Divergence of Expressing Definiteness between Mandarin and Cantonese

Joanna Ut Seong Sio  
Nanyang Technological University  
Singapore  
ussio@ntu.edu.sg

Sanghoun Song  
Nanyang Technological University  
Singapore  
sanghoun@ntu.edu.sg

1 Background

In this paper, we model the dialectal variation in definiteness in Mandarin (cmn) and Cantonese (yue) using the framework of HPSG (Pollard and Sag, 1994) and MRS (Copestake et al., 2005). There are 4 basic types of NPs in Chinese, as exemplified in Table 1.¹

<table>
<thead>
<tr>
<th>type</th>
<th>example</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEM-CL-N</td>
<td>這隻狗</td>
<td>‘this dog’</td>
</tr>
<tr>
<td>NUME-CL-N</td>
<td>三隻狗</td>
<td>‘three dogs’</td>
</tr>
<tr>
<td>CL-N</td>
<td>隻狗</td>
<td>‘a/the dog’</td>
</tr>
<tr>
<td>N</td>
<td>狗</td>
<td>‘a/the dog’ or ‘dogs’</td>
</tr>
</tbody>
</table>

As shown in the table, the interpretation of [CL-N] phrases and [N] phrases vary. They can be interpreted as definite (‘the X’), indefinite (‘a/an X’), or both. The whole range of interpretations are not available to all dialects, as we explain in more detail below.

2 Basic Properties

Unlike English, there are no articles (e.g. a, the) in Chinese indicating the definiteness value of an NP. The referential interpretations of some Chinese NPs are relatively flexible. Some surface forms can have two referential interpretations. In addition, dialects vary in terms of which surface forms are ambiguous.

[N] phrases can always be interpreted as having a kind reading across dialects, similar to a bare plural in English:

(1) a. 狗 喜歡 骨頭  
gǒu xǐhuān gǔtou  
dog like bones  
‘Dogs like bones.’ [cmn]

In Mandarin, bare nouns are ambiguous in terms of definiteness, as in (2a); In Cantonese, [CL-N] phrases are ambiguous, as in (2b) (Cheng and Sybesma, 1999; Sio, 2006).

(2) a. 我 看見 狗  
wǒ kāijiàn gǒu  
1SG see dog  
‘I saw a/the dog.’ [cmn]

b. 我 見到 隻狗  
ngo5 gin3dou2 zek3 gau2  
1SG see CL dog  
‘I saw a/the dog.’ [yue]

Phrases with demonstratives are always definite; [NUME-CL-N] phrases are always indefinite. A summary of definiteness interpretations of Mandarin and Cantonese NPs are presented in Table 2.

<table>
<thead>
<tr>
<th>type</th>
<th>Mandarin</th>
<th>Cantonese</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEM-CL-N</td>
<td>definite</td>
<td></td>
</tr>
<tr>
<td>NUME-CL-N</td>
<td>indefinite</td>
<td></td>
</tr>
<tr>
<td>CL-N</td>
<td>indefinite (in)definite</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>(in)definite indefinite</td>
<td></td>
</tr>
</tbody>
</table>

Generally, only definite noun phrases can appear in the subject position in Chinese (Chao, 1968; Li and Thompson, 1981; Lee, 1986, among many others).

Even though a [CL-N] phrase in Cantonese can be interpreted as either definite or indefinite, a [CL-N] phrase in the subject position can only be interpreted as definite. This is illustrated in (3a). The same applies to Mandarin bare nouns, which are only interpreted as definite (or kind) in the subject position as exemplified in (3b):

(3) a. 我 看見 隻狗  
wǒ kāijiàn zek3 gau2  
1SG see CL dog  
‘I saw a/the dog.’ [cmn]

b. 狗 見到 骨頭  
gǒu gin3dou2 gǔtou  
dog CL see bone  
‘Dogs see bones.’ [yue]

¹All examples used in this paper are written in traditional Chinese for ease of comparison between Mandarin and Cantonese.
3 Analysis

The previous section can be summarized as follows. First, there are four basic types of NPs in Mandarin and Cantonese, viz. [DEM-CL-N], [NUME-CL-N], [CL-N], and [N]; Second, [DEM-CL-N] phrases are always definite, and [NUME-CL-N] phrases are always indefinite; the last two types show a contrast in definiteness between Mandarin and Cantonese. Third, there exists a constraint on what can appear in the subject position: definite NPs only with one exception. Building upon these, this section models the properties of the four types of NPs in Mandarin and Cantonese within the framework of HPSG (Pollard and Sag, 1994) and MRS (Copestake et al., 2005).

3.1 Cognitive Status

Quite a few previous studies have dealt with definiteness and/or givenness using HPSG so far. The analysis proposed here is along the line of Borthen and Haugereid (2005) and Bender and Goss-Grubbs (2008). These studies address a property of referents within the HPSG formalism and propose cog-st (cognitive status), which specifies the relationship between referents and the common ground in discourse. This feature structure places a constraint on the availability of types of NPs in particular constructions.

The constraint has much to do with the morphosyntactic markers of expressing definiteness. Borthen and Haugereid (2005) and Bender and Goss-Grubbs (2008) argue that the binary distinction such as definite vs. indefinite is sometimes not precise enough to deal with the various types of definiteness in NPs. As exemplified in the previous section (and in many other human languages), NPs are often ambiguous, though a more specific meaning is provided up to the entire parse tree. Furthermore, language processing, as of now, normally does not go beyond a sentence (i.e. intrasentential). Contextual information can only be partially resolved in our language application. In other words, not all NP structures can be analyzed as two-fold (i.e., definite vs. indefinite) within the context of grammar engineering. Instead of the binary distinction, Borthen and Haugereid (2005) and Bender and Goss-Grubbs (2008) use the givenness hierarchy (Prince, 1981; Gundel et al., 1993). From right to left in Table 3, each type is exemplified in (8).

<table>
<thead>
<tr>
<th>Table 3: Givenness hierarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>In focus</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>it</td>
</tr>
<tr>
<td>this N</td>
</tr>
</tbody>
</table>

(8) a. I couldn’t sleep last night.
    b. 

For a noun phrase that cannot be interpreted as definite, putting it in the subject position would only lead to ungrammaticality:

(3) a. * jitterbug want cross road
    jitterbug want cross road
    ‘The dog wants to cross the road.’
    NOT ‘A dog wants to cross the road.’ [yue]

(4) a. * 土狗 妻 通 馬路
    jīōu yōu yāo guò mālú
    dog want cross road
    ‘The dog wants to cross the road.’
    NOT ‘A dog wants to cross the road.’ [cmn]

[NUME-CL-N] phrases are always indefinite. They can’t appear in the subject position (example (5), (6) and (7) are taken from (Li, 1998)):

(5) * 三個學生在校受傷了
    sān gè xuéshēng zài xuéxiào shòushāng le
    three students hurt at school SFP
    ‘Three students were hurt at school.’ [cmn]

The existential marker you ‘have, exist’ has to be added before the phrase to make it grammatical when appearing in the subject position:

(6) 有三個學生在校受傷了
    yǒu sān gè xuéshēng zài xuéxiào shòushāng le
    have three students hurt at school SFP
    ‘There are three students hurt at school.’ [cmn]

There is an exception to this restriction. When a [NUME-CL-N] phrase only denotes quantity, it could appear in the subject position (Li, 1998):

(7) 三個保姆就照顧
    sān gè bǎomǔ jiù zhāogù
    three babysitter only care
    ‘Three babysitters took care of you, only one child?’
    [cmn]
Along this line, Borthen and Haugereid (2005) provide an HPSG-based type hierarchy of cognitive status, which was then slightly refined by Bender and Goss-Grubbs (2008) as sketched out in (9).

This hierarchical approach to NP meanings enables us to represent partial information and thereby facilitates maintaining the phrase structure rules of forming NPs in a flexible way.

Building upon the type hierarchy provided in (9), Table 2 is now converted into Table 4.

<table>
<thead>
<tr>
<th>type</th>
<th>Mandarin</th>
<th>Cantonese</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEM-CL-N</td>
<td>uniq+fam+act</td>
<td></td>
</tr>
<tr>
<td>NUME-CL-N</td>
<td>type-id</td>
<td></td>
</tr>
<tr>
<td>CL-N</td>
<td>type-id</td>
<td>activ-or-less</td>
</tr>
<tr>
<td>N</td>
<td>activ-or-less</td>
<td>type-id</td>
</tr>
</tbody>
</table>

First, if a particular construction conveys only definite meaning, the phrase places the *uniq+fam+act* feature to the head noun as indicated in the second row in Table 4. Notice that in the *cog-st* hierarchy provided in (9) *uniq+fam+act* excludes the leftmost item and the rightmost item from its subtypes. The leftmost item *type-id* signals indefiniteness, and the rightmost item *in-focus* is used for pronouns. In this way, *uniq+fam+act* indicates that the NP can be evaluated as containing definiteness. Note also that ‘Activated’ and ‘Familiar’ in Table 3 are instantiated as NPs with demonstratives (i.e., *this N*, and *that N*). Since *uniq+fam+act* includes these meanings, [DEM-CL-N] in the second row of Table 4 is not inconsistent with the constraint. Second, if a particular construction conveys only indefinite meaning, the phrase is constrained as type-id. Notice that the *type-id* node in the *cog-st* hierarchy is exclusive of any definite meaning. Finally, if a particular construction is ambiguous (i.e. (in)definite), the cognitive status of the phrase is specified as *activ-or-less*, which excludes only *in-focus* from the subtypes. The other types in the bottom line, such as *type-id, uniq-id, familiar*, and *activated*, inherit from *activ-or-less*. This means that an NP whose value of cognitive status is *activ-or-less* can be interpreted as either indefinite or definite.

### 3.2 Phrase Structure Rules

In Table 4, note that Mandarin and Cantonese exhibit contrasting features in the fourth row and the fifth row whereas they share the same features in the second row and the third row. The constraints on such a divergence of expressing definiteness between Mandarin and Chinese are as follows.

First of all, Mandarin and Cantonese share the following lexical type of classifiers, in which the element of MOD goes for the head noun, the element of SPR (i.e. specifier) goes for demonstratives and numerals. For example, in 那隻狗 ‘this dog’, 那 and 狗 are constrained as SPR and MOD, respectively.

Classifiers signal [COG-ST *activ-or-less*] to the head noun, given that pronouns and proper names are normally associated with *in-focus* and seldom co-occur with classifiers. Recall that *in-focus* does not inherit from *activ-or-less*, as sketched out in (9). When classifiers are not specified by demonstratives and numerals (i.e, [CL-N]) in Mandarin, the NP involves an indefinite interpretation. This is constrained by a lexical rule, as presented in the AVM of (11). This rule makes the SPR list empty and places a constraint on the head noun’s cognitive status as *type-id* responsible for indefinite. A
sample derivation is given on the right side.

(11) \[
\begin{array}{l}
\text{no-spr-cl-lex-rule} \\
\text{MOD} \left[ \text{COG-ST type-id} \right] \\
\text{SPR} \langle \rangle \\
\text{ARGS} \left[ \text{classifier} \right] \\
\end{array}
\]

Note that this constraint is Mandarin-specific. Since the definiteness of the [CL-N] form in Cantonese is ambiguous, this rule is not necessary for Cantonese.

Mandarin and Chinese also differ in how bare NPs are constrained. Cantonese, in which the [N] form is not ambiguous, employs the following lexical rule for nouns. This rule functions the same as the rule presented in (11), but it takes nouns as its daughter. The rule is Cantonese-specific.

(12) \[
\begin{array}{l}
\text{no-cl-lex-rule} \\
\text{MOD} \left[ \text{COG-ST type-id} \right] \\
\text{SPR} \langle \text{noun} \rangle \\
\text{ARGS} \left[ \text{classifier} \right] \\
\end{array}
\]

Bare-np-phrase used in the parse trees of (11-12) is constrained as represented in the following AVM. This non-branching rule signals activ-or-less and introduces an existential quantifier (i.e. exist_q_rel) into the RELS list.

(13) \[
\begin{array}{l}
\text{bare-np-phrase} \\
\text{HD} \left[ \text{noun} \text{COG-ST} \text{LTOP} \text{INDEX} \right] \\
\text{C-CONT} \left[ \text{PRED} \text{ARG0} \text{RSTR} \right] \\
\text{RELS} \langle \text{exist_q_rel} \rangle \\
\text{HCONS} \langle \text{gqg} \text{HARG} \text{LARG} \rangle \\
\end{array}
\]

If the daughter of this phrase can have a more specific value of COG-ST, the value is unified. For instance, the daughters of bare-np-phrase in parse trees of (11-12) are constrained as [COG-ST type-id]. Because type-id is a subtype of activ-or-less, the COG-ST feature is unified as type-id (i.e. indefinite).

Finally, in order to disallow indefinite items to be used as subjects in Mandarin and Cantonese, the ordinary subj-head-phrase rule additionally includes one language-specific constraint as provided in (14).²

(14) \[
\begin{array}{l}
\text{subj-head-phrase} \\
\text{NHD} \text{COG-ST uniq-or-more} \\
\end{array}
\]

Note that uniq-or-more is mutually exclusive with type-id, as represented in the type hierarchy (9). For instance, the structures provided in (11-12) cannot take the subject position because their COG-ST feature is inconsistent with the constraint on subj-head-phrase.

4 Sample Derivations

This section provides two sample derivations in Cantonese and Mandarin, respectively. The sentences are listed in (15). The two sentences share almost the same meaning. The subjects are evaluated as conveying a definite interpretation, as only definite NPs can appear as subjects in Chinese.

(15) a. 隻狗走啊
zek3 gau2 zum2 laa3 CL dog run SFP
‘The dog ran.’ [yue]

b. 狗走啊
gau zˇou laa3 CL dog run SFP
‘The dog ran.’ [cmn]

Figure 1 representing (15a) shows the derivation of a Cantonese sentence, an intransitive verb taking a [CL-N] phrase as the subject. Even though [CL-N] phrases can be interpreted either as definite or indefinite in Cantonese, when appearing in the subject position, it can only be interpreted as definite. The Mandarin counterpart of this sentence would be ungrammatical as [CL-N] phrases can only be indefinite in Mandarin. In the MRS structure on the right side, the COG-ST value of the subject 隻‘dog’ is specified as uniq+fam+act. Note that the NP 隻‘CL-dog’ itself is assigned activ-or-less as the value of COG-ST, as shown on the tree. The value becomes more hierarchically specific when the NP is used as the nonhead daughter of subj-head-phrase: When the NP

²Since pronouns, proper names, and clausal subjects are not indefinite, this constraint does not affect other types of subjects.
is combined with the verb 走 ‘run’ to form a subj-head-phrase, the subject is assigned [COG-ST uniq-or-more], and this results in [COG-ST uniq+fam+act]. Note that uniq+fam+act multiply inherits from activ-or-less and uniq-or-more. As a result, the most left hand node type-id and the most right hand in-focus in the type hierarchy presented in (9) are excluded from a cognitive status of the NP. That is to say, the NP can be interpreted as (near) definite.

Figure 2 representing (15b) shows the derivation of a Mandarin sentence, an intransitive verb taking an [N] phrase as subject. Even though [N] phrases can be interpreted either as definite or indefinite in Mandarin, when appearing in the subject position, it can only be interpreted as definite. The Cantonese counterpart of this sentence would be ungrammatical as [N] phrases can only be interpreted as indefinite in Cantonese. The COG-ST of the subject 狗 ‘dog’ in the MRS representation is specified as uniq+fam+act in the same way as Figure 1. The subject is constrained as [COG-st activ-or-less] by bare-np-phrase and also [COG-st uniq-or-more] by subj-head-phrase. These two constraints are unified into [COG-ST uniq+fam+act].3

Due to space limitation, the derivations for the quantity reading for [NUME-CL-N] phrases and the generic reading for [N] phrases would only be discussed in the talk.

References


3Note that cog-st is hear-oriented. The speaker-oriented status is represented as [SPECI bool] (i.e. specificity) (Borthen and Haugereid, 2005; Bender and Goss-Grubbs, 2008).


