ON THE STATUS OF THE COMPLEMENTIZER WAA6 IN CANTONESE

Ka-Wai Yeung

ABSTRACT

Complementizers are generally known as function words that introduce a clausal complement, like that in English, for instance (Radford 1997). In many languages, complementizers are re-analyzed from verba dicendi, or verbs of ‘saying’ (Lord 1976; Frjzyngier 1991; Hopper and Traugott 1993; Lord 1993). This paper argues for the existence of a complementizer re-analyzed from a verb of ‘saying’ in Cantonese by providing a synchronic analysis of waa6. Waa6 has often been assumed to be a lexical verb in serial verb construction because of its following a ‘saying’ predicate or a cognitive predicate. However, this paper argues that waa6 is not always a verb, postulating that waa6 may have different meanings and subcategorizations in different situations, including waa6 meaning ‘say’ [__ (PP) CP] or [__ PP NP], the transitive verb waa62 meaning ‘blame’.

1 As is common practice for Mandarin and Taiwanese romanization, the paper use the Scheme for the Chinese Phonetic Alphbat (Hanyu Pinyin Fang’an) and Church Romanization, which was devised by Presbyterian missionaries in the 19th century in Taiwan. The tones are not indicated in the transcription, except when they are essential to the analysis. For Cantonese romanization, this article follows the Cantonese Romanization Scheme (Jyutping) designed by the Linguistic Society of Hong Kong (LSHK) in 1993. See Tang et al (2002). The Chaozhou data are directly cited from Xu and Matthews (2005), in which the romanization is based on Xu (2004).

* I am most grateful to Elaine Francis and Stephen Matthews for their constructive comments on the early drafts of this article. Thanks are also due to Sze-Wing Tang for providing some of the Cantonese examples and to Hilary Chappell for providing the Taiwanese Southern Min data in her article prior to publication. I am indebted to two anonymous reviewers for their comments and suggestions.
condemn’ [__ NP CP], and the complementizer waab selecting a clause [__ IP^2].
This proposal is supported by different tests, such as aspect marking and argument selection, confirming that the complementizer waab formally exhibits different properties from that of the verbs waab^1 and waab^2.

1. INTRODUCTION
1.1 Background to the Grammaticalization of Complementizers from verba dicendi

Complementizers, in general, are the words that introduce a complement clause (IP), indicating that such a word is a complement of some lexical head. Complementizers are typically selected by verbs, usually verba dicendi (i.e., verbs of ‘saying’), like say, tell, or cognitive verbs describing mentality, like think, doubt. For example in English, that in I thought that they would come is categorized as a complementizer.3

Heine (2002) describes grammaticalization in terms of a four-stage scenario. At Stage I, the ‘initial stage,’ the grammaticalized item is expressed with its ‘normal’ or source meaning. At Stage II, there should be a ‘bridging context’ that gives rise to an inference in favour of a new meaning that is a more plausible interpretation than the source meaning. Stage III concerns a ‘switch context,’ in which an interpretation in terms of the old source meaning is no longer allowed. Stage IV is the ‘conventionalisation’ of the grammaticalized items, where the target meaning no longer needs to be supported by the context that gave rise to it. The four-stage scenario does not occur in discrete stages, but in a continuum leading from Stage I to Stage IV and it has been found to be a helpful model in typological studies of the grammaticalization of

---

2 The paper assumes propositional sentences as maximal projections IP projected by the INFL. See Chomsky (1981).
3 The strict definition of complementizers is problematic. See Yeung (2003). While some believe complementizers are function heads that subcategorize a clausal complement IP (Radford 1997; Haegeman and Guéron 1999); and some categorize complementizers by their semantic functions (Wierzbicka 1988; Langacker 1991; Frajzyngier 1995; Givón 2001); some even totally reject the proposition that complementizers form a distinct word class (Hudson 1995; Hudson 2000). Still, these approaches all share the general consensus that complementizers subcategorize for a clausal complement, forming a constituent that acts as the complement for the preceding lexical item (Yeung 2003).
complementizers. Studies of grammaticalization have shown that that one of the paths of re-analysis by which complementizers develop is from *verba dicendi*\(^4\), as both complementizers and verbs potentially subcategorize for a clausal complement (IP). This potential constitutes the ‘right typological precondition’ for the grammaticalization of complementizers from verbs, in particular the ‘say’ verbs. Wu (2004:90-91) observes that ‘frequently this [grammaticalization] occurs when a language has serial verb constructions which allow for a sequence of two verbs of communication (one more specific, the second less specific) to become re-analyzed as a sequence of verb + complementizer [...]’. Xu and Matthews (2005) also suggest that the productive verb serialization in Sinitic languages provides a ‘relevant typological precondition’ for complementizer grammaticalization. These preconditions construct the possible context by which the target meaning can be established.

The first exhaustive study of the re-analysis of complementizers from ‘say’ verbs is by Lord (1976) who surveyed a number of African and Asian languages, mostly Niger-Congo languages, and found that 29 of them appeared to have a complementizer re-analyzed from ‘say’ verbs. Her work provides extensive cross-linguistic evidence that complementizers can be re-analyzed from ‘say’ verbs. (1) shows an example from the Kwa language Ewe, where the complementizer *bé* is re-analyzed from the verb *be*, meaning ‘say’. In (1a), the ‘say’ verb *be* takes a propositional complement. Yet, the ‘say’ verb *gblo* only takes an NP complement, as in (1b). If *gblo* takes a propositional complement, the complement clause must be introduced by a complementizer *bé*, as in (1c) and (1d).

(1)

a. me-*be* mewèè
   *I-say I-do-it*
   ‘I said, “I did it”.’ or
   ‘I said that I did it.’

b. me-*gblo*-e
   *I-say-it*
   ‘I said it.’

---

\(^4\) Regarding the sources of complementizers, Hopper and Traugott (1993) suggest several paths of grammaticalization, including derivation from accusative case morphemes, allatives, datives, deictic demonstratives, lexical nouns, and verbs of ‘saying’. 

---

Cantonese Complementizer *waab*
Apart from West African languages, the same phenomenon is found in other languages, such as the Creole languages Saramaccan, as in (2), and Sranan, as in (3) (Plag 1993; Veenstra 1996; Lefebvre 2001).

**Saramaccan:**

(2) a. Hén a táa: mi nángó
   and-then 3sg say 1sg Asp_go
   ‘And then he said: I am leaving.’

   b. A táki: táa: a náki hén
      3sg say say 3sg hit 3sg
      ‘He said that he had hit her.’
      (Veenstra 1996:155)

**Sranan:**

(3) a. Da’ Anansi taki, eng go proberi efi eng kan saki fa a nem
    ‘Then Anansi said, he would try if he could learn his name.’
    (Plag 1993:40)

   b. Mi sabi taki Marlon no fufuru na fowru
      I know SAY Marlon Neg steal the chicken
      ‘I know that Marlon didn’t steal the chicken.’
      (Plag 1993:36)

For Chinese, some studies (Chui 1994; Hwang 1998; Hwang 2000) have provided evidence to suggest that ‘say’ verbs are grammaticalized into complementizers introducing propositional complement clauses. These examples include shuo in Mandarin, kong in Taiwanese, and waat6 in Cantonese, as shown in (4).

(4) a. Mandarin:

   Liji hui-da shuo (*le) meiyou…
   Liji answer SHUO (*Asp) NO
   ‘Liji’s answers (that) “NO.”’

---

5 It is likely that these Creoles have the same phenomenon as a result of substrate influence from Niger-Congo languages. See Plag (1995).
Cantonese Complementizer \textit{waa6}

b. Taiwanese:
\begin{quote}
\textit{Goa siongsin kong} (*goe) \textit{I} \textit{Teng bo phian goa} \\
\textit{I believe KONG} (*Asp) \textit{he must Neg lie I} \\
\end{quote}

‘I believe that he didn’t lie to me.’

c. Cantonese:
\begin{quote}
\textit{John gong2 waa6 keoi5 hai6 mei5gwok3 lai4 gaa3} \\
\textit{John say WAA6 he be America come SFP} \\
\end{quote}

‘John said that he was from the U.S.’

(Hwang 1998)

These studies are primarily concerned with a diachronic analysis of the grammaticalization of complementizers from ‘say’ verbs, setting the stage for further investigation through a synchronic analysis of these complementizers.

1.2 The Aims and the Proposal of the Paper

Building on proposals from earlier diachronic studies of Chinese, this paper argues for the existence of a complementizer re-analyzed from a verb of ‘saying’ in Cantonese by providing a synchronic analysis of \textit{waa6}. It is tempting to analyze \textit{waa6} as a lexical verb in serial verb construction (SVC) because of its following a ‘saying’ predicate or a cognitive predicate. An SVC analysis is often assumed without any synchronic justification, such as in Matthews and Yip (1994:308), ‘\textit{gong}…\textit{wah} “talk…say” is a serial verb construction in which the second verb \textit{waa6} appears redundant.’ Also, Hwang (2000) analyses \textit{shuo} ‘say’ construction as ‘factive-complementizer \textit{shuo} “say” serial verb construction,’ even though she believes that \textit{shuo} performs the grammatical function as a complementizer following saying, informative, and cognitive verbs. This paper takes the alternative view that although the \textit{waa6} construction is historically derived from an SVC, and \textit{waa6} continues to be used as a serial verb in some contexts, there is a usage of \textit{waa6} that is fully grammaticalized as a complementizer and synchronically distinct from a serial verb.

This paper postulates that \textit{waa6} may have different meanings and subcategorizations in different constructions, including 1) \textit{waa6}, meaning ‘say’ which subcategorizes either an optional PP and an obligatory CP [ ___ (PP) CP], or selects a PP and an NP [ ___ PP NP]; 2) the transitive verb \textit{waa6}, meaning ‘blame/ condemn’ which subcategorizes an NP and an optional CP [ ___ NP (CP)], and 3) the
complementizer *waa6* as a clause introducer selecting a propositional sentence (___ IP). In addition, this paper argues that *waa6* lacks the syntactic characteristics typical of verbs, providing further evidence that *waa6* is not a serial verb in this usage. In the following sections, a number of syntactic and semantic arguments will be given to support these proposals.

**The Three *waa6*’s:**

- *waa61 ‘say’*:  
  - Intransitive Verb: ___ (PP) CP
  - Transitive Verb: ___ PP NP

- *waa62 ‘blame/condemn’*:  
  - Transitive Verb: ___ NP (CP)

- *waa63*:  
  - Complementizer: ___ IP

Having ascertained the complementizer status of *waa63*, the paper then looks into the possibility of further grammaticalization of *waa6* into a sentence-final particle (SFP) with reference to the case of Taiwanese *kong* (Simpson and Wu 2002; Wu 2004). In addition, it also contrasts the grammaticalization of *waa6* in Cantonese with other reanalyzed complementizers in the Sinitic languages, such as *tā* in Chaozhou and Taiwanese *kong*, showing the process of grammaticalization of *waa6* is still in its ‘younger’ stage. The complementizer status of *waa6* also sheds light on word order typology in Chinese. A postulation of how the fact that *waa6* is a head-initial complementizer relates to the unusual word order typology in Chinese will be given in the concluding section.

**2. A PRELIMINARY ANALYSIS BASED ON CANTONESE CORPUS DATA**

Before probing into the syntactic and semantic evidence, this section provides a brief sketch of the Cantonese data gathered from Cantonese corpora. These data provide some preliminary information on the different usages of *waa6*. Most of the data used in the section are taken from two corpora, the Cantonese Radio Corpus (CRC) by Francis, Yiu, Matthews, and Chu (in preparation), and the Hong Kong Cantonese Adult Language Corpus (HKCAC) by Law et al. (2002).
2.1 The Verbal Usage of *waa6*

The basic usage of *waa6* is as a ‘say’ verb, as in (5a). As a ‘say’ verb, *waa6* subcategorizes a CP (___ CP) and it exhibits several verbal properties, including aspect marking (5b) and (5c), selecting an optional PP goal argument (5d) and (5e), and following an auxiliary (5f).

(5) a. keoi5 waa6 [CP ji5hau6 dou1 m4 heoi3 lak3] 3sg WAA1 [CP forever all Neg go Part] ‘(S)he said that (s)he would never go again.’

b. keoi5 waa6_zo2 [CP ji5hau6 dou1 m4 heoi3 lak3] (HKCAC) 3sg WAA1_Asp [CP forever all Neg go Part] ‘(S)he said that (s)he would never go again.’

c. ngo5 mou5 waa6_gwo3 [CP lei5 co3] (CRC) 1sg Neg WAA1_Asp [CP 2sg wrong] ‘I didn’t say you’re wrong.’

d. keoi5 waa6 [pp bei2 ngo5 teng1], (HKCAC) 3sg WAA1 [pp to 1sg listen], [CP/NP loeng5ci3 lo1] [CP/NP twice Part] ‘(S)he told me that it happened twice.’ or ‘(S)he told me twice.’

e. zau6 waa6 [pp bei2 ngo5dei6 zil] (CRC) then WAA1 [pp to 1pl know] [CP/au5 me1 saang1ji3 ho2ji3 zou6] [CP have what business can do] ‘Then tell us what kind of business we can do.’

f. ngo5 gok3dak3 ngo5 jing1goi1 (CRC) 1sg think 1sg Mod waa6 bei2 keoi5 teng1 [NP li1go3 cing4fong3] WAA1 to 3sg listen [NP this situation] ‘I think I should tell him/her the situation.’

Sometimes, *waa6* is used in a metaphorical sense. This metaphorical extension bleaches the original verb of its meaning of ‘say’ and the verb meaning slowly alters from to that of a ‘cognitive’ sense, such as ‘think.’ Hwang (1998) suggests that the metaphor involved here is probably *SPEECH IS THOUGHT* (Chui 1994; Lord 1993). In (6), *waa6* may also be used as a cognitive verb, meaning ‘think.’
Apart from the ‘say’ verb (waab61), there is another verb waab62 meaning ‘blame’, ‘condemn’ or ‘criticize.’ The subcategorization of waab62 is different from waab61 (___ CP). It selects an NP, the patient argument, and a CP, (___ NP CP), as (7a) and (7b). The passive construction of waab62 (7c) is also common. It should be noted that for the waab62 construction, it is possible to have another waab before the complement clause, as in (7d). It is likely that the second waab in (7d) is the complementizer waab63. A detailed analysis will be provided in later sections.

(7) a. …waab6 [NP zi6gei2] [CP sat1wan4 aa3] (CRC)
    …WAA2 [NP oneself] [CP absent-minded Part]
    ‘…blamed yourself for being scatter-brained.’

b. tau4jat1nin4 zuau6hai6 waab6 [NPlei5] (CRC)
    first year then be WAA2 [NP 2sg]
    [CP ngo5 m4 zoi6fu4 keoi5 laa1]
    [CP 1sg Neg care 3sg Part]
    ‘In the first year, (s)he blamed me for not taking care of
    him/her.’

c. keoi5 bei2 jan4 waab6 [CP m4 sik1 jing1man2]
    3sg Pass people WAA2 [CP Neg know English]
    ‘(S)he is condemned for not knowing English.’

d. jan4dei6 waab6 [NP keoi5] waab6 [m4 sik1 jing1man2]
    people WAA2 [NP 3sg ] WAA [ Neg know English]
    ‘People condemn him/her for not knowing English.’

---

6 It is worthy noting a methodological point here, as Matthews (2003:8) notes that some of these sentences, when presented in isolation to a lay Cantonese speaker, produce a “puzzled reaction.” He observes that presenting such sentences out of the blue “is no way to test the productivity of a structure so dependent on discourse,” as many of the constructions belong to colloquial register and their use is closely tied to the discourse context. Therefore, this paper assumes the same methodological concern, stipulating that “(w)hile some aspects of the phenomenon can be discussed on the basis of single sentences, the need for such a discourse context should be borne in mind in the following discussion.”
2.2 The Clausal Complement-introducing waa6

In addition to the verbal usages of waa6, waa6 can also be used to introduce a complement clause. From the analysis of the preliminary corpus data, it is found that a number of predicates can co-occur with waa6 as the complement-clause introducer. These predicates fall into three main categories: ‘say’ predicates indicating any forms of speech; cognition and perception predicates describing the speaker’s mentality or perception; and informative predicates. This following list shows all the predicates found to co-occur with waa6 in the two corpora, with waa6 in a complement-introducing position. The list of actual instances in which waa6 were found is given in Appendix II.

<table>
<thead>
<tr>
<th>‘say’ predicates:</th>
<th>cognition and perception predicates:</th>
<th>informative’ predicates:</th>
<th>Other predicates:</th>
</tr>
</thead>
<tbody>
<tr>
<td>gong2 ‘say’</td>
<td>gok3dak1 ‘think’ / soeng2 ‘think’ /</td>
<td>tung1zi1 ‘inform’</td>
<td>hai6 ‘be’</td>
</tr>
<tr>
<td>tai4 / tai4dou3 /</td>
<td>‘believe’</td>
<td>se2 ‘write’</td>
<td></td>
</tr>
<tr>
<td>tai4kap6</td>
<td>‘plan’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘mention’ man6 ‘ask’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>daap3 ‘answer’</td>
<td>hei1mong6 ‘hope’ / lam2 ‘think’ /</td>
<td></td>
<td></td>
</tr>
<tr>
<td>giu3zou6</td>
<td>tai2 ‘see’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘be called’ ak1 ‘cheat’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>jyun3 ‘grumble’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>san3 ‘grouch’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>giu3 ‘shout’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>laai6 ‘blame’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>‘informative’ predicates:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>teoi1 ‘refuse’</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tam3 ‘comfort’</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

From the data, it can be seen that predicates that can co-occur with waa6 in the complement-introducing position are mostly the ones that

---

7 It should be noted that the same also applies to the Mandarin shi ‘be’ as in:

```
xin de da-yi shi shuo...
letter DE main-points be explain
'The main point of this letter is to narrate…'
```

(Hwang 1998)
potentially select a clausal complement, i.e., the ‘say’ predicates, the ‘cognition and perception’ predicates, and the ‘informative’ predicates. Even for hai6 ‘be,’ which is not in these three categories, hai6 ‘be’ is still a potential predicate selecting a clause as its complement. While in some cases, waa6 appears unambiguously with a complementizer reading, many of the cases in which the ‘say’ and informative predicates are followed by a waa6-clause may be ambiguous between serial verb and complementizer interpretations, i.e., the bridging contexts postulated by Heine (2002). In such cases, both old (SVC) and new (complementizer) interpretations are available, even though the new meaning is favoured.

Finally, the corpora even show waa6 with some English predicates. In code-mixing situations, speakers may use waa6 to introduce propositional complements after an English predicate that potentially subcategorizes a that-clause. Note that such the data are found in both our Cantonese corpora:

(8) ngo5 dou1 m4 EXPECT waa6 [zan1hai6 wui5 hai6] (CRC)  
1sg all Neg expect WAA [really Mod is]  
‘I didn’t expect that would really be true.’

(9) keoi5 ENCOURAGE ngo5 waa6 kei4sat6 laam4sing3 (HKCAC)  
3sg encourage 1sg WAA actually men  
dou1 jau5 hou2 zan1zi3 ge3 jau5ji4  
all have very sincere Poss friendship  
‘He encouraged me to believe that actually men also have sincere friendship.’

This phenomenon may suggest that intrinsically, waa6 may be understood and used as a complementizer. When speakers encounter a complement-taking predicate which potentially selects a complementizer that in English, they probably select the word waa6 as the complementizer. Even if the predicate actually does not allow a that-complement in English, like encourage in (9), the code-mixing speaker probably foresees its potential of taking a propositional complement and picks out waa6 to introduce the complement because of overgeneralization. Or, speakers may actually be forced to choose waa6.

---

8 The verb encourage only allows an NP followed by a that-clause in extrapositions, e.g.,  
It encourages me that she is doing well.
Cantonese Complementizer *waa6*

because it is the only complementizer available to follow a verb in Cantonese. While Example (9) may be ambiguous as to whether it takes a serial verb or complementizer reading, i.e., between ‘he encouraged me saying […]’ or ‘he encouraged me (to believe) that […].’ the serial verb reading in (8) is definitely incompatible with the context. This usage probably corresponds to Heine’s (2002) ‘switch context,’ in which the source meaning (verbal usage of *waa6*) is no longer compatible.

After the preliminary analysis of the Cantonese corpus data, we have seen the plausibility of analyzing *waa6* into three different words with different meanings, syntactic properties, and subcategorizations. From the fact that *waa6* may co-occur with various predicates, it is likely that *waa6*, in fact, functions as a complementizer that introduces a complement clause in some cases. To justify this postulation, more syntactic and semantic evidence is necessary. This will be the main focus of the following sections.

3. SYNTACTIC EVIDENCE

This section aims to evaluate the syntactic evidence that *waa6*, apart from its verbal usage, may function as a complementizer in some constructions. The first part focuses on the different argument selection properties of different senses of *waa6*, justifying them as separate lexical entries. The second part demonstrates the loss of the verbal behaviour of *waa6* in the complement-introducing position, which involves the inability of being aspect marked and undergoing verb-doubling. The third part will be a constituency test proving that *waa6* forms a constituent with its following IP.

3.1 Argument Selection

As a ‘say’ verb, *waa6*, can select an optional PP as its goal argument (10), such as *gong2* (11). However, in the complement-introducing

---

9 In code-mixing situations, it is likely that speakers select *waa6* in preference to *that* due to constraints on code-mixing which generally exclude ‘system’ morphemes (i.e., ‘closed-class’ words, such as specifiers, quantifiers, or complementizers. See Chan (1998).
position, the goal argument can only follow the first verb gong2, but not waa6 (as shown in (12) and (13)). The ungrammaticality of (13) shows that waa6 in (13) cannot be waa6i in a SVC with gong2 and that the verbal behaviour of waa6i, which takes a PP argument has been lost. It is therefore likely that waa6 in (12) is the complementizer waa6i.

(10) keoi5 waa6 [pp bei2 ngo5 teng1] lei5 wui5 lai4
    3sg say [pp to 1sg listen] you Mod come
    ‘(S)he said to me that you would come.’

(11) keoi5 gong2 [pp bei2 ngo5 teng1] lei5 wui5 lai4
    3sg say [pp to 1sg listen] you Mod come
    ‘(S)he said to me that you would come.’

(12) keoi5 gong2 [pp bei2 ngo5 teng1] waa6 lei5 wui5 lai4
    3sg say [pp to 1sg listen] WAA you Mod come
    ‘(S)he said to me that you would come.’

(13) *keoi5 gong2 waa6 [pp bei2 ngo5 teng1] lei5 wui5 lai4
    3sg say WAA [pp to 1sg listen] you Mod come

This argument selection property is a useful tool to identify the status of waa6 in many constructions appearing to be SVC’s and it will be used in later sections.

3.2 Loss of Verbal Behaviour

During the re-analysis process, semantic bleaching often occurs on a grammaticalized item and some of the syntactic properties exhibited by the original item are then lost (Lord 1976; Hopper and Traugott 1993). Regarding the case of waa6, it is observed that the verbal properties of waa6i, such as aspect marking and undergoing verb-doubling, are lost, as shown by the following examples.

Aspect marking

Generally, in SVC’s, either verb can be aspect marked, as (14) and (15). As a ‘say’ verb, waa6i can be aspect-marked by gwo3 or zo2 (17), just like other ‘say’ verbs such as gong2 (16). In (18), when waa6 follows a ‘say’ verb, gong2, only gong2 can be aspect-marked (18a), but not waa6 (18b).
13

Cantonese Complementizer *wa6*

(14) a. *ngo5 pui4 keoi5 sik6_gwo3/zo2 faan6*  
    1sg accompany 3sg eat_Asp  
    ‘I have had dinner with him/her.’

b. *ngo5 pui4_gwo3/zo2 keoi5 sik6 faan6*  
    1sg accompany_Asp 3sg eat  
    ‘I have had dinner with him/her.’

(15) a. *keoi5 heoi3_gwo3/zo2 king1 [IP gaa1 m4 gaa1 jan4gung1]*  
    3sg go_Asp discuss [IP add Neg add salary]  
    (bat1gwo3 mei6 jau5 kyu3t3ding6)  
    (but Neg have decision)  
    ‘(S)he has tried to discuss whether to have a pay rise (but there hasn’t been a decision).’

b. *keoi5 heoi3 king1_gwo3/zo2 [IP gaa1 m4 gaa1 jan4gung1]*  
    3sg go discuss_Asp [IP add Neg add salary]  
    (bat1gwo3 mei6 jau5 kyu3t3ding6)  
    (but Neg have decision)  
    ‘(S)he has tried to discuss whether to have a pay rise (but there hasn’t been a decision).’

(16) *keoi5 gong2_gwo3/zo2 [IP lei5 wui5 lai4]*  
    3sg say_Asp [IP 2sg Mod come]  
    ‘(S)he said you would come.’

(17) *keoi5 waa6_gwo3/zo2 [IP lei5 wui5 lai4]*  
    3sg WAA_Asp [IP 2sg Mod come]  
    ‘(S)he said you would come.’

(18) a. *keoi5 gong2_gwo3/zo2 waa6 [IP lei5 wui5 lai4]*  
    3sg say_Asp WAA [IP 2sg Mod come]  
    ‘(S)he said that you would come.’

b. *keoi5 gong2 waa6_gwo3/zo2 [IP lei5 wui5 lai4]*  
    3sg say WAA_Asp [IP 2sg Mod come]  
    ‘(S)he said that you would come.’

One possibility is to suggest that *waa6* in (18) has become a defective verb that cannot take any aspect markers, like the verb *hai6* ‘be’ (c.f. *hai6_zo2, *hai6_gwo3*). Another possibility is that in a complement-introducing position, *waa6* has lost its verbal property and become a complementizer. The confusion is expected and predicted in a bridging context, in which both serial verb and complementizer readings are possible.
Verb-doubling test

In Saramaccan and Sranan, we can observe a similar phenomenon, i.e., a lexical item serving both as a verb and as a that-type complementizer (refer to (2) and (3)). To distinguish between the verbal function and the complementizing function, Lefebvre (2001) uses the verb-doubling test, which is also known as ‘verb-topicalization’ (Matthews and Yip 1998) or ‘verb-fronting’ (Matthews 2003). In Creoles, the basic function of verb-doubling is for emphasis and concession. (Matthews 2003). As a verb, taki/táa may participate in the verb doubling phenomena, but as a complementizer it cannot. The contrast in grammaticality between (19a) and (19b) is shown in Lefebvre (2001):

Saramaccan:
(19) a. Verb:
   táa a táa: m’e nángó
   say 3sg say: 1sg-Neg Asp-go
   ‘He really said: I am not leaving.’
   b. Complementizer:
   * táa a táki táa á búnu
   say 3sg say say 3sg-Neg good
   (Lefebvre 2001:113)

Verb-doubling phenomenon is also a productive phenomenon in Cantonese syntax, carrying a similar emphatic or concessive sense as in the Creoles (Matthews and Yip 1998), as (20).

(20) fan3 zau6 fan3_zo2 hou2 loi6 laa1
    sleep then sleep_Asp very long Part
   (daan6hai6 zung6 mei6 fan3_zoek6)
   (but still Neg sleep_Asp)
   ‘He went to sleep some time ago (but hasn’t been asleep yet).’

In serial verb constructions, both the main verb and the serial verb may undergo verb-doubling, such as pui4 and sik6 in (21), as well as heoi3 and king1 in (22):
When the verb-doubling test is applied to *waa6*, it appears that *waa6* only allows verb-doubling in its verbal usage, as in (23a); but disallows verb-doubling in its complementizer usage, i.e., when it follows the predicate *gong2* ‘say’ (23b), or *lam2* ‘think’ (23c):

(23) a. Verb:

```
  waa6  keoi5 zau6  waa6_gwo3  waa6 [IP m4 dak1haan4]
  WAA  3sg    then    WAA_Asp  WAA [IP Neg free]  
  (daan6  hau6mei1  jau6  lei4_zo2)
  (but  eventually  but  come_Asp)
```

‘(S)he really said that (s)he was not free (but she came eventually).’
Complementizer:

b. \{gong2 /*waa6\}  keoi5 zau6 gong2  waa6  [wp m4 dak1haan4]
   \{say / *WAA\}  3sg then say  WAA  [wp Neg free]
   (daan6 hau6mei1  jau6  lei4_zo2)
   (but  eventually  but  come_Asp)
   ‘(S)he really said that (s)he was not free (but she came eventually).’

c. \{lam2 /*waa6\}  keoi5 zau6 lam2  waa6  [wp m4 dak1haan4]
   \{think / *WAA\}  3sg then think  WAA  [wp Neg free]
   (daan6 hau6mei1  jau6  lei4_zo2)
   (but  eventually  but  come_Asp)
   ‘(S)he really thought that (s)he was not free (but she came eventually).’

The test shows that waa6 in the complement-introducing position lacks verbal behaviour, to the extent that it cannot undergo verb-doubling. According to the results of this test, we may draw a similar conclusion to that in the Creole languages discussed, namely that Cantonese waa6 functions more like a complementizer than a serial verb in such a construction.10

3.3 Constituency Test

If waa6 is a complementizer, it must form a constituent, namely CP, with the following IP. Even though constituency tests do not directly prove the complementizer status of waa6, they are necessary for identifying waa6-clauses as single constituents. If waa6-clauses fail to be proven as constituents, the postulation of waa6 as a complementizer

10 However, it is observed that the second verb in certain Cantonese serial verb constructions may not undergo verb-doubling either, as in (i):

i. \{co5 /*faan1\}  ngo5  zau6  co5_gwo3  sailtit3  [faan1  uk1kei2]
   \{sit/back\}  1sg then sit_Asp  West Rail  [back home]
   ‘I really took the West Rail home.’

In (i), co5 is the first verb and faan1 is the second verb in the SVC. It is proposed that the unacceptability of (i) may be caused by the complement NP uk1kei2 ‘home,’ c.f.:

ii. \{co5 /*faan1\}  ngo5  zau6  co5_gwo3  sailtit3  faan1
   \{sit/back\}  1sg then sit_Asp  West Rail  back
   (heoi3  zau6  mei6  si3_gwo3)
   (go then Neg try_Asp)
   ‘I’ve tried taking the West Rail back home (but not traveling from home).’
selecting an IP complement will be undermined. In this section, coordination tests are carried out to investigate the constituent structure of waa6-clauses.

Assuming waa6 is a complementizer following predicates like gong2 ‘say’ or lam2 ‘think,’ it should form a CP constituent with the following IP. Example (24) shows the possible coordination of waa6-clauses (CPs). The coordinating conjunction tung4maai4 ‘and’ is optional as in Cantonese parallel coordination can be generally achieved by juxtaposition (Matthews and Yip 1994) However, the coordinating conjunction waak6ze2 ‘or’ is obligatory.

(24) a. With gong2 ‘say’:
   keoi5 gong2_gwo3 [CP waa6 m4 dak1haan4,] (tung4maai4)
   3sg say_Asp [CP(WAA) Neg free,] (and)
   [CP(waa6) m4 soeng2 gin3_dou2 ngo5,] so2ji3 m4 heoi3
   [CP(WAA) Neg want see_Asp 1sg,] so Neg go
   ‘(S)he said that (s)he was not free and that (s)he didn’t want to see me, so (s)he would not come.’

b. With lam2 ‘think’:
   ngo5 lam2_zyu6 [CP waa6 keoi5 m4 heoi3,] waak6ze2
   1sg think_Asp [CP(WAA) 3sg Neg go,] or
   [CP(waa6) ngo5 dak1haan4,] sin1 wui5 ceot1jin6
   [CP(WAA) 1sg free,] only_when Mod appear
   ‘I think I will be there if (s)he doesn’t go or if I am free.’

For the predicate waa62 ‘blame/condemn’ however, the sentences appear to be slightly worse than the examples in (24), probably because it becomes unclear whether the repeated waa6 is waa61, waa62, or waa63:

(25) a. With waa62 ‘blame/condemn’:
   ji4gaa1 di1 jan4 seng4jat6
   now Cl people always
   waa62 hoeng1gong2 ge3 daai6hok6saang1,
   condemn Hong Kong Poss university students,
‘Now, people always condemn the university students in Hong Kong for having a low ability in English and a low ability in Chinese.’

b. In passive voice:

‘Now, the university students in Hong Kong are always condemned for having a low ability in English and for having a low ability in Chinese.’

It is observed that these sentences sound better if a conjunction, like tung4maai4 ‘and,’ is added:

(26) a. With waa6, ‘blame/condemn’:

‘Now, people always condemn the university students in Hong Kong for having a low ability in English and a low ability in Chinese.’
b. In passive voice:

\[\text{hoeng1gong2 ge3 daai6hok6saang1} \]
\[\text{Hong Kong Poss university students} \]
\[\text{seng4jat6 bei2 jan4 waa6;} \]
\[\text{always Pass people condemn} \]
\[\text{[CP} \text{waa6 jing1man2 m4 hou2,} \]
\[\text{[CPWAA English Neg good,]} \]
\[\text{tung4maai4 [CP(waa6) zung1man2 m4 hou2]} \]
\[\text{and [CP(WAA) Chinese Neg good]} \]

‘Now, the university students in Hong Kong are always condemned for having a low ability in English and for having a low ability in Chinese.’

It is not clear why coordination without an explicit conjunction \text{tung4maai4} ‘and’ appears to be less grammatical. In cases without the second \text{waa6}, we have IP-, not CP-, conjunctions. Although the possible coordination of \text{waa6}-clauses does not directly prove the complementizer status of \text{waa6}, it ascertains the fact that \text{waa6} forms a constituent with the following IP and is consistent with the previous analyses of \text{waa6} as a potential complementizer.

4. SEMANTIC EVIDENCE

After demonstrating the syntactic evidence, this section will provide some semantic evidence that supports the proposal that \text{waa6} is a complementizer, rather than a verb in SVC with other predicates. The co-occurrence of two \text{waa6}’s in sentences will be first discussed, suggesting that there may be two different \text{waa6}’s with different semantic/pragmatic functions. The second part analyzes the occurrence of \text{waa6} following non-factive predicates, a context where \text{waa6} cannot be interpreted as a ‘say’ verb.

4.1 The Co-occurrence of \text{waa6}’s

For SVC, the co-occurrence of different senses of the same predicate is found to be problematic, such as\textsuperscript{11}:

\textsuperscript{11} It is suggested that the longer the intervening elements between the two verbs, the
Even though Hwang (1998) proposes that *waa6 may have different meanings in different situations, such as as a ‘saying’ verb, or as a complementizer, she suggests that the co-occurrence of *waa6’s in a series is not possible, as in:

(28) keoi5 waa6 (*waa6) heong1gong2 hou2 deifong1
     3sg say (*WAA) Hong Kong good place
     ‘(S)he said that Hong Kong was a nice place.’
     (Hwang 1998)

The appearing ungrammaticality of the immediate co-occurrence of *waa6’s may be due to the ‘Obligatory Contour Principle’, which states that consecutive phonological identical/ similar features are banned in underlying representations. Chappell (to appear) relates this observation with Chao’s (1968) rule of haplology in Chinese linguistics, whereby ‘two consecutive instances of the same syllable in Mandarin simplex clauses result in the omission of the second’. The co-occurrence of *waa6’s can only be grammatical in some structures, such as with intervening aspectual markers, particles, pauses. These examples are shown in (29) to (31).

With aspectual marker intervening:
(29) keoi5 waa6 gwo3 waa6 [ip m4 dak1haan4 lai4]
     3sg say Asp WAA [ip Neg free come]
     ‘(S)he said that (s)he was not free to come.’

With particle intervening:
(30) keoi5 waa6 le1 waa6 [ip m4 dak1haan4 lai4]
     3sg say Part WAA [ip Neg free come]
     ‘(S)he said that (s)he was not free to come.’

more acceptable sentence becomes. (Tang 1998a)

---

Cantonese Complementizer waa6

With pause:

(31) keoi5 waa6 <p> waa6 [IP m4 dakhaan4 lai4]
    3sg say <p> WAA [IP Neg free come]

‘(S)he said that (s)he was not free to come.’

It is interesting to note that in other Sinitic languages, such as Chaozhou and Taiwanese, the immediate co-occurrence of the verbal ‘say’ and the re-analyzed complementizer ‘say’ is accepted without any intervening elements, which may reflect a more ‘advanced’ stage of grammaticalization of Chaozhou tã and Taiwanese kong:

(32) a. i tà tà “lu₅ tsio sê tso m tui”
    3sg say say 2sg this way do not right

‘(S)he said “What you did was wrong”.’
(Xu and Matthews 2005)

b. lin ban-chek-à kah goâ kông kông
goa oân-â ù choè siân-sù là
1sg also have do good deed Part

‘Your youngest uncle told me that I had also done some good deeds.’
(Chappell to appear)

Apart from these constructions with intervening elements, it is also possible to have an immediate surface adjacency of waa6 without intervening elements, at least in the case of transitive verb waa6; ‘blame/condemn’, as (33). Note that it is only a case of surface adjacency, as assuming a movement analysis for passivization, there is a trace left between the two waa6’s.

Immediate (surface) adjacency without intervening elements:

(33) keoi5, bei2 jan4 waa6 ti, [CP waa6 m4 sikl jing1man2]
    3sg, Pass people condemn ti, [CP WAA Neg know English]

‘(S)he is condemned in that she doesn’t know English.’

The two superficially adjacent waa6’s may be made possible by the trace left after the NP-movement. It is also possible that the lexical sense
of the transitive \textit{waa6} is maximally distinct from the semantically bleached complementizer \textit{waa6}, facilitating their juxtaposition.

It might be argued that the second \textit{waa6} is not necessarily \textit{waa61}, but actually \textit{waa61} `say' in a serial construction with \textit{waa62} `blame/condemn'. Yet, this possibility can be ruled out by the ungrammaticality of (35) where \textit{waa6} is not able to take the PP argument. (Please refer back to (12)-(13).)

(34) \textit{keoi5 bei2 jan4 waa6 waa6} [\textit{m4 sik1 jing1man2}]
\begin{itemize}
\item 3sg Pass people condemn WAA
\item \textit{Neg know English}
\end{itemize}
`(S)he is condemned for not knowing English'

(35) *\textit{keoi5 bei2 jan4 waa6 waa6} [\textit{pp bei2 keoi5 teng1}]
\begin{itemize}
\item 3sg Pass people condemn WAA
\item to 1sg listen
\item \textit{m4 sik1 jing1man2}
\item \textit{Neg know English}
\end{itemize}
`(S)he was condemned by people, telling her that she doesn’t know English.'

The ungrammaticality of (35) implies that the second \textit{waa6} that introduces the complement clause cannot be there in its verbal usage, but as a complementizer (\textit{waa63}).

4.2 Non-factive Predicates

Non-factive predicates presuppose the complement clause as being false (Ludlow 1996:406-461). For example, in (36a), in order to make the non-factive reading felicitous, there exists a presupposition that proposition (36b) must be false.

(36) a. He thought that she went to the church. (non-factive reading)
   b. She went to the church.

Owing to its occurrence following non-factive predicates, such as \textit{ji5wai4} `think,' \textit{wai4ji4} `suspect,' and \textit{lam2zyu6} `think,' \textit{waa63} cannot be interpreted as a `say' verb. For example, (37a) only presupposes (37b) as false, but not (37c). Therefore, the complement proposition should be (37b), not (37c); hence, it forces \textit{waa6} to be given a complementizer reading, like the English \textit{that}, instead of a verbal meaning (\textit{waa61} `say'). This corresponds to the `switch context' stage of the grammaticalization
Cantonese Complementizer *waa6*

pathway (Heine 2002), in which the new context is incompatible with
the source meaning, i.e., the verbal meaning of *waa6* as ‘say’ is
incompatible within the context.

(37) a. *ngo5 zung6 ji5wai4 waa6* [lei5 m4 dak1haan4 tim1]
   1sg  again  think  WAA  [2sg  Neg  free  Part]
   ‘I thought that you were not free.’
b. *lei5 m4 dak1haan4*  (F)
   2sg  Neg  free
   ‘You were not free.’
c. *ngo5 waa6 lei5 m4 dak1haan4* (T/F ??)
   1sg  say  2sg  Neg  free
   ‘I said that you were not free.’

Sometimes, *waa6* sentences with a potential non-factive reading may be
thought to be ambiguous, as in (38a). Some may argue that (38a) may
have two possible readings and the verbal interpretation of *waa6* is still
salient. In fact, (38a) only entails the non-factive reading of *waa6* as
being a complementizer, i.e., only the presupposition of *heoi3 bak1ging1*
‘go to Beijing’ is false, as demonstrated in (38b). It does not presuppose
the falsity of the proposition in (38c) *waa6 heoi3 bak1ging1* ‘say to go to
Beijing’. Thus, *waa6* following a non-factive predicate, as in (38a), rules
out the possible verbal interpretation of *waa6*.

(38) a. *ngo5 lam2_zyu6 waa6* [ip heoi3 bak1ging1]
   1sg  plan_Asp  WAA  [ip  go  Beijing]
   ‘I planned to go to Beijing (but I didn’t).’
   (Reading 1: non-factive, *waa6* as complementizer)
   * ‘I planned to say that I would go to Beijing.’
   (Reading 2: *waa6* as ‘say’)
b. *ngo5 lam2_zyu6 waa6* [ip heoi3 bak1ging1],
   1sg  plan_Asp  WAA  [ip  go  Beijing],
   dim2zi1 mou5  cin2
   but  Neg  money
   ‘I planned to go to Beijing, but I have no money.’
   (Presupposition: *heoi3 bak1ging1* ‘go Beijing’ is false)
c. *ngo5 waa6* [ip heoi3 bak1ging1]  (T/F ??)
   1sg  WAA  [ip  go  Beijing],
   ‘I said I went to Beijing.’ (T/F ??)
The occurrence of *waa6* with non-factive predicates rules out the possible verbal interpretation of *waa6*. It provides the switch context (Stage III), in which the source meaning (as a ‘say’ verb) is no longer compatible with the non-factive context of the sentence; hence foreshadowing that the target meaning (as a complementizer) provides the only possible interpretation.

5. FURTHER GRAMMATICALIZATION OF *WAA6* INTO A SENTENCE-FINAL PARTICLE (SFP)

Heine (2002) hypothesizes a four-stage grammaticalization pathway in which various contexts are involved. The four-stage grammaticalization pathway is represented as follows:

Stage I II III IV  
Initial stage > Bridging context > Switch context > Conventionalization

The pathway for ‘say’-complementizer in Sinitic languages may possibly be (Xu and Matthews 2005):

I II
Initial stage > Bridging context >
‘say’ verb Quotative marker
(following speech act verbs)

III IV
Switch context > Conventionalization
Complementizer Complementizer
(following cognition/perception verbs) (following epistemic modals)

For some complementizers developed from ‘say’ verbs, all stages in the entire pathway have been observed, such as *tã* in Chaozhou:

(39) a. **Initial Stage (I):** *tã* as a lexical verb

\[
i \quad siå \quad tã  
\]

3sg first say

‘He will say (it) first.’
b. Bridging context (II): tã as a quotative marker

\[
\begin{align*}
\text{3sg say say } & \text{2sg this way do not right} \\
\text{‘(S)he said “what you did was wrong”.'}
\end{align*}
\]

c. Switch context (III): tã as a complementizer following cognition/perception verbs

\[
\begin{align*}
\text{1sg think *(say) tomorrow should go.to.work} \\
\text{‘I thought I had to go to work tomorrow.’}
\end{align*}
\]

d. Conventionalization (IV): tã as a complementizer following epistemic modals

\[
\begin{align*}
\text{3sg can say three nights three days not.need sleep} \\
\text{‘(S)he can go without sleep for three days and three nights.’}
\end{align*}
\]

(Xu and Matthews 2005)

Unlike the Cantonese waa6, the Chaozhou tã is in a more ‘advanced’ stage of grammaticalization. For example, the immediate co-occurrence of tã’s is acceptable without any intervening elements, as (39b); whilst the co-occurrence of Cantonese waa6 is problematic, and is acceptable only with an intervening element. (Please refer back to §4.1 on the co-occurrence of waa6’s.) Moreover, the complementizer usage of the Chaozhou tã is conventionalized when it follows a cognition/perception predicate where it becomes an obligatory element, as in (39c). The Cantonese waa6, on the other hand, is never observed to be obligatory as a complementizer. For the case of the Taiwanese kong, Chappell (to appear) suggests that it appears to be ‘well-advanced’ in terms of the stages of grammaticalization. Apart from the possibility of immediate co-occurrence (32b), she observes that kong may follow preceding emotion, psych-verbs and possibly other stative verbs, allowing ‘what might be normally non-complement taking verbs.’ For instance, hoanlo kong ‘be anxious that’ and hoa-hi kong ‘to rejoice that’ are found in Taiwanese corpora, but not in the two Cantonese corpora cited in this study. If the Chaozhou tã or the Taiwanese kong is said to be fully

13 The psych-verbs described in Chappell are not the cognition/perception verbs discussed in the paper. They are the ‘verbs’ that are more adjective-like and depict personal emotions. Examples of these verbs include hoanlo ‘be anxious’, or hoa-hi ‘be rejoice’. 

25
grammaticalized from a ‘say’ verb to a complementizer, Cantonese *waa6* should be regarded as in its ‘earlier’ stages of grammaticalization. As examined in the paper, there have been many clear instances of *waa6* in a bridging context (Stage II) serving as quotative markers following speech act verbs. Some examples appear to suggest instances of *waa6* in a switch context (Stage III), in which the verbal meaning of *waa6* is disallowed. These include the examples of *waa6*-clauses following cognition/perception predicates, especially the non-factive predicates. Instances of *waa6* in switch contexts indicate a later stage of grammaticalization. Yet the grammaticalization of *waa6* is still quite far from being fully grammaticalized, as *waa6* is still an optional element, but not a conventionalized obligatory complementizer, when introducing the clausal complement.

The previous sections have confirmed the complementizer status of Cantonese *waa6* as distinct from its verbal usage. As mentioned in §1, complementizers re-analyzed from verbs of ‘saying’ exist not only in Cantonese: the Taiwanese *kong* is also a complementizer re-analyzed from ‘say’ verbs (Hwang 1998). Simpson and Wu (2002) and Wu (2004) have worked on the syntax of Taiwanese *kong* and by means of phonological evidence 14, they postulate that the Taiwanese *kong* is further grammaticalized into a sentence final particle carrying an emphatic assertion sense. Examples of *kong* as a complementizer followed by its complement IP clause (as in (40)) and *kong* as a sentence final particle (as in (41)) are shown as follows15:

\[
\text{(40) a. } A\bullet-hui \ siong\bullet-sin \bullet \ kong \bullet \ A\bullet-sin \ m\bullet \ \text{lai} \\
A-hui \ \text{believe} \ \text{KONG} \ A-sin \ \text{Neg} \ \text{come} \\
\text{‘A-hui believes that A-sin is not coming.’}
\]

---

14 Simpson and Wu (2002) and Wu (2004) make use of phonological evidence involving ‘tone sandhi,’ which is a well-known property of southern Min dialects such as Taiwanese. It is observed that a syllable may not undergo tone sandhi if it occurs sentence-finally. However, the sentence-final *kong* is found to have undergone tone sandhi which indicates, according to their analysis, that the complement IP clause of *kong* has actually moved up to the left of *kong*, leaving *kong* in the surface sentence-final position.

15 The symbol of a dot • following the word indicates the tone sandhi change.
b. A-hui liau•chun•kong• A•-sin si• tai•pak• lang
A-hui thought KONG A-sin is Taipei person
‘A-hui thought that A-sin is from Taipei.’
(Wu 2004:94-95)

(41) a. A•-hui siong• A•-sin m• lai kong•
A-hui think A-sin Neg come KONG
‘A-hui thinks A-sin is not coming.’

b. A•-hui siong• kong• A•-sin m• lai kong•
A-hui think KONG A-sin Neg come KONG
‘A-hui thought that A-sin is not coming.’
(Simpson and Wu 2002:80)

Similarly, in a recent study on the Mandarin shuo, Wang et al (2003) provide a descriptive study of shuo and suggest that when ‘shuo is attached to the initial or final positions of an utterance,’ it functions as a ‘discourse marker,’ expressing ‘the emotional state of the speaker’ (Wang et al 2003:482).

(42) a. (SK-II) shuo meitian zhi shui yi ge xiaoshi,
(SK-II) say everyday only sleep one Cl hour
..ni xiangxin ma?
you believe SFP
‘It is said that you only need to sleep for one hour (with SK-II), do you believe it?’

b. wo yiwei ta ershiwu liu shuo16
I thought he twenty-five -six say
‘I thought he was twenty-five or -six say.’

Although the Cantonese waa6 does not appear in the sentence final position as the Taiwanese kong does, some cases with waa2 in the sentence final position indicating an interrogative sense were found in the corpus data:

---
16 The author has confirmed the sentence-final usage of shuo in Mandarin with native Mandarin speakers. It has been reported that the usage likely originated from Taiwanese Mandarin speakers, but has been spread to mainland China via Internet and entertainment programmes. The usage can be occasionally found on Internet and newsgroups in mainland China. The same usage is reported in Chappell (to appear).
This case of waa2 appears to occupy a similar sentence final position to the Taiwanese kong and it is phonologically similar to the complementizer waa6. However, unlike the Taiwanese kong, waa2 does not have the same pronunciation as waa6 and it does not signal any emphatic sense, but rather an interrogative sense. Phonologically, waa2 has a rising tone and waa6 is a low tone word. In Cantonese, there are many cases of tone 2 words deriving from low tone words, e.g., jan from jan4 ‘person.’ It is possible that waa2 carries an interrogative function and serves as an echo question particle. Clearly, both the SFP waa2 and the complementizer waa63 are derived from the lexical verb waa61. The question is whether the two pathways of grammaticalization are related (i.e., whether the SFP waa2 is historically derived from the complementizer waa63). At this stage, to determine the grammatical status of waa2 simply by its reference to the case of the Taiwanese kong might be incautious.

Tang (1998b) has suggested that waa2 is an echo question particle occupying the COMP position and that waa2 cannot occur in subordinating clauses. It appears that waa2 occurring in matrix clauses and waa6 introducing subordinating clauses are in complementary distribution. Tang suggests that the omission of SFP’s (including waa2) in embedded clauses can be accounted for by the possibility of having a head-initial complementizer waa6 in embedded clauses. This proposal requires further research to ensure its justification. Further research is also needed to investigate the relation of waa2, if any, to the complementizer waa6; and its typological relationship to the Taiwanese kong. Being diachronically related in the sense that both waa2 and waa63 appear to be derived from the verb waa61, the relation of waa6 and waa2 may serve as a good starting point for the investigation of the relation between head-initial complementizers introducing subordinating clauses and head-final complementizers occurring in matrix clauses.
6. SOME CONCLUDING REMARKS: FROM HEAD-INITIAL COMPLEMENTIZERS TO WORD ORDER TYPOLOGY IN CHINESE

This paper has ascertained the status of *waa6* as a complementizer and introduced a new member to the Chinese COMP category. In the studies of Chinese complementizers, it has been proposed in the literature that relative clause markers, like *de* in Mandarin and *ge3* in Cantonese are complementizers (Cheng 1986; Law 1990; Ning 1993; Xu 1997). Unlike *waa6*, which is a head-initial complementizer selecting propositional clauses as its complements, relative clause markers *de* or *ge3* can be selected by nouns and introduce noun-complement clauses. The clauses selected can be modifying clauses, as in (45), and also complement clauses, as in (46).

(45)
a. *de* or *ge3* selects a modifying clause:

\[
[ei zuotian zai wuli yong bi xie xin] \text{de reni}
\]
\[\text{yesterday in room with pen write letter DE man}\]
'the man who wrote the letter with a pen in the room yesterday.'

b. \[NP [ei seng4jat6 wan6dung6 ge3] sai3lou6za2i2\]
\[all-day exercise children\]
\[bei2gaau3 gin6hong1\]
'Children who always exercise are healthier.'

(46)

*de* or *ge3* selects a complement clause:

\[NP [ta zuotian zai wuli yong bi xie xin]\]
\[NP [he yesterday in room with pen write letter]\]
\[de qingkuang] hen qiguai\]
'DE situation very strange'
'The situation that he wrote a letter with his pen in the room yesterday was very strange'
b. [[hok6saang1 faan2deoi3 haaau6zoeng2 ge3] wut6dung6] 
student oppose principal activity
jyut6 lai4 jyut6 koeng4liit6
more come more strong

‘The activities that the students engage in to oppose the principal are getting stronger.’

At least in this usage in (46), de or ge3 appears to be a complementizer which is head-final and selected by a noun. If waa6 has been proven to be a head-initial complementizer, it seems to conflict with the previous assumptions in the literature that Chinese complementizers are head-final. One possibility could be that the conflict between head-initial and head-final complementizers is related to the conflict between head-initial VP’s and head-final NP’s. In Chinese, as NP’s tend to be head-final and as de/ ge3 are relativization markers which link a modifying/ complement phrase with the head noun, it is expected that de/ ge3 are head-final. For waa6, however, as verbs are head-initial within the VP and as waa6 is reanalyzed from a ‘say’ verb, selecting an IP propositional complement, waa6 is expected to be head-initial.

This proposal is predictable in view of theories on word-order typology, such as the Branching Direction Theory (BDT) proposed by Dryer (1992) or the principle of Early Immediate Constituents (EIC) developed by Hawkins (1994). Both theories suggest that regarding word-order in a language, consistent left-branching like (47a) or right-branching structures like (47b) are preferred over a mixture of left- and right-branching (47c-d).

17 The ‘Early Immediate Constitutes’ (EIC) is defined in Hawkins (2001:5) as follows: ‘The human parser prefers linear orders that minimize CRD’s (by maximizing their IC-to-nonIC [or IC-to-word] ratios), in proportion to the minimization difference between competing orders.’
According to Hawkins (2002:101), for cognitive processing reason, languages prefer ‘minimal domains.’ Since the mixed word orders ‘involve larger viewing windows (CRDs)\(^{18}\),’ they are ‘dispreferred cross-linguistically.’

Contrary to typological predictions, the structure of NP in Chinese is inconsistent with the structure of VP. However, the branching structures within NP and VP appear to be internally consistent. According to the BDT and the EIC, consistent left-branching or right-branching structures are preferred over a mixture of left- and right-branching. Thus, if Chinese NP tends to be head-final (i.e., a left-branching structure as in (47a)), the complementizer position is expected to precede the head noun, as shown in (48):

\(^{18}\)The definition for a ‘Constituent Recognition Domain’ (CRD) in Hawkins (2001:5) is as follows: ‘The CRD for a phrasal mother node M consists of all non-terminal and terminal nodes dominated by M on the path from the terminal node that constructs the first IC on the left to the terminal node that constructs the last IC on the right.’ For further discussion of the parsing approach proposed by Hawkins, see Hawkins (1990, 1994, 2001, 2002).
(48) Left-branching structures in Chinese NP’s:

```
    COMP
   / \  
Noun
```

Similarly, if Chinese VP’s are head-initial (i.e., right-branching structure as in (47b)), the complementizer is expected to be head-initial (i.e., following the head verb), as shown in (49):

(49) Right-branching structures in Chinese VP’s:

```
   Verb
  /   
COMP
```

The next question that naturally follows is whether typologically the co-existence of head-initial and head-final complementizers in a language is possible. Primus (2001) constructs a table showing the basic details associating with head-initial and head-final patterns:
The table shows the ideal distinction between head-initial and head-final languages and it is noted that many languages are found to be less consistent with the shown pattern. For example, English is observed to be a less consistent head-initial language as it has word order patterns such as adjective – noun (AN), genitive/possessor – noun (GN), and adverb – adjective (AdvA); where Finnish is categorized as a less consistent head-final language as it is a language with SVO, AuxV, and NRel/RelN word order patterns. Regarding Chinese, it is similar to Finnish in that it also appears to be a less consistent head-final language with SVO, AuxV, article—noun and adjective–object word order patterns. According to Primus (2001, pp.856), ‘relative and complement clauses and complementizers are bad patterners in head-final languages,’ and which depart quite often from the cross-categorial harmony shown in Table 1. Regarding the typological word-order pattern of complementizers and relative/ complement clauses in head-final languages, Primus (2001) observes that:

‘…head-initial languages will have almost only CompS order whereas head-final languages will have both CompS and SComp.’

(Primus 2001:861)
Japanese is an example of a verb-final language with head-final complementizers; while Persian is a verb-final language with head-initial complementizers. However, Primus (2001) does not indicate whether the co-existence of CompS and SComp in the same language is possible. Siewierska (1998:791-810) lists a total number of 37 European languages that employ both NRel and RelN for NP’s containing relative clauses. Examples of these languages include Basque, Georgian, Gothic, Latin, Finnish, and Hungarian. The possible co-existence of NRel and RelN structures in the same language may imply a possibility of having a language with both head-initial and head-final complementizers, but there is not a single language reported to have both head-initial and head-final complementizers. Still, it is believed that Chinese may be the only possible candidate for having both types of complementizers because it is the only language reported to be head-initial in VP’s, but head-final in NP’s for relative clauses/noun-complement structures. Dryer (1992:82-87) examines the word-order properties in 625 languages from various language genera including Africa, Eurasia, Southeast Asia & Oceania, Australia-New Guinea, North America, and South America, and the only language found to have the order of verb—object (VO) and the order of relative clause—noun (RelN) is Chinese. Hawkins (1994) makes a similar observation:

‘(a language) combining verb-initial VP’s and prepositions with pronominal relatives, are exceedingly rare…Mandarin Chinese is the only language known to me that provides solid evidence for these structures…’

(Hawkins 1994:268)

Hence, it would not be a surprise if Chinese were the only language family found to have both head-initial and head-final complementizers. In such a case, de or ge3 may be the head-final complementizer selected by a noun to introduce the noun-complement. As relative/complement clause—noun structures in Chinese are head-final (RelN), the complementizer position within the NP is preferably at the right edge of the complement clause (i.e., head-final). On the other hand, waat6 reanalyzed from a ‘say’ verb might be a head-initial complementizer selected by a verb to introduce the verb-complement. As Chinese has head-initial VP’s, the complementizer position within the VP should be at the left edge of the complement clause (i.e., head-initial).
The proposal of having the head-initial complementizer *waab* in the VP, as opposed to the head-final complementizer *de* or *ge3* in the NP, sheds light on the head-initial and head-final word order differences in Chinese VPs and NPs, providing further supporting evidence for theories in word order typology, such as the Branching Direction Theory (BDT) (Dryer 1992) and the principle of Early Immediate Constituents (EIC) (Hawkins 1994), which predict consistent left-branching or right-branching structures in languages.

This paper has successfully distinguished the complementizer status of *waab* from its verbal status by various synchronic tests, including argument selection, aspect marking, verb-doubling, coordination tests, and proposition negation in non-factive predicates. Further, even though in its earlier stage of grammaticalization, it has proposed that in some cases, *waab* is a head-initial complementizer in the VP introducing IP complements. This paper has not only introduced a new member to the Chinese COMP category, it has also shed light on other possible research areas. For example, this paper suggests a possible case of the grammaticalization pathway where sentence-final particles, like *waab2*, are derived from complementizers, like *waab6*. In light of *waab6* being a head-initial Chinese complementizer, the paper relates to the special word order typology in Chinese and provides supportive evidence for theories in word order typology (e.g., BDT and EIC). It is hoped that the findings of this paper will be inspiring and stimulate further research in the studies of Chinese complementizers derived from ‘say’ verbs.

**APPENDIX I**

**ABOUT THE CORPORA:**

The Cantonese Radio Corpus (Francis, Yiu, Matthews, and Chu, in preparation)

This corpus contains samples of adult spoken Cantonese from four radio talk programs in Hong Kong, with a total of 43,283 lines of text transcribed and coded for syntactic analysis. The corpus constitutes part of a project entitled ‘Lexical and contextual effects on the grammar of syntactic categories’ which is funded by the Committee on Research and Conference Grants of the University of Hong Kong. The principal investigator on the project is Elaine J. Francis, the co-investigator is Stephen Matthews, and the research assistants are Winnie S.M. Yiu and Gene Y.F. Chu. The data for the corpus were collected by tape recording.
several Hong Kong radio talk programs and selecting a sample of four programs which span a variety of genres, including celebrity interviews, political and economic discussions, and relationship issues. A pre-publication version of the CRC was used for this research, and a corrected version is due to be submitted to the online TalkBank database.

The HKCAC Corpus (Law, Fung, and Leung, 2002)
The development of the Hong Kong Cantonese adult language corpus (HKCAC) was made possible by a grant from the Research Grant Council of Hong Kong (# HKU 5190/98H) to Sam-Po Law, Suk-Yee Fung, and Man-Tak Leung. The database consists of orthographic and phonetic transcriptions of phone-in programs and forums broadcast on the radio. The recordings, totaling more than eight hours, were made during the period from November 1998 to February 2000. Sixty-nine speakers in addition to the program hosts were recorded. The corpus has approximately 170,000 characters.

APPENDIX II
LIST OF SENTENCES CONTAINING INSTANCES OF WA46:
Note: This is not an exhaustive list of all instances encountered. For each predicate, only one or two instances are listed.

'say' predicates:
gong2 'say'
tau4sin1 lei5dei6 gong2 wa46 sing1ga3bo1 ge3 ji1liu4 ne1 just now 2pl say WAA Singapore Poss medical SFP

'Just now, you talked about the medical facilities in Singapore…'

(HKCAC)
ngo5 jat1lou6 dou1 mou5 tung4 keoi5 gong2 wa46
1sg always all Neg with 3sg say WAA

ngo5 tung4 faan1 keoi5 jat1cai4
1sg accompany Asp 3sg together

'I never told her that we were back together.'

(CRC)
Cantonese Complementizer `waa6`

**king1 `discuss`**
```
kam4jat6 ngo5 tung4 jat1go3 jan4  king1 gan2 waa6
```
yesterday 1sg with one person **discuss Asp WAA**
```
zik1 ji4gaal ge3 ngan4hong4 zung6 seoi1 m4  seoiljiu3 fan1hong4 nel
```
that now Poss bank still need Neg need branch **SFP**
```
‘Yesterday, I was talking to someone about whether banks still need branches nowadays.’
```

(HKCAC)

**tai4/tai4dou3/tai4kap6 `mention`**
```
keoi5 jat1zou2  tai4 gwo3  waa6  m4  jiu3
```
3sg early **mention Asp say** Neg want
```
‘she’d always mention that she didn’t want it.’
```

```
gam2 kei4sat6  baa1zi3gam1 jik6 dou1  tai4dou3  waa6  nel
```
then actually Bokhary also all **mention WAA SFP**
```
gam1ci3  lai4  gong2 ji5  dou3  hin3faat3  so2  jung4heoi2
```
this time for say already reach constitution that allow
```
ge3  gik6haan6 laa3
```
Poss limit **SFP**
```
‘Actually, Mr Justice Bokhary also mentioned that for this incident, it had already reached the limit of the constitution.’
```

(HKCAC)

**jan1wai6 colco1  tai4kap6 waa6  zel**
```
because in the beginning **mention WAA SFP**
```
```
hat1jyu4 paak3to1 hou2  m4  hou2  nel  gam2
```
why not dating good Neg good **SFP SFP**
```
‘In the beginning, I mentioned, “how about dating? Is that okay?”’
```

(CRC)

**tai4ceot1 `suggest`**
```
maa4maal zik6zeng4  tai4ceot1  waa6  lei4fan1  lo1
```
Mother actually **suggest WAA divorce SFP**
```
‘Mother actually suggested a divorce.’
```

(HKCAC)
I asked him/her whether (s)he had re-arranged them after (s)he put them back.

But, (somebody) answered me, “Because…I didn’t…I didn’t…I haven’t…”

I answered that I had to be calm.

(S)he answered, “Yes!”

I lied to him, saying that the weather report had said that the temperature would drop to 10 degrees.

Cantonese Complementizer \textit{waa6}

\begin{quote}
\textit{ngo5} \textit{ak1} \textit{keoi5} \textit{waa6} \textit{aa3} Ann \textit{faan1} \textit{zou2} heong1 haa2 wu1 lou5 muk6 cai4 1sg lie 3sg WAA Ah Ann back Asp hometown Urumqi ‘I lied to him/her that Ah-Ann had gone back to her hometown Urunqi.’


\textit{sing4jing6 ‘admit’}
\begin{itemize}
\item \textit{keoi5 jat1zik1 dou1 mou5 can1 hau2 sing4jing6 gwo3 lo1, 3sg all-along too Neg in-person admit Asp SFP}
\item \textit{waa6 keoi5 go2 go3 leoi5 pang4 jau5 faan1} \textit{WAA} 3sg that girlfriend back ‘He never actually admitted that he really was going back to his ex-girlfriend.’
\end{itemize}

\textit{(CRC)}

\textit{giu3 zou6 ‘called as’}
\begin{itemize}
\item \textit{lei1 dou6 dou3 m4 hai6 giu3 zou6 waa6} \textit{gam3 hung2 bou3} here all Neg be call as WAA that scary ‘This place cannot be considered all that scary.’
\end{itemize}

\textit{(HKCAC)}

\textit{gaai2 sik1 ‘explain’}
\begin{itemize}
\item \textit{keoi5 gaai2 sik1 waa6 zii6 gei2 sam1 ceng4 m4 hou2} 3sg explain WAA oneself feeling Neg good ‘(S)he explained that (s)he was not in a good mood.’
\end{itemize}

\textit{(A-kuan. 1988. \textit{A JAN ri ji ‘Jan’s Diary’, v.3 Hong Kong: Yau Wo, pp.107)}

Diana \textit{gaai2 sik1} \textit{waa6,} ‘so2 ji3 jat6 bom2 zai2 zung3 di1 Diana explain WAA, so Japanese grow Poss ping4 gwo2 pou4 ai4 zii2 gam3 daai6 lap1 ze1.’ ‘apple grape so big Cl SFP’

\textit{Diana explained, “That’s why the apples and grapes grown by the Japanese are so big.”}


\end{quote}
‘Then, she grumbled about my father in front of me, saying that she’d already mentioned that she didn’t want it.’


‘At first, I wanted to refuse him/her saying that I would not go…’


‘My auntie groused to my mother that my second elder female cousin had given up a chance to study overseas for the sake of marriage.’


‘(She) always groused, “If my husband let me go to Japan to do that just once, I could also be that stunning.”’

Cantonese Complementizer waa6

tam3 ‘comfort’
ngo5 zung6 tam3 keoi5 waa6 lai4_gan2 lei1go3 lai5baai3
1sg even comfort 3sg WAA come Asp this week
wan2 maan3 tung4 keoi5 jat1cai4 heoi3 gaa1
find night with 3sg together go street
‘I even comforted him, saying that I would go out with him one night sometime this week.’


giu3 ‘shout’
ngo5 hoi1_zoek6 dang1 teng1_dou2 jit6 gaa1ze1 giu3 waa6,
1sg turn on_Asp light listen_Asp two sister shout WAA
“m4 hou2 hoi1 dak1 m4 dak1?”
‘Neg good turn on can Neg can?’
‘I turned on the lights and heard my second elder sister shout, “Could you not turn on the lights?”’


keoi5 gin3_dou2 ngo5 gin3_dou2 keoi5 zung6 daai6 seng1 giu3 waa6,
3sg see_Asp 1sg see_Asp 3sg even big voice shout WAA
“aa3fun1, lei5 mai5 soeng5 lai4 aa3…”
‘Ah-Fun 2sg Neg up come SFP…”
‘When (s)he saw me see him/her, (s)he actually shouted, “Ah-fun, don’t come forward…”’

Hong Kong: Crown Publishing, pp. 42)

laai6 ‘blame’
jin4hau6 zau6 laai6 di1 gei3ze2 waa6 hou2 hou3dou6 ne1
after then blame Cl reporter WAA very report SFP
‘And then you put the blame on the reporters.’

(CRC)
‘cognition and perception’ predicates:
gok3dak1 ‘think/ believe’
keoi5 seng4jat6 gok3dak1 waa6
3sg always think WAA
lei1go3 neoi5jai2 hou2 min6sin6 hou2 min6sin6
this girl very familiar very familiar
‘(S)he always thought that this girl looked very familiar.’

(HKCAC)
actually all think WAA


‘Actually, I have thought that, say, if we might go out some time.’

but unfortunately Neg car

‘I wanted to get a taxi downstairs, but unfortunately there were none about.’

Then, (somebody) wants to tell us that we should regulate our work.

‘I thought if you save up, why don’t we put the money at my place and I will keep it.’

‘I don’t think that I will be together with the boy who’s been dating me for two weeks.’
Cantonese Complementizer *waa6*

**tai2 ‘see’**

* ji4gaa1 lei5 tai2_haa3 waa6 si5min6 soeng6 gam3 siu1tiu4 now 2sg see_Asp WAA market up so downturn
tung4maai4 gam3 do1 jan4 zap1lap1 and so many people close business

‘Now you see that the economy is in a downturn and many businesses are closing down.’ (HKCAC)

**‘informative’ predicates**

**tung1zi1 ‘inform’**

* daan6hai6 go2go3 course le1 zau6 cat1jyut6 but that course SFP then July
sin1zi3 tung1zi1 waa6 sau1_zo2 ngo5 only inform WAA take_Asp 1sg

‘But the course only informed me in July that I was approved for admission.’ (HKCAC)

**se2 ‘write’**

* keoi5 se2_zyu6 waa6 hou2 zung1ji3 ngo5 lo1, 3sg write_Asp WAA very love 1sg SFP,
peng4si4 keoi5 mon5 gong2 gaa3 usually 3sg Neg say SFP

‘He wrote that he loved me very much. He didn’t say so.’ (HKCAC)

**Other predicates**

**hai6 ‘be’**

* hai6 jau5 po2 do1 ge3 hai6 order le1, ze1 maai5pun2 laa1haa2, be have quite many Poss be order SFP, that's buyer SFP,

‘There are quite a lot of orders, that’s buyers, who are buying at the market.’ (HKCAC)
‘What I most worried about is that actually in the problem we were discussing…’

‘My understanding is that in these years, they actually have been in touch.’

‘Even if Bokhary’s interpretation is that that should be the end…’

REFERENCES


Chappell, Hilary. to appear. Variation in the grammaticalization of verba dicendi in Taiwanese Southern Min and other Sinitic languages.


Cantonese Complementizer waa6


*Ka-Wai Yeung*
Department of Chinese
The University of Hong Kong
Pokfulam Road, HKSAR
kkwyueung@graduate.hku.hk
論廣東話之補語連詞：「話」

楊家慧
香港大學中文系

補語連詞(Complementizer)泛指連接子句補述語之詞(見英語中的that)，據類型語言學前人研究，表述說話之動詞於多種語言中均有語法化成補語連詞之現象。本文以橫向斷代方式分析廣東話「話」字，提出前述現象亦見於廣東話。「話」常見於各種表述言語或思想的謂語之後，一向被認為連動句式中的第二動詞，然而，本文以為「話」不一定用作動詞，於不同情境下，「話」之語意各異，其次語類制約亦有所不同。譬如：話1，意「說」[(PP) CP]/[(PP NP)；話2，意「責怪/指摘」[(NP CP)；而話3，則為補語連詞，用作連接子句補述語[(IP)。根據其表現之形式句法特徵，如：時體結構、論元選擇關係，本文論證了話3於作補語連詞時，形式句法上有別於說話動詞話1與及物動詞話2。