> | <i>min<sup>6</sup>chin<sup>4</sup></i> | <i>gong<sup>2</sup></i> | <i>wa<sup>6</sup></i> | =, |
| then                   | be:at                  | this                   | CL                     | chief:minister                           | face-front                             | tell                    | SAY <sub>THAT</sub>   |    |

<sup>28</sup> Emotion verbs form part of the corpus of naturally occurring examples in Southern Min collected by Liu (1996). The corpora used by Yeung are equivalent to 8 hours for one corpus (HKCAC corpus: phone-in programs on radio) and a total of 43,283 lines of text for the other (The Cantonese radio corpus). These databases also include interesting cases of code-mixing whereby *waa<sup>6</sup>* co-occurs with English verbs: *encourage X waa<sup>6</sup>* and *expect waa<sup>6</sup>*. The corpora used by Chui (1994: 2) include 10 telephone conversations and data from TV programs about daily life (no quantification given).

<sup>29</sup> I am indebted to Ka-Wai Yeung for making these data available to me. While a large number of modal verbs combine with *waa<sup>6</sup>* in the database, this is not a case of the complementizer usage, but rather that of an irrealis marker, mirroring the situation for vernacular Beijing Mandarin. In contrast to this, emotion verbs do not occur in the database with *waa<sup>6</sup>*. However, examples such as *fun<sup>1</sup>hei<sup>2</sup> waa<sup>6</sup>* 歡喜話 ‘be happy that’ or 煩惱話 *faan<sup>4</sup>nou<sup>5</sup> waa<sup>6</sup>* ‘be worried that’, can be heard in use by certain generations, namely younger speakers, and in certain areas of Hong Kong (Ka-Wai Yeung, pers. comm.). One of the anonymous reviewers similarly points out that *daam<sup>1</sup>sam<sup>1</sup> wa<sup>6</sup>* 擔心話 ‘to be worried that’ is possible and consequently is of the view that this difference in verb classes may boil down to a matter of frequency. Further research with larger corpora is called for.

<Q	你	千	祈	唔	好	呀	,
<Q	lei <sup>5</sup>	chin <sup>1</sup>	kei <sup>4</sup>	m <sup>4</sup>	ho <sup>2</sup>	a <sup>3</sup>	,
	2SG	thousand:	pray	NEG	good	PRT	

‘So he, Pooi Sang, said in front of the chief minister (that): “You mustn’t do this ...’  
(*Tale of the reborn lady at the red flowering plum* 222-224)

For Cantonese, the looser combination of the speech act or cognition verb with its following complementizer *waa*<sup>6</sup> (= *wa*<sup>6</sup>) is already evident in the fact that a direct object, as in (12) above, or aspectual modification may intervene, as shown in example (15) with the experiential aspect: V<sub>1</sub>-Aspect-*waa*<sup>6</sup>. As pointed out in §4.2 above, there were no such examples with aspectual modification occurring between V<sub>1</sub> and V<sub>2</sub> in my Southern Min texts:

(15)	我	有	諗	過	話	你	儲	不如	擺
	ngo <sup>5</sup>	jau <sup>5</sup>	lam <sup>2</sup> -gwo <sup>3</sup>	waa <sup>6</sup>	lei <sup>5</sup>	cou <sup>5</sup>	bat <sup>1</sup>	jyu <sup>4</sup>	baai <sup>2</sup>
	1SG	have	think-ASP	SAY <sub>COMP</sub>	2SG	save:up	why:	not	put
	喺	我	度	啦	咁	我	keep-住 [...]		
	hai <sup>2</sup>	ngo <sup>5</sup>	dou <sup>6</sup>	laa <sup>1</sup>	gam <sup>2</sup>	ngo <sup>5</sup>	keep-zyu <sup>6</sup> [...]		
	at	1SG	place	SFP	so	1SG	keep-ASP		

‘I thought that if you save up, why don’t we put the money at my place and I would look after it [...]’ (Yeung 2006: 32; example taken from Cantonese Radio Corpus)

In fact, 14/34 examples in Yeung’s appendix show either a direct object, an aspect marker or a modal particle intervening between the V<sub>1</sub> (= speech act or cognition verb) and the V<sub>2</sub> = *wa*<sup>6</sup>. Of the 20 examples of V<sub>1</sub>V<sub>2</sub> complexes, only four contain cognition verbs as V<sub>2</sub>. This allows us to conjecture that Cantonese is just arriving at the switch context of Stage III in the creation of a complementizer, given that the combined databases represent a well-balanced sample of actual usage in terms of both genre and size.

A further difference concerns the use of the lexical verb *wa*<sup>6</sup> ‘say’ with its related complementizer. Unlike Southern Min, it cannot directly precede it, according to Yeung (2006: 16). Instead, the two instances of *wa*<sup>6</sup> are separated by aspectual or other kinds of particles, if not a pause.<sup>30</sup> Below, one of her examples is reproduced with an intervening particle, *le*<sup>1</sup>, as representative of the constraint. This could also be realised by a brief pause:

(16)	佢	話	呢	話	[ <sub>IP</sub> 唔	得閒	嚟]
	keoi <sup>5</sup>	waa <sup>6</sup>	le <sup>1</sup>	waa <sup>6</sup>	[ <sub>IP</sub> m <sup>4</sup>	dak <sup>1</sup>	haan <sup>4</sup>
	3SG	say	PRT	WAA	[ <sub>IP</sub> NEG	free	come]

‘(S)he says, that (s)he’s not free to come.’

<sup>30</sup> The V<sub>1</sub>V<sub>1</sub> structure of *wa*<sup>6</sup> *wa*<sup>6</sup> is however possible when *wa*<sup>6</sup> means ‘to blame’. See example (25) in Yeung (2003).

In contrast to this, the majority of the examples of *kóng kóng* in the Southern Min texts, that is, eight out of ten (8/10), are found in the same intonation unit and form a  $V_1V_2$  complex without any intervening elements.<sup>31</sup>

(17) 恁 尅叔仔 共 我 講 講 ,  
*lín ban-chek-à kah goá kóng kóng*  
 2SG:PL youngest:uncle COM 1SG say SAY<sub>THAT</sub>

我 還也 有 做 善事 啦 .  
*goá oân-á ū chò siān-sū là*  
 1SG also have<sub>PFT</sub> do good-deed PRT

‘Your youngest uncle told me that I had also done some good deeds.’ (*Fate* 77-78)

Note that there are no tone sandhi rules affecting the Cantonese (semi-)complementizer *wa*<sup>6</sup>, (see Chui 1994), unlike the case in Southern Min where it potentially indicates this new grammatical function of *kóng* (cf. §4.2 on tone sandhi from high falling to the high level tone *kong*).

Hence, insofar as the databases are representative of the spoken language, Cantonese Yue appears to be in the early stages of developing its complementizer use of the verb *wa*<sup>6</sup>, straddling the bridging and context stages II and III, given that combination into a single but still separable verb unit is only just emerging with cognition verbs (that is, Stage III). The main evidence supporting this analysis regards the possibility of inserting material between the  $V_1$  and *wa*<sup>6</sup> ‘say’, that is, elements such as direct objects and addressee PPs, or aspect and pause particles, described just above. The younger age of the Cantonese complementizer is also mirrored in the restriction on the co-occurrence of semi-complementizer *wa*<sup>6</sup> to mainly speech act and cognition verbs. In spite of this, Cantonese also exemplifies an intriguing case of polygrammaticalization of SAY since other pathways are in evidence such as the clause-final usage of *wa*<sup>2</sup> (< *wa*<sup>6</sup>) in echo questions,<sup>32</sup> treated in Chui (1994) and the formation of complex markers of subordination, including the conditional (Hwang 1998: §2.2).

## 5. Complementizers in other Sinitic languages

The section aims to give a brief presentation of what is known about complementizers in other Sinitic languages, based on either published studies using discourse data for two non-standard spoken varieties of Mandarin, namely Beijing and Taiwanese Mandarin, or on my own research using a database of transcribed oral texts for Xiang, Gan and Wu, if not on published narrative materials in the case of Hakka and Jin.

First, Standard Mandarin *shuō* 說 ‘say’ is compared with the cognate verbs used in two other important varieties of Mandarin, Beijing and Taiwanese, turning up an extremely

<sup>31</sup> For the remaining 2/10 examples of *kóng kóng* 講講 ‘say SAY<sub>THAT</sub>’, one instance involved a break in the intonation unit and the other an interruption by another speaker, followed by a subsequent resumption of the turn.

<sup>32</sup> The clause-final use of *wa*<sup>6</sup> in echo questions obligatorily has a different tone, namely, high rising *wa*<sup>2</sup> which may be the result of its interaction with the rising intonation contour of echo questions (see Matthews & Yip 1994: 318-9).

interesting contrast in usage, whereby the two latter non-standard varieties show the complementizer usage. Second, developments in the use of the quotative verb *koŋ*<sup>51</sup> 講 as a semi-complementizer in Sixian Hakka are also briefly described.

Finally, an overview is given for Sinitic languages in which SAY verbs do not appear to have undergone any of the grammaticalization processes, reporting on a larger discourse study of SAY verbs in Chappell (in prep.). These are the Changsha dialect of Xiang, the Nanchang dialect of Gan, the Shanghainese dialect of Wu and the Huojia dialect of Jin.

### 5.1. Standard Mandarin or *pǔtōnghuà*

*Pǔtōnghuà* 普通话, ‘the common language’ of China, is a variety of Northern Mandarin, modelled on the pronunciation of educated Beijing speech, the lexicon of Northern Chinese dialects and the grammar of modern vernacular literature (Chen 1999). As the official language of China, it is taught in the education system and used in the media, among other domains, and is intended to serve as a kind of *lingua franca* transcending the ensemble of Chinese dialects and languages. It can be viewed as a model for the formal register of the Northern Mandarin dialect group. Since Standard Mandarin is frequently chosen as the representative language for the Sinitic taxon in numerous typological studies, including the *World Atlas of Language Structures* (Haspelmath et al, 2005), it thus deserves inclusion for treatment here.

The main verb of saying is *shuō* 說 in Standard Mandarin. Although this verb can be optionally used to introduce a direct or indirect quote, as described in Li and Thompson (1981: 601–602), it does not appear to be used as a full-fledged complementizer, according to other studies of its polyfunctionality such as those in Meng (1982) and Y. Liu (1986). Reference grammars such as Chao (1968: 108–110) do not mention the phenomenon either, while analyses such as Cheng (1991) and Paris (1996: 34) support the view of zero marking, as far as subordinating conjunctions are concerned. Adding further grist to the mill, if one elicits a translation of the Southern Min text given in (10), and presented in (18) below, native speakers of Mandarin will not translate the complementizer *kóng* by its functionally equivalent lexical verb, *shuō* ‘say’. In fact, they do not translate it at all:

#### (18) Standard Mandarin translation of a Southern Min discourse example

我	認爲	(*說)	我	從來	沒	做過	善事
<i>Wǒ</i>	<i>rènwei</i>	(* <i>shuō</i> ),	<i>wǒ</i>	<i>cónglái</i>	<i>méi</i>	<i>zuò-guo</i>	<i>shàn-shì</i> .
1SG	think	(*SAY <sub>THAT</sub> )	1SG	never	NEG	do-EVD	good-matter

‘I thought I’d never done any good deeds.’ (*Elicited example*)<sup>33</sup>

To probe this issue a little further, let us first examine some data from a contemporary novel from China, written in a colloquial style, where non-prescriptive uses may inadvertently turn up. This is the novel *Huózhe* 活著 [To live] by Yu Hua (1994) which contains a large proportion of dialogue.

The following example is instructive because it contrasts the use of direct and indirect speech in the words of Jia Zhen, the wife of the storyteller, Fugui. This relates to the period of the

<sup>33</sup> This section on Standard Mandarin, including the examples, was discussed at length with students and colleagues in my doctoral seminar at the Ecole des Hautes Etudes en Sciences Sociales in 2007. With a common will, the three native speakers in this class produced the translation in (18) as being the most felicitous, thus conforming appropriately to the norms of usage for an official language.

communes and production teams in China, mid-20<sup>th</sup> century, when peasants were allocated points for work achieved, according to the heaviness of the manual labour. Jia Zhen in the past has normally received 8 points but now as she becomes daily weaker with an unidentified illness, she has been allocated only 4.

(19)	她	總	覺得	自己	還	能	干	重	活,
	<i>Tā</i>	<i>zǒng</i>	<i>juéde</i>	<i>zìjǐ</i>	<i>háí</i>	<i>néng</i>	<i>gàn</i>	<i>zhòng</i>	<i>huó,</i>
	3SG	always	think	self	still	can	do	heavy	work
	幾次	都	去	對	隊長	說，			
	<i>jǐcì</i>	<i>dōu</i>	<i>qù</i>	<i>duì</i>	<i>duìzhǎng</i>	<b>shuō<sub>1</sub>,</b>			
	several-times	all	go	to	team-leader	<b>say<sub>1</sub></b>			
	說	她	也	知道	自己	有	病,		
	<b>shuō<sub>2</sub></b>	<i>tā</i>	<i>yě</i>	<i>zhīdao</i>	<i>zìjǐ</i>	<i>yǒu</i>	<i>bìng,</i>		
	<b>say<sub>2</sub></b>	3SG	also	know	self	have	illness		
	可	現在	還	能	干	重	活,	說:	
	<i>kě</i>	<i>xiànzài</i>	<i>háí</i>	<i>néng</i>	<i>gàn</i>	<i>zhòng</i>	<i>huó,</i>	<b>shuō<sub>3</sub>:</b>	
	but	now	still	able	do	heavy	work	<b>say<sub>3</sub></b>	
	«等	我	真	干不動	了	再	給		
	“ <i>děng</i>	<i>wǒ</i>	<i>zhēn</i>	<i>gàn-bu-dòng</i>	<i>le</i>	<i>zài</i>	<i>gěi</i>		
	wait	1SG	really	do-NEG-move	PFT	again	give		
	我	記	四	分	吧.»				
	<i>wǒ</i>	<i>jì</i>	<i>sì</i>	<i>fēn</i>	<i>ba.</i> ”				
	1SG	record	four	points	PRT				

‘She continued to believe that she could do heavy work and even went several times to talk to the brigade leader (saying) that she knew she was ill but could still do heavy work. She said: “Wait until I’m really unable to do it, before you record four points for me again.”’ (Yu 1994: 152, my translation and glossing – HC)

The first use of ‘say’ is the lexical verb use ‘to talk’ while the second and third uses introduce indirect and then, direct speech, respectively. The second occurrence of *shuō* looks like a partially grammaticalized use of ‘say’, similar to the bridging stage II of the semi-complementizer: (*tā*) ... *duì duìzhǎng shuō<sub>1</sub>, shuō<sub>2</sub>* ‘she talked to the team leader, saying...’. Could it possibly be indicative of the first signs of a similar development of *shuō* in Standard Mandarin into a complementizer, despite the fact that native speakers spontaneously remark upon the oddness of this usage.<sup>34</sup>

<sup>34</sup> Unbidden, several members of seminar audiences on this topic who were native speakers of Mandarin, queried the use of *shuō<sub>2</sub>* which seemed strange to them. It is interesting that their objection appears to uphold prescribed usage for standard Mandarin grammar; and, as it turns out, not to conform at all with trends in the spoken variety of Beijing Mandarin where the complementizer usage abounds (see Fang, 2006 and discussion in §5.2). Ka-Wai Yeung

Since the grammar of Standard Mandarin is codified as being based on the literature of Northern Mandarin, this reaction is not surprising from a stylistic point of view. Hence, it appears that this syntactic device is not generally recognized as being a permissible one for introducing complement clauses in Standard Mandarin, or *pǔtōnghuà*, as spoken in mainland China. Given the prescriptive norms that are in force for the standard language, data from one of the non-standard Mandarin dialects, Beijing Mandarin and from the regional variety of Taiwanese Mandarin prove revelatory with regard to this matter.

## 5.2. *Beijing Mandarin*

The colloquial variety of Mandarin, as spoken in Beijing, is not identical to *pǔtōnghuà*, or Standard Mandarin, as several studies have recently illustrated in terms of its grammar, for example, Chirkova & Lamarre (2005) or for its phonology (Chen 1999: chapter 3.2).

A new study by Fang Mei (2006) on Beijing Mandarin contrasts in a remarkable way with the paucity of examples of this phenomenon for Standard Mandarin. She shows that Beijing Mandarin uses the verb *shuō* 說 ‘say’ in many nonlexical functions, focussing on its use as a complementizer and as a topic marker in her study. According to Fang, these two main functions can be distinguished according to their distributional patterns and specific syntactic features, for which she uses a corpus of contemporary spoken data as well as historical materials.

Similarly to Southern Min, the complementizer usage occurs in between the matrix clause and the linked clause. In general, it directly follows the main verb which may belong to the speech act or perception domain, to more abstract verb classes such as cognition, or even to emotion verbs such as *dānxīn shuō* 担心说 ‘be worried about’, just as in Southern Min. It does not co-occur with modal verbs, however, as a complementizer. Hence, Beijing Mandarin has clearly reached Stage IV.

### (20) Complementizer usage in Beijing Mandarin:

我 总是 觉得 说, 生活里 缺了 点儿 什么。

wǒ zǒngshì juéde shuō, shēnghuó-lǐ quē-le diǎnr shénme

1SG always feel SAY<sub>THAT</sub> life-in lack-PFV little something

‘I’ve always felt that there is something a little lacking in my life.’ (= example (1) in Fang 2006: 109; my translation and glossing – HC)

From an historical point of view, Fang (2006: 111) points out that, in late 19<sup>th</sup> century literature, the quotative use of *shuō* 說 with speech act and perception verbs can already be observed, reflecting the vernacular language of the capital, Beijing, at the end of the Qing dynasty. Regrettably, no statistics are provided on the different uses.

---

confirmed a similar subjective reaction from her Hong Kong students who found the use of Cantonese *waa*<sup>6</sup> as a complementizer to be ‘improper’ or ‘disfluent’, and certainly not the kind of language that would be found in more formal registers such as news reports or speeches. From a sociolinguistic point of view, this is even more fascinating than the Mandarin case, for as she remarks, Cantonese does not hold the same position as Standard Mandarin as an official language: it is not studied at school in Hong Kong according to any prescriptive grammar.

### 5.3. Taiwanese Mandarin

In another important regional variety of Mandarin, Taiwanese Mandarin, a similar development can be found as to that in Beijing Mandarin whereby the high frequency verb of saying, *shuō* 說, is used as a complementizer in the identical configuration: NOUN<sub>SUBJECT</sub> - VERB<sub>1</sub> - *shuō* + CLAUSE. The formal standard for Taiwanese Mandarin is not exactly the same as that for mainland *pǔtōnghuà* or Standard Mandarin, since the two varieties have diverged somewhat, particularly in the last half of the 20<sup>th</sup> century which witnessed strained, if not hostile, cross-strait relations, and limited contact.

Note that Taiwanese Mandarin does not choose the verb cognate with *kóng*, *jiǎng* 講 in Mandarin, to create the new syntactic category but instead makes use of its functional equivalent, *shuō* 說. An example from conversational discourse follows, reproduced from Huang (2003: 440):

(21) Taiwanese Mandarin:

260.	.. 他們	不	知道	說、		
	<i>tāmen</i>	<i>bù</i>	<i>zhīdao</i>	<i>shuō</i>		
	3PL	NEG	know	SAY <sub>THAT</sub>		
261.	.. 係裏面	最近	有	什麼樣	的	事情。
	<i>xì-lǐmian</i>	<i>zuìjìn</i>	<i>yǒu</i>	<i>shémeyàng</i>	<i>de</i>	<i>shìqing</i> .
	department-inside	lately	have	what:kind	LIG	matter

‘They didn’t know what was happening in the department.’

Although complementizer usage with verb classes is more or less parallel to those found in Southern Min and Beijing Mandarin, Cheng (1991) and Liu (1996), in particular, claim that Taiwanese Mandarin is more conservative in its use of the grammatical functions of *shuō* than Southern Min is for *kóng*, pointing out that bilinguals may even use Southern Min *kóng* as a complementizer with Taiwanese Mandarin verbs that they feel should not collocate with *shuō*; for example, the Mandarin verb *xiǎodé* ‘know’ may be so used: *xiǎodé kóng* 曉得 *kóng* ‘to know that’ in preference to *xiǎodé shuō* 曉得說. According to the data presented in these studies on Taiwanese Mandarin, *shuō* can act as a complementizer to at least the verb classes of speech act, cognition and modality. Thus, it appears to have just reached the conventionalization stage, V, but is not as advanced as Taiwanese Southern Min. Further research will shed more light on this matter.

### 5.3. Sixian Hakka

In the Sixian dialect of Hakka spoken in Taoyuan county in northwestern Taiwan, a semi-complementizer usage of its SAY verb, namely, *koŋ<sup>31</sup>*, cognate to Southern Min *kóng*, is frequently found in use as the second verb in V<sub>1</sub>V<sub>2</sub> series, in 15 stories collected and transcribed by Yang (1957). These data are precious in providing a window on Hakka as spoken 50 years ago, in light of the paucity of discourse materials currently available for this group of dialects.

(22)	就	同	掌牛	阿哥	參詳	講：			
	<i>ts'iu</i> <sup>55</sup>	<i>t'uŋ</i> <sup>11</sup>	<i>tʃoŋ</i> <sup>31</sup> - <i>ŋiu</i> <sup>11</sup>	<i>a</i> <sup>24</sup> - <i>ko</i> <sup>24</sup>	<i>ts'am</i> <sup>24</sup> - <i>sioŋ</i> <sup>11</sup>	<i>koŋ</i> <sup>31</sup> ：			
	then	COM	cowherd	brother	discuss	say			
	“細	阿哥	唉！	你	攪	個	該	尾	蛇
	<i>se</i> <sup>cc</sup>	<i>a</i> <sup>24</sup> - <i>ko</i> <sup>24</sup>	<i>ə</i> <sup>0</sup>	<i>ŋi</i> <sup>11</sup>	<i>kau</i> <sup>31</sup>	<i>kai</i> <sup>0</sup>	<i>kai</i> <sup>55</sup>	<i>mui</i> <sup>24</sup>	<i>ʒa</i> <sup>11</sup>
	little	brother	PRT	2SG	play	LIG	that	CL	snake
	來	賣	分	我	好	麼？”			
	<i>loi</i> <sup>11</sup>	<i>mai</i> <sup>55</sup>	<i>pun</i> <sup>24</sup>	<i>ŋai</i> <sup>31</sup>	<i>ho</i> <sup>31</sup>	<i>mo</i> <sup>11</sup>			
	come	sell	give <sub>DAT</sub>	1SG	good	Q			

‘(The fisherman seeing the snake in tears) then negotiated with the cowherd, (saying):  
 “Hey Little Brother! How about selling me that snake you’re playing around with?”  
 (Yang 1957 : 161, my glossing and translation - HC)

There are numerous examples in these 15 written narratives of *koŋ*<sup>31</sup> as a quotative marker or a semi-complementizer (14/158), introducing either direct or indirect speech, that is, mainly with speech act verbs such as *tap*<sup>2</sup>-*in*<sup>55</sup> *koŋ*<sup>31</sup> 答應講 ‘answer-say’; *ham*<sup>55</sup>-*koŋ*<sup>31</sup> 喊講 ‘call-say’; *mun*<sup>55</sup>-*koŋ*<sup>31</sup> 問講 ‘ask-say’ and *t'aŋ*<sup>24</sup>-*to*<sup>0</sup> *koŋ*<sup>31</sup> 聽倒講 ‘hear-say’. Its semi-complementizer status is seen in the fact that a direct object may be inserted between the verb *mun*<sup>55</sup> ‘to ask’ and the quotative verb *koŋ*<sup>31</sup>, for example. This suggests that the Sixian dialect of Hakka is conservative, remaining at the bridging stage II, and that in the 1950s, this dialect had not advanced as far as Cantonese Yue currently has, that is, by adding cognitive verbs to its scope for complementation. A larger study of the contemporary linguistic situation would be needed to pinpoint the progression of grammaticalization in this dialect over the last 50 years.

#### 5.4. Sinitic languages without complementizers

Detailed examination of discourse data from four further non-standardized languages representing the Xiang, Wu, Gan and Jin branches of Sinitic did not turn up any highly grammaticalized usages for their *verba dicendi*.<sup>35</sup> Given that the same genre of oral narratives was analysed for each of the representative languages, the data were deemed to be fully comparable to those used in the examination of Southern Min and Cantonese Yue. None of the representative varieties of these four different Sinitic groups showed any evidence, however, of either complementizer or semi-complementizer usage for the relevant SAY verb. Although larger

<sup>35</sup> A corpus of informal oral narratives was used in all cases. Due to the limited size of their respective corpora, the findings for Nanchang Gan and Huojia Jin necessarily remain the least robust at this stage, awaiting confirmation from a larger corpus of materials (see Appendix 2 for tabulations of these data).

samples of discourse data are certainly needed to confirm these trends, the present study is indicative of differences between Sinitic languages in the degree of grammaticalization of high frequency verbs such as the SAY class.

This notwithstanding, apart from their principal uses as lexical and quotative verbs, compound subordinating conjunctions formed with SAY and coding conditional ‘if’ and consequence ‘therefore’ were identified in Shanghainese, Mandarin and Jin, not to mention topic-marking functions in Shanghainese Wu, Nanchang Gan and Changsha Xiang, a negative comparative use in Changsha Xiang and a rhetorical function to express surprise in Nanchang Gan and Hakka. Excluding compound conjunctions which code conditional and consequence clauses, the other uses did not show any of the symptomatic signs of grammaticalization.

## 6. Conclusion

This analysis has suggested that in terms of synchronic description, the grammaticalization of SAY verbs into complementizers is taking place concurrently in several Sinitic languages but to different degrees. The main finding of the quantitative discourse analysis is that while Southern Min is close to completing the grammaticalization process, Cantonese Yue is situated at an earlier stage on the pathway, that of a semi-complementizer about to become a full-fledged complementizer used with cognition verbs. In other words, for Southern Min, only obligatory use of *kóng* is missing in the wake of conventionalization of meaning for its complementizer (Stage V), while Cantonese Yue is on the verge of crossing the threshold into the switch context stage III. Beijing Mandarin has reached Stage IV, while Taiwanese Mandarin appears to have advanced nearly as far as Southern Min, according to the available discourse studies. In contrast to this, Sixian Hakka evidences just the semi-complementizer usage with speech act verbs (Stage II).

In contrast to these Sinitic languages, standard Mandarin or *pǔtōnghuà* as well as representative varieties of Xiang, Gan, Wu and Jin do not show signs of any such development, even for the semi-complementizer usage. This is of significant interest in itself, given that the texts consulted all exemplify the same genre of the oral narrative. Furthermore, where different types of oral texts are available, such as for Southern Min, the use of complementizers does not appear to be more characteristic of one genre than another, occurring in both multi-party conversations and in single-party narratives.

Once more large-scale discourse studies are undertaken, it would be expected that complementizers based on verbs of saying, not to mention other grammaticalized uses of SAY verbs, are going to be found in many other Chinese languages and dialects. Serendipitously, one stumbles upon such instances, while searching for other syntactic phenomena, for example, the use of *jā tsǎ̃ koŋ* ‘quarrel-say’ in Ningbo Wu (Forrest 1965: 345).

Finally, the primary importance of using spoken discourse materials as the main source of data can only be reiterated, as a corollary to the entire problematic engaged upon: the use of a complementizer is more likely to be detected in colloquial discourse materials than in elicited data or in reference grammars for standard varieties of languages. This has far-reaching consequences for methodology in current typological studies.

\*\*\*\*\*

*Correspondence address:*

Centre de Recherches Linguistiques sur l'Asie Orientale, Ecole des Hautes Etudes en Sciences Sociales (EHESS), and Centre National de la Recherche Scientifique (CNRS), 54 boulevard Raspail, 75270 Paris cédex 06, France.  
eMail : Hilary.Chappell@ehess.fr

*Acknowledgements:*

I would like to express my indebtedness to the three anonymous reviewers of *Linguistic Typology* for two sets of copious and detailed comments which helped to considerably reshape this article; my gratitude to Tom Güldemann for kindly making available his pre-publication manuscript on quotatives in African languages (2001), illuminating for its in-depth typological overview of the expanded category of quotative verbs; to Yunji Wu for sharing her transcriptions of narratives in the Changsha dialect of Xiang; and to Ka-Wai Yeung for generously offering me the use of her Hong Kong Cantonese database materials, not to mention her article on Cantonese complementation prior to its publication (now Yeung 2006).

I would also like to thank the following colleagues for comments, critique, and their insights on several earlier versions of this paper during its lengthy gestation from an originally shorter paper on just Southern Min: Sacha Aikhenvald, Mengistu Amberber, Ang Ui-jin, Isabelle Bril, Nick Enfield, Fang Mei, Alexandre François, Christine Lamarre, Charles Li, Chinfu Lien, Hsiu-ying Liu, Bill McGregor, Mei Tsu-lin, Alain Peyraube, Stéphane Robert, Laurent Sagart, James Tai, Hongyin Tao, Sandy Thompson, Catherine Travis, Martine Vanhove and Anna Wierzbicka. It goes without saying that I remain responsible for the final interpretations (but not for any oblivion on their part!).

Recording and transcription work for the Southern Min and Cantonese texts and the research trips to China, Hong Kong and Taiwan, originally benefited from funding from two Australian Research Council Large Grants for the projects “A semantic typology of complex syntactic constructions in Sinitic languages” (A59701190, 1997-1999) and “Medieval Chinese and the diachronic syntax of Southern Min: 16<sup>th</sup>–20<sup>th</sup> centuries” (A00106262, 2001—2003), the latter jointly held with Alain Peyraube, EHESS-CNRS, France. Finally, Imogen Yu-chin Chen and Catherine Ng Chan Kam Chi, who worked on these projects in Melbourne, are to be respectively thanked for their assistance with the data compilation and transcription of the Taiwanese Southern Min and Hong Kong Cantonese narratives and conversations.

*Abbreviations:* ASST assertive use of the particle *ge*<sup>3</sup>; CL classifier; COM comitative marker; COMM communication verb; COMP complementizer; COND conditional marker; CONJ conjunction; CRS currently relevant state marker; DAT = dative preposition; DIR directional complement; DM discourse marker; DO direct object; EVD evidential modality marker; IMP imperative marker; FUT = future, <I... J > Japanese insert; LOC locative preposition; LIG linker (genitive, relative clause or attributive marker); <MC ... MC > Mandarin Chinese insert; NEG negative adverb; NOM nominalizer (suffix or particle); OM object marker; PRT discourse particle; PFT perfect or change of state aspect marker; PFV perfective aspect marker; PL plural; POSS possessive pronoun; PROG progressive aspect marker; PRT discourse particle; PURP = purposive; Q interrogative marker; < Q ... Q > quotation; REDUP reduplication; SFP sentence-final particle; SG singular; SUBJ subject; TOP topic marker.

*Transcriptions:* The transcriptions of recorded material in Taiwanese Southern Min and Hong Kong Cantonese follow the system devised for natural conversation and oral narratives by Du Bois and co-researchers at the University of California at Santa Barbara. The intonation unit is treated as the basic unit of conversation; it is indicated simply by means of a new line in the examples from the texts reproduced above.

An intonation unit (IU) can be defined as a single coherent contour over a group of words, identifiable in terms of pause phenomena; resetting of pitch at the beginning of the I.U. and acceleration towards the end of it. IUs can be separated by just a fraction of a second or up to several seconds. Cognitively, they represent the expression of one new idea or ‘single focus of attention’ and correspond closely to the clause in written English (though not necessarily in Chinese languages).

For reasons of authenticity, the actual pronunciation is always given, while the accepted citation form is added to its right and indicated by /X/. Hence, the transcriptions will include non-standard forms, mistakes, disfluencies, repairs, shortened and assimilated syllables. Hesitations and pauses are marked as well as other kinds of vocal noises, including, for example, laughing with ‘@@@’. The full list of symbols is provided in Du Bois et al (1993).

For the purposes of my project on the Typology of Sinitic Languages, I have introduced an extra line with the Chinese characters, the practice adopted in my edited volume on Sinitic synchronic and diachronic linguistics (2001a). I also use make use of a slightly broader transcription practice than Du Bois et al for convenience of reading;

and have taken the liberty of omitting some of the diacritics and symbols in the original transcriptions for ease of presentation here.

*Romanizations:* For **Taiwanese Southern Min**, I use the Church Romanization, as is current practice. This system uses the following tone diacritics: **a** (no marking) = high level 55; **á** = high falling 51; **à** = low falling 21; **â** = mid rising 25; **ã** = low level 33; and two ‘entering tones’ which are found in syllables with final stops: **aC** = low checked 2, **aC** = high checked 5. Complex tone sandhi affects these seven basic tones of Taiwanese, the tonal value being determined by the position of a syllable in a tone group as either juncture or context. This is, however, not indicated in the transcription.

For the Southern Min characters, I have aimed to use, wherever possible, the correct etymological source (*běnzì* 本字) or if not, the *kunyomi* 訓讀 or ‘trained reading’ method where the character is chosen with the right meaning for the gloss, in the spirit of Yang (1991). There is some difficulty in being consistent in this practice, given that the language does not yet have a standardized written form.

The **Cantonese** examples from Yeung (2006) use the Yale system while my own data employs the Sidney Lau system. The **Hakka, Xiang, Shanghainese, Gan** and **Jin** examples are rendered in the International Phonetic Alphabet. The **Mandarin** examples are transcribed by means of the *pīnyīn* romanization system, adopted in 1958 for transcription and language pedagogy purposes by the Chinese government.

Appendix 1: The corpus of spoken discourse materials used in the quantitative analysis

The quantified discourse analysis is based on two sets of recordings: one of Southern Min, as spoken in Taiwan and the other of Cantonese, as spoken in Hong Kong. The transcriptions of these conversations and spoken narratives in Taiwanese Southern Min and Hong Kong Cantonese follow the system devised for natural conversation and oral narratives by Du Bois and colleagues at the University of California at Santa Barbara (see Du Bois et al 1993). The intonation unit is treated as the basic unit of conversation, a unit of discourse with prosodic, syntactic and cognitive ramifications (see Chafe 1993).

Further information is summarized in Table A1.1 and Table A1.2 on the genre and content of the recordings.

(i) Taiwanese Southern Min

The three transcriptions for Southern Min include two narratives and one family conversation.

Table A1.1 : Southern Min corpus

<b>Title</b>	FATE	JAPANESE	JESSE	TOTAL
<b>Length in minutes</b>	14:06	26:23	17:48	58:17
<b>N° Intonation units</b>	796	1216	831	2843
<b>Genre</b>	Family conversation in Taipei	Narrative by Fang Laoshi (Other interlocutors make remarks)	Narrative by Jesse Chen (A second interlocutor makes remarks)	
<b>N° of interlocutors</b>	6	4	2	
<b>Content</b>	<i>Various:</i> 1. <i>Principal subject</i> News on the latest fortune-telling results for members of the family, on the basis of divination carried out by the youngest uncle ; 2. <i>Also discussed:</i> sister and her family in Australia, stockmarket losses, changing jobs	1. <i>Japanese history:</i> Rise of General Toyotomi; 2. <i>Also discussed:</i> life in Taiwan under the Japanese occupation	<i>Reminiscences</i> 1. <i>Childhood stories:</i> summer holidays, running a family business, waitering experiences 2. <i>As an adult:</i> Lost overnight on a mountain	

## (ii) Hong Kong Cantonese

The Cantonese corpus includes two narratives which recount traditional romances that are the basis of popular Chinese operas, favourites chosen by the speaker, Anna Fung Shuk-han.

Table A1.2: Cantonese corpus

NARRATIVE	<i>Tale of the Reborn Lady at the Red Flowering Plum</i> 再世紅梅記	<i>Balcony Rendezvous</i> 樓台會	TOTAL
<b>Length in minutes</b>	7:24	13:42	21:06
<b>N° Intonation units</b>	416	754	1,170
<b>Genre</b>	Narrative	Narrative	
<b>N° of interlocutors</b>	One-party	One-party	
<b>Content</b>	Traditional Chinese opera story of a reincarnated damsel , a scholar-hero and a villainous chief minister	Traditional Chinese opera story of the star-crossed lovers, Liang Shanbo and ZhuYingtai	

Appendix 2: Other Sinitic languages – *Say* verbs in oral narratives

## (i) Changsha dialect of Xiang, Hunan province

The corpus comprises an oral narrative of 90 minutes, entitled ‘Travelling’ *li<sup>41</sup> iəu<sup>13</sup>* 旅游 (=LY), and relates interesting and unusual incidents during the speaker’s travels in China. This was recorded and transcribed by Wu Yunji in Changsha for her project on Xiang grammar. A second collection of 3 short oral narratives, all based on folktales, is taken from Li Yongming (1991: 613-650) (=CF).

Table A2.1: The functions of *kan<sup>41</sup>* 讲 ‘say’ in Changsha Xiang narratives

Function	Gloss	LY 90 mins	CF 38 pages	<i>Total</i>
1. Lexical	‘speak’, ‘talk about’, ‘tell’	58	8	66
2. Quotative	‘say [QUOTE]’	87	9	96
3. Topic	‘talking about X’	3	0	3
4. Negative comparative	不讲 ‘not like’ <i>pu<sup>24</sup> kan<sup>41</sup></i>	2	0	2
<i>Total</i>		<i>150</i>	<i>17</i>	<i>167</i>

## (ii) Shanghainese dialect of Wu, Shanghai city, Zhejiang province

The corpus consists of two sets of oral narratives on cultural life in Shanghai and Australia, recorded in Shanghai by Xu Ziliang and in Melbourne by Shi Shiqing respectively for my project on the typology of Sinitic languages, as well as transcriptions of Shanghainese oral narratives based on folktales published in Xu and Tang (1988).

Table A2.2: The functions of the verb *kā<sup>34</sup>* 講 ‘say’ in Shanghainese Wu

Title Function	Shanghai conversations 15:37 mins	‘The three excesses of Australia’ 8:30 mins	Oral narratives in Xu and Tang (1988) (57 pages)	Total
1. Lexical	3	26	48	77
2. Quotative	1	0	25	26
3. Topic	1	3	4	8
4. Conditional	0	0	7	7
5. Other compound conjunctions	1	2	1	4
Total	6	31	85	122

## (iii) Sixian Hakka of Taoyuan, Taiwan

These 15 oral narratives by the same speaker are taken from Yang Shifeng 1957: (133-200). Each one recounts a folktale. Of all the additional discourse material consulted to enlarge the sample of Sinitic languages, this is the only set where I found the semi-complementizer usage of a SAY verb in  $V_1V_2$  structures.

Table A2.3: Functions of the verb *koŋ*<sup>51</sup> 講 ‘say’ in 15 Sixian Hakka stories

<i>Function</i>	<i>Total</i>
Lexical	47
Quotative	97
$V_1V_2$ semi-complementizer	14
<i>Total</i>	158

These data were collected in the 1950s by Yang Shifeng and represent the largest corpus of Hakka materials available from the second half of the 20<sup>th</sup> century. For this reason, we make use of them in the survey, as to our knowledge no other corpora of this size are available at present. Sixian Hakka is, importantly, the unofficial prestige variety of Hakka (Chappell and Lamarre 2005, chapter 1). It can be surmised that if the semi-complementizer use of *koŋ*<sup>51</sup> 講 ‘say’ was possible in the 1950s, then a higher degree of grammaticalization may well have been reached by now, verification for which would require further fieldwork.

Note that, in this volume by Yang, there are a further 11 stories recounted by a speaker of Hailu Hakka, a different dialect, not surveyed here.

## (iv) Nanchang dialect of Gan, Jiangsu province

These five narratives were recorded and transcribed by Laurent Sagart in Nanchang in 1999 for a reference grammar of Nanchang Gan, as part of my project on the typology of Sinitic languages (see Sagart 1999). The narratives entitled ‘Nanchang Weather’, ‘New Year’ and ‘Noodles’ are factual accounts respectively of the climate, celebrations at Chinese New Year and the recipe for how to make the special local dish of Nanchang noodles. The remaining two recount folktales. Given the small size of the corpus available, the findings can only be taken as indicative.

Table A2.4: The functions of the verb *wa*<sup>6</sup> 話 ‘say’ in Nanchang dialect of Gan

Narrative / Function	Nanchang Weather	Scholar and the Maiden	New Year	Turnips	Noodles	Total 17:43 mins
1. Lexical	1	2	1	8	2	15
2. Quotative	0	4	1	4	6	14
3. Rhetorical ‘Would you believe ...!’	0	0	2	1	0	3
Total	1	6	5	13	8	32

(v) Huojia dialect of Jin, Northern Henan province

For the Jin dialect group, I make use of two oral narratives transcribed in He Wei (1989: 123-130), a reference grammar of the Huojia dialect of Jin as spoken in northern Henan province.

Given the small size of the corpus available, the findings can only be taken as indicative. The first story, ‘Wolf’, is a folktale akin to ‘Red Riding Hood’ while the second relates changes in village life post-1949.

Table A2.5: Functions of the verb *ʂueʔ³* 說 ‘say’ in the Huojia dialect of Jin

Story Function	Wolf 狼 (2 pages)	Changes in our village 俺庄儿的变化 (6 pages)	Total (8 pages)
1. Lexical	0	10	10
2. Quotative	10	0	10
3. Conditional uses	0	3	3
Total	10	13	23

## References

- Amberber, Mengistu. 1995. Functional verbs in predicate formation: Event-type hierarchy and grammaticization. In Jocelyn Ahlers et al (eds.), *Proceedings of the Twenty-first Annual Meeting of the Berkeley Linguistics Society*, 1-11.
- Bisang, Walter. 1992. *Das Verb im Chinesischen, Hmong, Vietnamesischen, Thai und Khmer*. Tübingen: Gunter Narr Verlag.
- Bisang, Walter. 2001). Finite vs. non finite languages. In Haspelmath et al (eds.), 2001, 1400-1413.
- Bolinger, Dwight L. 1972. *That's that*. The Hague : Mouton.
- Bril, Isabelle. 2002. *Le Nélémwa (Nouvelle Calédonie)*. Collection *Langues et cultures du Pacifique* N° 16. Paris: Edition Peeters.
- Bresnan, Joan. 1970. On complementizers: Towards a syntactic theory of complement types. *Foundations of Language* 6: 297-321.
- Bresnan, Joan. 1979. *Theory of complementation in English syntax*. New York/London: Garland.
- Chafe, Wallace. 1993. Prosodic and functional units of language. In Edwards & Lampert (eds.), 1993, 33-43.
- Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Chappell, Hilary. 1997. The grammaticalization of the verb *kóng* 'say' in Taiwanese Southern Min: A case of reanalysis. Paper presented in January 1997 at the Lake Arrowhead Conference on *Morphological change in Chinese*, University of California at Los Angeles and the University of California at Santa Barbara.
- Chappell, Hilary. (ed.). 2001a. *Sinitic grammar: synchronic and diachronic perspectives*. Oxford: Oxford University Press.
- Chappell, Hilary. 2001b. Language contact and areal diffusion in Sinitic languages: problems of typology and genetic affiliation, In: Alexandra Aikhenvald & R.M.W. Dixon (eds.), *Areal diffusion and genetic inheritance: problems in comparative linguistics*, 328-357. Oxford: University Press.
- Chappell, Hilary. 2002. The universal syntax of semantic primes in Mandarin Chinese. In Cliff Goddard & Anna Wierzbicka (eds.) *Meaning and universal grammar – Theory and empirical findings*. Amsterdam: John Benjamins, 1: 243–322.
- Chappell, Hilary. In prep. *Typology of Sinitic languages: morphosyntax and grammar*. Oxford: Oxford University Press.
- Chappell, Hilary & Christine Lamarre. 2005. *A grammar and lexicon of Hakka: Historical materials from the Basel Mission Library*. Collection *des Cahiers de Linguistique – Asie Orientale* 8, Paris: École des Hautes Études en Sciences Sociales.
- Chappell, Hilary, Li Ming & Alain Peyraube. 2007. Chinese linguistics and typology: the state of the art. *Linguistic Typology* 11.1: 187-211.
- Chappell, Hilary & Alain Peyraube. 2006. Polygrammaticalization of SAY verbs in Sinitic languages: a diachronic study. Plenary talk, July 24th, 2006 at the 20<sup>ème</sup> Journées de Linguistique Asie Orientale, CRLAO, CNRS, Paris.
- Chen, Ping. 1999. *Modern Chinese: History and sociolinguistics*. Cambridge: Cambridge University Press.

- Cheng, Robert L 鄭良偉. 1991 Taiyu yu Taiwan Huayu-li de ziju jieou biaozi 台語與台灣華語裏的句結構標誌 [The complementation markers *kóng* ‘say’ and *k’uà* ‘see’ in Taiwanese and Taiwanese Mandarin]. In: *Di San-jie Shijie Huayu Wenjiaoxue Yantaohui* 第三世界華語文教學研討會 [The Third World Symposium on Chinese Culture and Education]. Taipei. (Reprinted in Cheng (ed.) 1997, vol. II: 105 – 132.)
- Chirkova, Katia & Christine Lamarre. 2005. The paradox of the construction [V *zai* NP<sub>LOC</sub>] and its meanings in the Beijing dialect of Mandarin. *Cahiers de Linguistique – Asie Orientale* 34.2: 169-220.
- Chui, Kawai. 1994. Grammaticization of the saying verb *wa* in Cantonese. *Santa Barbara Papers in Linguistics* 5.1-12.
- Crowley, Terry. 1989. Say, c’est, and subordinate constructions in Melanesian Pidgin. *Journal of Pidgin and Creole Languages*. 4.185-210.
- Dixon, R.M.W. 1991. Complement clauses. Chapter 7 in *A new approach to English grammar, on semantic principles*. Oxford: Clarendon Press.
- Du Bois, Jack, Stephan Schuetze-Coburn, Danae Paolino & Susanna Cumming. 1993. Outline of discourse transcription. In Edwards & Lampert (eds.), 1993, 45-89.
- Ebert, Karen. 1986. Reported speech in some languages of Nepal. In Florian Coulmas (ed.), *Direct and indirect speech* (Trends in Linguistics 31), 145-159. Berlin: Mouton de Gruyter.
- Ebert, Karen. 1991. Vom verbum dicendi zur Konjunktion – Ein Kapitel universaler Grammatikalisierung. In Walter Bisang & Peter Rinderknecht (eds.), *Von Europa bis Ozeanien- von der Antonymie zum Relativsatz. Gedenkschrift für Meinrad Scheller*. (Arbeiten des Seminars für Allgemeine Sprachwissenschaft 11.), 77-95. Zürich: Universität Zürich.
- Edwards, J. A. & M. D. Lampert (eds.). 1993. *Talking data: transcription and coding in research*. New Jersey and London: Lawrence Erlbaum Associates.
- Fang, Mei 方梅. 2006. Beijinghua-li ‘shuo’ de yufahua – cong yanshuo dongci dao congju biaoji 北京话里《说》的语法化 – 从言说动词到从句标记 [Grammaticalization of *shuō* (say) in Beijing Mandarin : from lexical verb to subordinator] *Zhongguo Fangyan Xuebao* 中国方言学报 1.107-121.
- Forrest, R.A.D. 1965. *The Chinese language*. London: Faber & Faber (2nd edition).
- Frajzyngier, Zygmunt. 1991. The *de dicto* domain in language. In Traugott & Heine (eds.), 1991, 1.219-251.
- Frajzyngier, Zygmunt. 1995. *A functional theory of complementizers*. In Joan Bybee & Suzanne Fleischmann (eds.), *Modality in grammar and discourse*, 473-502, Amsterdam/Philadelphia: John Benjamins.
- Frajzyngier, Zygmunt. 1996. *Grammaticalization of the complex sentence: A case study in Chadic*. Amsterdam: John Benjamins.
- François, Alexandre. 2002. Araki: a disappearing language of Vanuatu (Pacific Linguistics N° 522). Canberra: Australian National University.
- Givon, Talmy. 1980. *On understanding grammar*. New York: Academic Press.
- Güldemann, Tom . 2001. *Quotative constructions in African languages: a synchronic and diachronic survey*. Leipzig: Universität Leipzig, Institut für Afrikanistik habilitation thesis.
- Güldemann, Tom . 2002. When ‘say’ is not say: the functional versatility of the Bantu

- quotative marker *ti* with special reference to Shona. In Tom Güldemann & Manfred von Roncador (eds.), *Reported discourse: a meeting ground for different linguistic domains* (Typological Studies in Language 52), 253-287. Amsterdam: John Benjamins.
- Harris, Alice & Lyle Campbell. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
- Haspelmath, Martin, Ekkehart König, W. Österreicher & W. Raible (eds.). 2001. *Language typology and language universals: an international handbook*. Berlin: Mouton de Gruyter.
- Haspelmath, Martin, Matthew Dryer, David Gil & Bernard Comrie (eds.). 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge University Press.
- He, Wei. 1989. *Huojia fangyan yanjiu 獲嘉方言研究* [Research on the Huojia dialect]. Beijing: Shangwu.
- Heine, Bernd. 2002. On the role of context in grammaticalization. In Ilse Wischer & Gabriele Diewald (eds.), *New Reflections on Grammaticalization*, 83-101. Amsterdam: John Benjamins.
- Heine, Bernd, Ulrike Claudi & Friederike Hünemeyer. 1991. *Grammaticalization: a conceptual framework*. Chicago: University of Chicago Press.
- Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Heine, Bernd & Tania Kuteva. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Heine, Bernd & Tania Kuteva. 2007. *The genesis of grammar*. Oxford: Oxford University Press.
- Hock, Hans Heinrich. 1982. The Sanskrit quotative: A historical and comparative study. *Studies in the Linguistic Sciences* 12.37-96.
- Hopper, Paul. 1991. On some principles of grammaticalization. In Traugott & Heine (eds.), 1991, 1. 17-36.
- Hopper, Paul & Elisabeth Closs Traugott. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.
- Horie, Kaoru. 2001. Complement clauses. In : Martin Haspelmath et al (eds.), 2001, 979-993.
- Hu, J. Pan, H. & Xu, L. 2001. Is there a finite vs. nonfinite distinction in Chinese? *Linguistics* 39.1117-1148.
- Huang, James C.-T. 1998. *Logical relations in Chinese and the theory of grammar*. New York: Garland.
- Huang, Shuanfan 黃宣繁. 1993. *Yuyan, shehui yu zuquan yishi 語言, 社會與族權意識* (*Language, Society and Ethnic Identity*). Taipei: Crane Publishing Co. Ltd.
- Huang, Shuan-fan. 2003. Doubts about complementation: a functionalist analysis. *Language and linguistics* 4(2).429-455.
- Hwang, Jya-Lin. 1998. A comparative study on the grammaticalisation of saying verbs in Chinese. In Chaofen Sun (ed.), *Proceedings of the 10<sup>th</sup> North American Conference on Chinese Linguistics*, 574-584. Stanford: Stanford University.
- Klamer, Marian. 2000. How report verbs become quote markers and complementizers. *Lingua* 110.69-98.
- Langacker, Ronald. 1977. Syntactic reanalysis. In Charles N. Li (ed.), *Mechanisms of syntactic change*, 57-139. Austin and London : University of Texas Press.

- Leok, Har Chan. 1982. *Information structure and sentence mood in Hokkien syntax*. Manoa: University of Hawaii doctoral dissertation.
- Li, Audrey Yen-hui. 1990. *Order and constituency in Mandarin Chinese*. Dordrecht: Kluwer.
- Li, Charles N. 1988. Grammaticalization in Hmong: Verbs of saying. Paper presented at the 21st International Sino-Tibetan Languages and Linguistics Conference. Lund.
- Li, Charles N. & Sandra A. Thompson. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Li, Yongming 李永明. 1991. *Changsha fangyan 长沙方言*(The Changsha dialect). Changsha: Hunan Chubanshe.
- Liu Hsiu-ying 劉秀英 1996. Minnanhua shuo-hua dongci kóng zhi gongneng yanbian ji yuyi tantao 閩南話說話動講之功能演變及語義探討 [Exploration of the functional development and semantics of the speech act verb kóng ‘say’ in Southern Min]. Ms. Hsinchu: Institute of Linguistics, National Tsing Hua University, Taiwan.
- Liu, Yuehua 刘月华. 1986. Duihu-zhong ‘shuo’, ‘xiang’, ‘kan’ de yizhong teshu yongfa 对话中‘说’‘想’‘看’的一种特殊用法 [The special use of say, think and see in conversation. ] *Zhongguo Yuwen* 3.168-172.
- Lord, Carol. 1976. Evidence for syntactic reanalysis: From verb to complementizer in Kwa. In S. Steever, Carol Walker & S. Mufwene (eds.), *Papers from the parasession on diachronic syntax*, 179-191. Chicago: Chicago Linguistic Society.
- Lord, Carol. 1993. *Historical change in serial verb constructions* (Typological Studies in Language 26). Amsterdam: John Benjamins.
- Matthews, Stephen & Virginia Yip. 1994. *Cantonese: A comprehensive grammar*. London: Routledge.
- Matisoff, James. 1991. Areal and universal dimensions of grammatization in Lahu. In Traugott & Heine (eds.), 1991, 2.383-454.
- Meng, Cong 孟琮. 1982. Kouyu ‘shuo’ zi xiaoji 口语‘说’字小集 [A mini-collection of the uses of ‘say’ in the spoken language] *Zhongguo Yuwen* 5: 337-346.
- Munro, Pamela. 1978. Chemehuevi ‘say’ and the Uto-Aztecan quotative pattern, *Selected Papers from the 14th Great Basin Anthropological Conference*, D. R. Tuohy, (ed.), Ballena Press Publications in Archaeology, Ethnology, and History 11, 149-71. Socorro, NM: Ballena Press.
- Munro, Pamela. 1982. On the transitivity of ‘say’ verbs. In Paul J. Hopper & Sandra A. Thompson (eds.), *Studies in Transitivity* (Syntax and semantics 15) , 301-318. New York/ London: Academic Press.
- Noonan, Michael. 1985. Complementation. In Timothy Shopen (ed.), *Language Typology and syntactic description*, 2. 42-140. *Complex constructions*. Cambridge: Cambridge University Press.
- Noonan, Michael. 2006. Direct speech as a rhetorical style in Chantyal. *Himalayan Linguistics Journal* 6.1-32.
- Paris, Marie-Claude. 1996. La subordination en chinois standard: quelques contraintes d’agencement. In Claude Muller (ed.), *Dépendance et intégration syntaxique, subordination, coordination, connexion*. (Linguistische Arbeiten, vol 351), 198-206. Tübingen: Max Niemeyer.
- Peyraube, Alain. 1996. Recent issues in Chinese historical syntax. In C–T James Huang, and Y.–H. Audrey Li (eds.), *New Horizon in Chinese Linguistics*, 161-214. Dordrecht: Kluwer Academic Publishers.

- Plag, Ingo. 1992. From speech act verb to conjunction : the grammaticalization of *taki* in Sranan. *Journal of Pidgin and Creole Languages* 7(1). 55-73.
- Ramsey, S. Robert. 1987. *The languages of China*. Princeton: Princeton U.P.
- Ransom, Evelyn. 1988. The grammaticalization of complementizers. *Proceedings of the Berkeley Linguistics Society* 14.364-374.
- Sagart, Laurent. 1999. *Notes on the Nanchang dialect (Sketch grammar and transcriptions)*. Manuscript, Melbourne: Department of Linguistics, La Trobe University.
- Saxena, Anju. 1988. The case of the verb 'say' in Tibeto-Burman. *Proceedings of the Berkeley Linguistics Society* 14.375-388.
- Saxena, Anju. 1995. Unidirectional grammaticalization: diachronic and synchronic evidence. *Sprachtypologie und Universalienforschung* 48.350-372.
- Simpson, Andrew & Zoë 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.
- Sweetser, Eve. 1990. *From etymology to pragmatics*. Cambridge: Cambridge University Press.
- Traugott, Elizabeth Closs & Bernd Heine (eds.). 1991 *Approaches to grammaticalization*. Amsterdam: John Benjamins.
- Truong, Văn Chinh. 1970. *Structure de la langue vietnamienne*. Publications of the Centre Universitaire des Langues Orientales vivantes, 6th series, vol. X. Paris : Librairie orientaliste Paul Geuthner.
- Xu Baohua & Tang Zhenzhu 许宝华、汤珍珠. 1988. *Shanghai Shiqu Fangyanzhi 上海市区方言志 [A record of the dialect of Shanghai City]* Hong Kong: Joint Publishers.
- Xu, Huiling & Stephen Matthews. 2007. Cong dongci dao ziju jiegou biaoji: Chaozhou fangyan he Taiwan Minnan hua dongci 'shuo' he 'kan' de xuhua guocheng 從動詞到子句標記: 潮州方言和台灣閩南話動詞 '說' 和 '看' 的虛化過程 [From verb to complementizer: the grammaticalization process for the verbs 'say' and 'see' in the Chaozhou dialect and Taiwanese Southern Min]. *中國語文研究 Zhongguo Yuwen Yanjiu [Studies in Chinese Linguistics]* 23.61-72.
- Wierzbicka, Anna. 1988. The semantics of English complementation in a cross-linguistic perspective. Chapter 1 in *The semantics of grammar*. Amsterdam: John Benjamins.
- Wu, Xiu-zhi Zoë. 2004. *Grammaticalization and language change in Chinese: A formal view*. London: Routledge Curzon.
- Yang Hsiu-fang. 1991. *Taiwan Minnanyu Yufagao 臺灣閩南語語法稿 [A grammar of Taiwanese Southern Min]*. Taipei: Ta An Press.
- Yang, Shih-feng 楊時逢. 1957. *The Hakka dialect of Taoyuan, Taiwan*. 臺灣桃園客家話方言. Taipei : The Institute of History and Philology Monographs, Series A, No. 22.
- Yeung, Kawaii K.Y. 2003. The grammaticalization of complementizers from verbs of 'saying': the case of Cantonese *waa6*. In Y-H. Lin (ed.), *Proceedings of the 15<sup>th</sup> North American Conference on Chinese Linguistics*, 487-503.
- Yeung, Ka-wai. 2006. On the status of the complementizer *waa6* in Cantonese. *Taiwan Journal of Linguistics* 4.1: 1-48.
- Yu Hua 于華. 1994. *Huózhe 活着*. [To live]. Hong Kong: Boyi Publishers.