Relative clauses and nominal modifiers in Cantonese

Dominic Yu

2006 May 3

1 Introduction

Cantonese relative clauses are occasionally cited in the literature. In virtually every case (the exception being Matthews and Yip (2001), which deals specifically with types of Cantonese relative clauses), Cantonese is used as evidence supporting a particular analysis of Mandarin. Dikken and Singhapreecha (2004), footnote 22, is a typical example:

Cantonese ge is the direct counterpart of Mandarin de, having essentially the same distribution. A direct comparison . . . (all taken from Cheng 1997b) should make this clear.

What it fails to mention is that Cantonese has not one, but three constructions for forming relative clauses (examples taken from Matthews and Yip (2001, 280)):

(45) ngo5 sik1 ge3 hok6saang1
    lsg know PRT student
    ‘the student(s) I know’.

(46) ngo5 sik1 go2 di1 hok6saang1
    lsg know that CL.pl student
    ‘the students I know’.

(47) ngo5 sik1 go2 di1 ge3 hok6saang1
    lsg know that CL.pl PRT student
    ‘the students I know’.
In addition to the $ge3$ construction which looks similar to Mandarin $de$ (45), there is also a construction that uses demonstrative + classifier (46). Matthews and Yip note that (47) is a new “hybrid” construction that combines the other two, used in such formal contexts as classical music stations on the radio and job interviews.

One may be tempted to excuse this oversight by Dikken and Singhapreecha; they were only talking about Mandarin, after all. But an incomplete comparison may be worse than no comparison at all. In some cases, direct comparison leads to borderline ungrammatical examples. Here is one from Alsagoff and Ho, Chee Lick (1998) (copied verbatim from original article)²:

\begin{align*}
(5) & \quad \text{Nie wode mama de neige nanhaizi hen huaidan.} \quad \text{[Mandarin]} \\
& \quad \text{pinch my mother RP that child very naughty} \\
& \quad \text{‘That boy who pinched my mother is very naughty’} \\
(7) & \quad \text{Mit ngo mama ge go-go namzai ho kuai} \quad \text{[Cantonese]} \\
& \quad \text{pinch my mother RP that boy very naughty} \\
& \quad \text{‘That boy who pinched my mother is very naughty’}
\end{align*}

Matthews and Yip (2001, 273) note that this type of construction “rarely occurs [in Cantonese] and is perceived as clumsy at best”. Examples such as (5) and (7) are misleading in suggesting that all relative clauses across all Chinese dialects can be analyzed in exactly the same way, which can be dangerous if one is attempting to draw conclusions about cross-linguistic syntactic structures. I will attempt to shed some light on the issue by developing a formal analysis of relative clauses in Cantonese.

The structure of this paper is as follows: in section 2, we look at some interesting facts about Cantonese classifiers. In section 3, we set up the basis for our discussion by sketching an analysis for the noun phrase in Cantonese. In sections 4-6, we go on to examine several different analyses for relative clauses in Cantonese, and we will find that none of them are completely satisfying. It is my hope that laying out the facts as I have and considering the merits and faults of

---

¹This form appears to be more basic and is acquired by children earlier, according to Lee et al. (1995), cited in Matthews and Yip (1994, 277).

²These examples are troubling for two more reasons: one is the complete lack tones in the transcription, which unfortunately appears to be standard practice in Mandarin syntax; the second is the actual Cantonese transcription which is rife with errors, conflating short and long /a/ and the two vowels /o/ and /ou/.
different analyses as applied to Cantonese will contribute to the development of a more appropriate model in the future.

2 Various functions of the classifier in Cantonese

When listening for Cantonese classifier relative clauses, one is struck by the wide variety of functions that the classifier plays, all of them very similar to the relative clause:² ³ ⁴

(1) bun2 syu1
   CL  book
   ‘the book’

(2) keoi5 bun2 syu1
   3sg  CL  book
   ‘his/her book’

(3) keoi5 go2 bun2 syu1
   3sg  that  CL  book
   ‘that book of his/hers’

(4) sap6ji6-hou6 go2 coeng4 (hei3)
   twelve-date  that  CL  (show)
   ‘the performance on the twelfth’

³LSHK romanization is used here, which is approximately IPA except for the following: z [ts]; c [tsʰ]; ng [ŋ]; yu [y]; oe [œ]; eo [ø]; consonants are unaspirated/aspirated (not voiced/voiceless). Tones are as follows: 1. si 2. si ũ 3. si 4. si 5. si 6. si.
⁴Our classifier picture is also complicated by appositive noun phrases, for example:

(i) aa3-zon1  go2 go3 seoi1 zai2 . . .
   VOC-John that  CL  bad  son
   ‘that bastard John . . . ’

Notice that this is surface-similar to (3), and in fact one could construct a context where it could be interpreted as possessive, viz. ‘John’s stupid kid’ (also, the possessive reading would be more natural if the demonstrative were not there). We will not explore appositive nominals here; see Del Gobbo (1999) for a discussion of data from Mandarin.
In these examples, we see the use of a bare classifier used as a definite marker (1); possessive use of the classifier with and without a demonstrative (2, 3); a time adjunct (4); adjectival modification (5); and relative clauses (6, 7). Gil (2001) has noted that in Hokkien, a dialect of the Southern Min group of Sinitic, the marker e24 shows a similar range of features, including the use of a number before the classifier, and suggests that rather than separate numeral, possessive, adjectival, and relative clause e24s, there is only one e24 which serves an attributive function. The intuition seems to be that all of these nominal modifiers attach to the noun in the same way, and that a satisfactory analysis should somehow unify these different functions in a single type of structure. We will return to the question of whether there should be one single analysis for all these uses of the classifier below.

3 Noun phrases in Cantonese

Before we venture on to formal analyses of Cantonese relative clauses, it may be prudent to sketch out an analysis of Cantonese noun phrases as a first step.

---

5I believe that the number is part of the nominal projection here, for both Hokkien and Cantonese. In Cantonese, one cannot insert other things such as demonstratives in between the number and the classifier, e.g. saaml bun2 syu1 ‘three books’ > *saaml go2 bun2 syu1 (attempted: those three books, analogous to (3)), but numbers (and, indeed, quantifiers such as gei2 ‘a few’) can generally be freely inserted in between the demonstrative and classifier in all the examples above (but not di1, which is inherently plural and cannot have a number preceding it). I will not speak further of Hokkien here, except to point out that Gil (2001, 110) himself notes that numerals are different: of all the pre-e24 elements, tone sandhi only occurs in the case of numerals.
3.1 Order of elements in the noun phrase

Matthews and Yip (1994, 88) lists the order of elements in the noun phrase as follows:

(8) demonstrative – numeral – classifier – adjective – (ge) – noun

There is more structure to the noun phrase than this list implies, of course. First, 
\textit{ge}3 cannot occur without some sort of phrase to its left. (8) indicates that this
position should be filled by an adjective, but in fact it can be an adjective, noun
phrase (indicating possession), or relative clause (Matthews and Yip: 1994,
109).\(^6\) Furthermore, there is reason to believe that adjectives can appear in a
syntactic position next to the noun (as opposed to getting there via a pre-syntactic
word formation process), in addition to modification via \textit{ge}3—Paul (2005) gives
a parallel discussion for Mandarin and \textit{de}.\(^7\) Thus, a fuller structure for Cantonese
might look like this:

(9) \[
\text{DP} \\
\text{D} \\
\text{NumP} \\
\text{Num} \\
\text{ClP} \\
\text{Cl} \\
\text{XP} \\
\text{phrase \textit{ge}3 Adj. NP} \\
\]

Exactly what X might be (e.g., it might be N, if we believe everything is adjoined
 to N) will be explored below. Meanwhile, we must also keep in mind all the
modifier phrases that could potentially go in front of the noun phrase, which we
listed in section 2. That is, there are also things that can attach to the top of the
tree in (9).

\(^6\)To be fair, the co-occurrence of demonstrative + classifier with a possessive \textit{ge}3 does seem odd. I am as of yet unable to construct a context where it would sound natural.

\(^7\)When applied to Cantonese, Paul (2005)’s analysis leads to a very interesting result: there are actually four (!) places where adjectives can show up: (1) next to the noun via morphology, (2) next to the noun via syntax, (3) attached to the noun via \textit{ge}3, and (4) attached to the noun phrase next to the classifier.
In contrast to Cantonese, Mandarin seems to have two possible orders for elements in noun phrases, which Li and Thompson (1981, 124) list as follows:

(10) a. associative phrase + classifier/measure phrase + relative clause + adjective + noun
    b. associative phrase + relative clause + classifier/measure phrase + adjective + noun

Here, “associative phrase” means noun phrase + de (indicating possessive and possessive-like meanings); “classifier/measure phrase” means demonstrative + numeral + classifier; and “relative clause” includes adjective + de as well as relative clause + de. The two orderings are exemplified below (examples taken from Li and Thompson (1981, 124)).

(108) d. wǒ de nèi ge zhù zài Měiguó de hǎo péngyou
     I GEN that CL live at America NOM good friend
     that good friend of mine who lives in the United States
f. wǒ de zhù zài Měiguó de nèi ge hǎo péngyou
     I GEN live at America NOM that CL good friend
     that good friend of mine who lives in the United States

One question that immediately comes to mind is this: why does Mandarin have two possible orders, but Cantonese only one? In order to find the answer, let us compare the two orders: notice that in both, the associative (noun phrase + de) phrase comes first; the difference lies in the position of the classifier/measure phrase. This is an important clue to the structure, since it is the classifier/measure phrase that contains the demonstrative, which is usually taken to go in the D slot of the DP. In other words, if we assume that the demonstrative is the head of the DP, we find that a full DP can appear to the right of de in Mandarin, whereas Cantonese ge3 only allows an NP.

Note that this ability of de to have DPs attached to its right is in addition to its ability to have NPs. For example, the structure of (108d) must be as follows:

---

8I’ve added hǎo ‘good’ to (108f) to make the examples completely parallel.
9Perhaps this is related to why a possessive in the middle of the phrase is odd, as I noted in footnote 6.
10This is not entirely accurate: ge3 followed by demonstrative + classifier can occur, but my impression is that this occurs in only in formal/academic/political registers, where influence from Mandarin is heavy.
Thus, the structure of the Mandarin noun phrase is be exactly the same as the one for Cantonese, with the substitution of de for ge3 in (9). The crucial difference lies outside, in how modifiers of the DP attach. Mandarin requires de; Cantonese attaches modifiers with no intervening particle.

Thus far, we have shown that ge3 is limited to NP modification; another piece of evidence shows that the NP is actually obligatory. Matthews and Yip (1994, 111) note that the Cantonese classifier relative “enables relative clauses to be constructed which would be awkward or impossible using ge, such as those with a demonstrative phrase as the head”:

(11) gaau3 nei5 taan4 kam4 go2 go3?
    teach 2sg play piano that CL
    ‘The one who teaches you piano?’

(12) * gaau3 nei5 taan4 kam4 ge?
    teach 2sg play piano GE
    Attempted: ‘The one who teaches you piano?’

In (11), a relative clause modifies the “headless” noun phrase go2 go3 ‘that one’. In (12), I have substituted the alternative relativizing element, ge3, for demonstrative + classifier. The fact that this sentence is not grammatical shows that these two constructions, while seemingly similar, cannot be analyzed in the same way: go2 go3 is a noun phrase that can be modified, but ge3 is a linker that requires an NP on its right side.

11 Compare with a Mandarin example from Li and Thompson (1981, 575):

(2) [DP zhòng shúguǒ de ] hěn nán guòhuó
grow fruit NOM very difficult make:living
It is difficult for fruit growers to make a living.
3.2 Definiteness in the noun phrase

Another issue that must be discussed is that of definiteness. In this, the classifier plays a crucial role. Cheng and Sybesma (1999) give an extensive analysis of Cantonese and Mandarin; we will summarize the Cantonese facts here, along with examples:

<table>
<thead>
<tr>
<th>Structure</th>
<th>Definiteness</th>
</tr>
</thead>
<tbody>
<tr>
<td>(bare) noun</td>
<td>generic</td>
</tr>
<tr>
<td>CL + noun</td>
<td>preverbally: definite; postverbally: definite OR indefinite</td>
</tr>
<tr>
<td>Num + CL + noun</td>
<td>indefinite</td>
</tr>
<tr>
<td>dem + Num + CL + noun</td>
<td>definite</td>
</tr>
</tbody>
</table>

(13) ngo5 zung1 ji3 maau1

1sg  like cat

‘I like cats.’

(14) zek3 maau1 hou2 zing6

CL  cat very quiet

‘The cat is very quiet.’

(15) ngo5 maa5-zo2 zek3 maau1

1sg buy-PERF CL  cat

‘I bought a cat.’ OR ‘I bought the cat (which we’ve been talking about).’

(16) ngo5 maa5-zo  jat1 zek3 maau1

1sg buy-PERF one CL  cat

‘I bought a cat.’

(17) ngo5 maa5-zo2 go2 zek3 maau1

1sg buy-PERF that CL  cat

‘I bought that cat.’

Again, the key here is the classifier. In a way, it acts like an English determiner ([the, and a if we consider a a determiner). There are different ways to analyze this. Del Gobbo (1999) analyzes this in terms of feature movement; on the other hand, Cheng and Sybesma (1999) actually consider the classifier to be the locus of definiteness in the noun phrase, and not D.12 The specific analysis will not

---

12Cheng and Sybesma (1999, 527) mysteriously state, “For Tang [1990] (but not for us), demonstratives are Ds.” Unfortunately, I could not figure out what they thought demonstratives should be.
concern us here; what is important is that it is the classifier that gives the noun phrase definiteness. This importance of this issue will become apparent when we discuss the antisymmetry analysis of Cantonese relative clauses below.

4 A Traditional Analysis

A traditional analysis of Cantonese relative clauses can be made in an analogous fashion to Mandarin. For Mandarin, \textit{de} is taken to be a complementizer that takes a sentential (or in Minimalism terms, TP) complement to its left (such a view is summarized in Wu (2000, 96 ff.)). This is just like the traditional analysis for English \textit{that}: an empty operator moves to SpecCP, and the entire CP is adjoined to the noun phrase. There are differences from English: first, the complementizer \textit{de} follows the TP; second, the CP is adjoined to the left of the noun phrase, reflecting the prenominal position of the relative clause. The following tree is adapted from Wu (2000, 98):

\begin{equation}
\begin{tikzpicture}
  \node (NP) {NP}
    child {node (CP) {CP}
        child {node (C) {C'}
            child {node (op_i) {op}}
        }
    }
    child {node (NP) {NP}
        child {node (shu) {shù}}
        child {node (C) {'book'}}
    }
    child {node (TP) {TP}
        child {node (C) {C}}
        child {node (de) {de}}
    }
    child {node (wou_zuo_tian_mai_t) {wǒ zuótiān mǎi t_i}}
  child {node (I_yesterday_bought_t) {I yesterday bought t_i}}
\end{tikzpicture}
\end{equation}

In this analysis, the left-branching complementizer, while certainly odd in an otherwise right-branching, SVO language, is thought to be less odd for two reasons. First, Mandarin verbs for ‘say’, ‘believe’, etc. take sentential complements with no overt complementizer; perhaps there is some empty C at

\footnote{Interestingly, Cheng and Sybesma (1999) point out that adding a number “undoes” the definiteness, and that subsequently adding a demonstrative “restores” it.}
the right side of the TP that we can’t see. Moreover, there exist sentence-final particles, including the question particle ma, that seem to support the idea of a left-branching C in Mandarin.

Similarly, Cantonese has no overt complementizers after its verbs for ‘say’ and ‘believe’. In addition, it has a rich system of sentence-final particles to indicate speech-act types, evidentiality, and affective/emotional coloring (Matthews and Yip: 1994, 328). Thus, this seems like a not-unreasonable analysis. However, it is not without its flaws.

Wu (2000) gives three arguments against this analysis for Mandarin; these arguments apply equally well to Cantonese. First, other languages which do have complementizers with rightward complements (such as English) have sentence-final question particles (e.g. eh?, right?), suggesting that these particles might not be best analyzed as C in the first place; second, these particles can only occur in matrix clauses, not embedded ones; and third, words such as ‘if’ (Cantonese jyu4gwo2, Mandarin rúguó), usually analyzed as C, show up before the TP (on the left). To accommodate this, we would need two types of C in Chinese, each with different directionality.

Furthermore, Mandarin de (and possibly Cantonese ge3) can modify a full DP, as pointed out in section 3.1. This is in contrast to English relative clauses, where the CP is taken to adjoin to the NP.

Finally, it is unclear how this analysis would be applied to classifier relatives in Cantonese, where there are two problems: first, the relative clause modifies a DP; and second, there appears to be no overt complementizer outside of that DP. Thus in the tree below, it is unclear what the internal structure of the CP would be—which side of the TP is the null C on?

---

14 “Double conjunctions” such as Cantonese jyu4gwo2 ... ge3 waa2 (waa2 meaning ‘say/speech’) and the analogous Mandarin rúguó ... de huà are an interesting complication which I will not discuss here.
5 An Antisymmetry Analysis

With these problems in mind, we turn to an alternate analysis, first formulated by Kayne (1994) and applied to Mandarin by Simpson (2002) and Wu (2000). Under this analysis, the CP is the direct complement of the determiner, and the NP being relativized is moved to SpecCP. Here is an English example:

Following the D-CP hypothesis for Cantonese, we would come up with the following two structures corresponding to classifier and ge3 relatives, respectively:
Here, *ge3* is not a complementizer, but a determiner (Wu (2000) and Simpson (2002) give detailed explanations of why this must be the case for this structure under this analysis). Also note that in both structures, the entire TP must be moved to SpecDP to achieve the correct linear order. Moreover, in (21), what is moved is not the NP, but the CIP (or even the NumP, since that is what lies between DP and CIP in the nominal projection), a consequence of assuming the
nominal projection that we have been using, along with a strict D-CP structure.

This analysis does seem to solve some of the problems brought up for the traditional analysis. Specifically, all of the C’s (which are conveniently null) are right-branching, addressing all of Wu (2000)’s concerns.

However, the problems are numerous. First, the fact that *de* and *ge3* are placed in the D slot seems to be incompatible with the possibility of demonstrative + numeral + classifier appearing after them. Simpson (2002) gets around this problem by suggesting that demonstratives are generated somewhere lower than D (Q0 or SpecQP). A related problem is the fact that demonstrative + numeral + classifier sequence can appear above the structure in (22), as shown in (9) and (108d). Here, Simpson (footnote 7) suggests that these elements may be generated below D and optionally raised above it, not to SpecDP since it is already filled by the TP, but some functional/deictic projection above DP; alternatively, they are base generated in that functional projection. The movement account would seem to be an incorrect account of Cantonese, where the demonstrative rarely occurs after *ge3*. Moreover, Tang (to appear, 24) points out that in Mandarin, certain modifiers cannot follow the dem + num + CL sequence, suggesting that the two structures are semantically different, not merely the result of an optional movement. The same is true in Cantonese:

(23)  kei4taa1 go2 loeng2 go3 hok6saang1
      other that two  CL  student
      ‘those other two students’

(24)  * go2 loeng2 go3 kei4taa1 ge3 hok6saang1
      that two  CL other  GE  student

Another problem that arises is the question of what motivates movement of the TP to SpecDP. Simpson (2002, 12) gives an explanation that may seem reasonable at first (here, “Chinese” means “Mandarin”):

Such an analysis also has the advantage that it accounts for the unexpected pre-nominal positioning of relative clauses in Chinese in terms of the special properties of a single lexical item, i.e. the enclitic requirements of the D0 element *de*, rather than attributing it to a general parameter of the language. Dryer (1992) observes that Chinese is essentially unique in being the only language attested anywhere with a basic V-O word order and also pre-nominal relative
clauses, RC-N structures elsewhere being found to occur only in very regular O-V, (descriptively) head-final languages. This exceptional fact about Chinese clearly must have some explanation, and it would seem most appropriate to assume that it relates to the exceptional properties of a single lexical element rather than being the outcome of the general typological nature of Chinese, whose overall V-O head-initial character would clearly lead one to expect only post-nominal relative clauses to occur.

This explanation, however, must be false, since Cantonese has pre-nominal relative clauses with no exceptional lexical item to “do” the typologically odd word order. (The claim that Mandarin is unique in this regard is also false, since besides Cantonese, Hokkien and probably most other Chinese languages also have V-O word order and pre-nominal relative clauses.) In classifier relatives, the relative clause is attached directly to the noun phrase with no linking particle, unless you consider the demonstrative + classifier sequence to be a linker of this type. However, the idea that the word order is caused by a special version for the lexical item *go* \textsuperscript{2} ‘that’ which c-selects for a DP complement and causes movement of the TP to its specifier seems highly implausible to me. For one thing, there is no other evidence that would lead us to believe there are two different kinds of *go*, one relativizing and the other not. Moreover, under this analysis, it is the CIP (or NumP) that is moved to SpecCP, not NP as in the English example. This leads us to a problem having to do with the classifier, which we turn to next.

As we pointed out in section (3.2), the locus of definiteness in the noun phrase seems to lie in the classifier, and various methods of accounting for this have been proposed, such as feature movement. This is important in light of the fact that in classifier relatives, the demonstrative is sometimes optional, especially with the plural classifier *di* \textsuperscript{1} (this is example (37) in Matthews and Yip (1994)):

(25) keoi5 se2 di1 je5 jau5 mou5 jung6 gaa3
    3sg write CL.pl stuff have not:have use PRT
    ‘Is the stuff he writes any use?’

If there is movement or feature-checking going on between the classifier and the D slot here, how would it happen under this analysis? It seems that what is a local A-movement in a simple noun phrase (without modifiers) would be analyzed here as A-movement, crossing a CP boundary. This may be clearer if
one refers to the tree in (21): if the $D_{go2}$ is not there, the Cl, or features from Cl, must Agree or Move to D. Perhaps this is allowed because the CIP is sitting in SpecCP, from which raising is allowed, but it does seem unusual that it is not the entire node that sits in SpecCP that moves, but an element inside it.

6 A Unified Approach to Nominal Modifiers?

Thus far, we have looked at a left-branching complementizer approach, which is suboptimal because of its unusual directionality; and an antisymmetry approach, which is strange in a number of ways, including positing movement which seemingly has no motivation. In addition, if we assume that adjectival and possessive modifiers are actually relative clauses with some kind of null predicate/possessive verb, both of these analyses lead to very complex, elaborate structures for what intuitively are very simple constructions. Is there a satisfactory solution out there?

Unfortunately, I have not yet come upon one. Dikken and Singhapreecha (2004) propose that Mandarin $de$ is a semantically empty linker that causes Predicate Inversion of a small clause embedded in a DP, similar to French $de$ and Thai $thi$. This account is able to give a uniform structure to possessive, adjectival, and relative clause modifiers; however, it is difficult to apply this to Cantonese classifier relatives, where there is no meaningless linker particle to “do” the predicate inversion.

7 Conclusion

The “correct” analysis would hopefully capture the facts and intuitions that I have laid out above. Clearly, modifiers (possessive/attributive, adjectival, relative clauses) can modify things without an intervening linker particle, as the case of Cantonese shows. Such modification is usually understood to be formed by adjunction, which is what the traditional analysis goes after. The mechanism by which the modifier and modify-ee are linked together is a question which we unfortunately have not been able to answer here. The antisymmetry approach tries to hard-code this link via movement, but in a way it tries too hard without actually furthering our understanding of the phenomena involved.
The crux of the problem lies in the fact that Cantonese has two ways of linking modifying phrases (noun, adjectival, and clausal) to the noun phrase. One, the particle ige3, can be analyzed in a very similar (but not identical) way as Mandarin de. The other seems to employ no overt linker outside of the classifier, which is usually obligatory in noun phrases anyway. It seems that some sort of adjunction analysis is the most promising route; I await the day that we have a fuller understanding of adjunction.

References


