Divergence in expressing definiteness between Mandarin and Cantonese

Joanna Ut-Seong Sio
Nanyang Technological University

Sanghoun Song
Incheon National University

Proceedings of the 22nd International Conference on
Head-Driven Phrase Structure Grammar

Nanyang Technological University (NTU), Singapore

Stefan Müller (Editor)

2015

Stanford, CA: CSLI Publications

pages 177–194

Sio, Joanna Ut-Seong & Sanghoun Song. 2015. Divergence in expressing definiteness between Mandarin and Cantonese. In Stefan Müller (ed.), *Proceedings of the 22nd International Conference on Head-Driven Phrase Structure Grammar, Nanyang Technological University (NTU), Singapore*, 177–194. Stanford, CA: CSLI Publications. DOI: 10.21248/hpsg.2015.11.

Abstract

In this paper, we model the dialectal variation in the expression of definiteness in Mandarin and Cantonese adopting the Head-Driven Phrase Structure Grammar (HPSG) framework (Pollard & Sag, 1994) and Minimal Recursion Semantics (MRS) (Copestake et al., 2005).

1 Introduction

Definiteness is a grammatical category that applies to noun phrases. A noun phrase is definite if there is sufficient information in the context for the hearer to identify the referent. Identifiability is a pragmatic notion relating to the assumptions made by the speaker on the cognitive status of a referent in the mind of the addressee in the context of utterance (Chen, 2004).

Unlike English, there are no articles (e.g. *a*, *the*) in Chinese indicating the definiteness value of a noun phrase. The referential interpretations of some Chinese noun phrases are flexible and thus ambiguous given appropriate contexts. In addition, dialects vary in terms of which surface forms are ambiguous. Amongst seven Chinese dialectal groups (viz., Northern, Wu, Xiang, Gan, Hakka, Yue and Min (Yuan, 1983), the present work focuses on Mandarin (abbreviated as 'cmn' in examples), which is a member of the Northern Group, and Cantonese (abbreviated as 'yue' in examples), which is a member of Yue.

2 Basic Data

2.1 Four Basic Types of Noun Phrases in Chinese

Table 1: Definiteness

type	example	Mandarin	Cantonese
DEM-CL-N	這隻狗	definite	
NUME-CL-N	三隻狗	indefinite	
CL-N	隻狗	indefinite	(in)definite
N	狗	(in)definite	indefinite

Noun phrases (NPs) in Chinese come in four basic forms: [DEM-CL-N], [NUME-CL-N], [CL-N] and [N]. [DEM-CL-N] phrases are always definite in Chinese while

[†]We are thankful to Francis Bond for his help and comments on this paper. We would like to thank participants of the Linguistic Analysis Design (LAD) sessions (DELPH-IN '15 summit), attendees of the HPSG 2015 conference as well as all others who had given their useful input. This research was supported in part by the MOE Tier 2 grant *That's what you meant: a Rich Representation for Manipulation of Meaning* (MOE ARC41/13). The research reported here is also supported by the project 'Grammar Matrix Reloaded: Syntax and Semantics of Affectedness' (MOE 2013-T2-1-016) funded by the Ministry of Education (MOE) in Singapore.

[NUME-CL-N] phrases are always indefinite. The definiteness interpretation of [CL-N] and [N] phrases vary depending on the dialect. Bare noun, [N], can always have a kind reading. In Mandarin (cmn) and Cantonese (yue), the definiteness interpretations of noun phrases are presented in Table 1 (Cheng & Sybesma, 1999; Sio, 2006).

The definiteness of a noun phrase can affect its distribution. Generally, only definite noun phrases can appear in the subject or topic position in Chinese (Chao, 1968; Lee, 1986; Li & Thompson, 1989, among others). Even though a [CL-N] phrase in Cantonese can be interpreted as either definite or indefinite, a [CL-N] phrase in the subject or topic position can only be interpreted as definite. This is illustrated in (1a) and (1b).[1] The same applies to Mandarin bare nouns, which are only interpreted as definite (or kind) in the subject or topic position as exemplified in (2a) and (2b).[2]

(1) a. 隻 狗 要 過 馬路。
zek3 gau2 jiu3 gwo3 ma5lou6
CL dog want cross road

'The dog wants to cross the road.' [yue]

b. 隻 狗 冇 人 要 呀。
zek3 gau2, mou5 jan4 jiu3 aa3
CL dog, no one want SFP

'That dog, no one wants it.' [yue]

(2) a. 狗 要 過 馬路。
gǒu yāo guò mǎlù
dog want cross road

'The dog wants to cross the road.'
NOT 'A dog wants to cross the road.' [cmn]

b. 狗, 我 不 想 要 了。
gǒu, wǒ bù xiǎng yào le
dog I not want have SFP

'The dog, I don't want to have it (anymore).' [cmn]

For a noun phrase that cannot be interpreted as definite, putting it in the subject or topic position would lead to ungrammaticality. This applies to [CL-N] phrases in Mandarin, (3a), (3b) and bare nouns in Cantonese, (4a), (4b), with the exception of a kind reading, (5).

[1] CL: CLassifiers, SFP: Sentence Final Particle

[2] The examples presented in (1a) and (2a) are taken from Cheng & Sybesma (1999).

(3) a. *隻 狗 要 過 馬路。
zhī gǒu yāo guò mǎlù
CL dog want cross road [cmn]

b. *隻狗, 我 不 想 要 了。
zhī gǒu, wǒ bù xiǎng yàoo
CL dog I not want have

'The dog, I don't want to have it.' [cmn]

(4) a. *狗 要 過 馬路。
gau2 jiu3 gwo3 ma5lou6
dog want cross road [yue]

b. *狗, 冇 人 要 呀。
gau2, mou5 jan4 jiu3 gaa3
dog, no one want SFP

'The dog, no one wants it.' [yue]

(5) 狗 鍾意 食 骨頭。
gau2 zung1ji3 sik6 gwat1tau4
dog like eat bone

'Dogs like to eat bones.' [yue]

[NUME-CL-N] phrases do not show distributive differences between Mandarin and Cantonese with respect to definiteness. They are indefinite in both dialects. However, the distribution of [NUME-CL-N] phrases regarding the subject/topic restriction is more intriguing than the other types of noun phrases. Li (1998) argues that [NUME-CL-N] phrases have two interpretations: quantity-denoting (concerning quantity) or individual-denoting (concerning the existence of certain individuals). A [NUME-CL-N] phrase cannot appear in the subject or topic position unless it has a quantity-denoting reading. We will illustrate the contrast with the subject position using Mandarin data. In (6a), the subject is a [NUME-CL-N] phrase, and it is ungrammatical unless *you* 'have' is added in the front, as in (6b). [3]

(6) a. *三 個 學生 在 學校 受傷 了。
sān gě xuéshēng zài xuéxiào shòushāng le
three CL student at school hurt SFP

'Three students were hurt at school.' [cmn]

b. 有 三 個 學生 在 學校 受傷 了。
yǒu sān gě xuéshēng zài xuéxiào shòushāng le
have three CL student at school hurt SFP

'There are three students hurt at school.' [cmn]

[3](6a) and (6b) are taken from Li (1998).

(7), on the other hand, is grammatical. (7) is a non-episodic sentence and the [NUME-CL-N] phrase in (7) has what Li (1998) calls a quantity-denoting reading. It indicates the rice-eating capacity of (any)'three people' rather than the existence of three specific individuals.

(7) 三 個 人 可以 吃 得 完 一 桶 飯。
sān gè rén kěyǐ chī dè wá yī tǒng fàn
three CL person can eat to.the.extent finish one bucket rice

'Three people can finish one bucket of rice.' [cmn]

Adding *you* 'have' in (7) will make it ungrammatical as *you* 'have' asserts the existence of individuals and thus is only compatible with an individual-reading.

In addition to *you* 'have', it is also possible to save a sentence with a [NUME-CL-N] phrase as the subject by adding *dou* 'all' (Li, 1998). *Dou* 'all' ranges over an entire set of individuals and gives rise to a universal quantification reading. This is illustrated in (8). [4]

(8) 三 個 學生 都 來 這 裡 了。
sān gà xuéshēng dōu lái zhè lǐ le
three CL student all come this place SFP

'Three students all came here.' [cmn]

2.2 The Definite Article and Demonstratives

There are generally 6 situations where the English definite article is used (Lyons, 1977; Hawkins, 1978; Chen, 2004):

(9) a. Situational: Bring me **the hammer**.

b. Anaphoric: I saw a man pass by with a dog. **The dog** was very small and skinny, but **the man** was very large.

c. Shared knowledge: Be quiet. Do not wake up **the baby** (who is sleeping in the next room).

d. Uniqueness: Mary is **the smartest student in my class**.

e. Association: John went to a wedding last weekend. **The bride** was beautiful.

f. With an establishing relative clause: Do you know **the student who slapped the principal in the last Christmas party**?

[4] Example (8) is taken from Li (1998).

In the situational use in (9a), by using the definite article, the speaker indicates to the addressee that he will be able to identity the hammer in the context of the utterance. The use of the definite article in (9b) is anaphoric. The referents of 'the dog' and 'the man' are introduced into the universe of discourse by the previous sentence. In (9c), the definite article is used because the referent, 'the baby', is shared knowledge. In (9d), the definite article is used because the referent is unique (a superlative). (9e) illustrates a case of identifiability via association. The mention of 'a wedding' triggers the identifiability of all the things that are related to 'a wedding' (e.g. bride, cake, etc.) by association. In (9f), the identifibility of the student comes from the post-nominal relative clause. Hawkins (1978) calls the relative clause an 'establishing' relative clause. It establishes the identity of the referent.

The use of demonstratives fall into four major types (Himmelmann, 1996; Chen, 2004):

(10) a. Situational: Could you carry **this huge bag** for me?

b. Discourse Deictic: Your wife is not answering the phone. **This** is not good.

c. Anaphoric: There is a shopping mall about a block from here. You won't find anything interesting in **that mall** though.

d. Recognitional: It was filmed in California, **those dusty kind of hills that they have out here by Stockton and all.**[5]

Demonstratives in situational use is different from the situational use of definite article in that in the former, the subject in question must be visible to the addressee. Consider the following two sentences (Chen, 2004):

(11) a. Beware of the dog.

b. Beware of that dog.

(11b) is felicitous only if the dog is visible. In fact, the implication that there is a dog supposedly visible in the surrounding makes it a much scarier sign.

Demonstratives primarily encodes spatial notions (e.g. proximal vs. distal with respect to the speaker). They are most natural in a contrastive environment, explaining their incompatibility with unique objects:

(12) a. The sun is so bright.

b.??That sun is so bright.

[5]Example (10d) is taken are from Himmelmann (1996).

The anaphoric use of demonstratives involves the transference of spatial notions to the temporal dimensions (Lyons, 1977, p. 670). Deictic location is reinterpreted as location in the universe of discourse. The anaphoric use of the demonstratives is much less common in comparison with their deictic use. When the demonstratives are used anaphorically, it is often with a contrastive sense (Chen, 2004).

The recognitional use is when the speaker does not know with certainty whether a referent is identifiable enough for the addressee. In such situations, the speaker usually prefers a definite expression, which presume some familiarity on the part of the addressee with the referent, rather than using an indefinite expression which treats the referent as non-identifiable (Chen, 2004).

In Mandarin, there are two demonstratives, proximal and distal. Both demonstratives appear in two related forms:[6]

(13) a. proximal: zhè, zhèi

b. distal: nà, nèi

Both forms of the demonstratives can be added directly to a noun (Cheng & Sybesma, 2015):

(14) 這 孩子 眞 頑皮。
zhè/zhèi háizi zhēn wánpí
this/that child really naughty

'This child is very naughty.' [cmn]

The only distributional difference is that *zhèi* and *nèi* cannot constitute a phrase. They cannot appear alone, neither as subjects, (17) nor as objects, (16). Unlike *zhè* and *nà*, which can be used alone as subjects, (15), though not as objects, (17) (Chao, 1968, p. 649).

(15) 這/那 也 不 要 緊。
zhè/Nà yě bù yàojìn
this/that also NEG matter

'This/That also doesn't matter.' [cmn]

(16) *我 要 這/那。
wǒ yāo zhè(zhèi)/nà(nèi)
1SG want this/that

Intended reading: 'I want this/that.' [cmn]

[6] It is generally believed that *zhèi* is historically *zhè* + *yī* 'one' and *nèi* is *nà* + *yī* 'one'

(17) 這/那 是 什麼？
zhèi/*nèi shì shénme
This/That BE what

Intended reading: 'What is this/that?' [cmn]

In Cantonese, the proximal demonstrative is *lei1* 'this' and the distal demonstrative is *go2* 'that'. Unlike Mandarin, in Cantonese, the demonstratives cannot stand alone and it cannot combine with the noun directly.

[DEM-CL-N] phrases are definite in both Mandarin and Cantonese, and they have similar grammatical properties. Chen (2004) claims that demonstratives in Mandarin have developed some functions which are typically served by definite articles in languages like English. He claims that the Mandarin demonstratives can be used anaphorically in a non-contrastive environment in (18), in situation of shared general knowledge in (19), association in (20) and with an establishing relative clause as in (21).[7]

(18) 有 一 個 獵人 養 著 一 隻 狗。
yǒu yī gè lièrén yǎn zhe yī zhī gǒu
have one CL hunter keep PROG one CL dog
這 隻 狗 很 懂事。
zhè zhī gǒu hěn dǒngshì
this CL dog very intelligent

'There was a hunter who had a dog. That dog was very intelligent.' [cmn]

(19) 這 天氣 眞 怪， 十二 月 了，可 一點 不 冷。
zhè tiānqì zhēn guài shièryuè le kě yī diǎn bù lěng
this weather really strange December SFP but one little.bit not cold

'The weather is really strange. It is December now, but it is not cold at all.' [cmn]

(20) 他 買 了 一 輛 舊 車，那 輪胎 都 磨平 了。
tā mǎi le yī liàng jiù chē nà lúntāi dōu mópíng le
3SG buy PERF one CL old car that tire all rub.flat SFP

'He bought an old car. All the tires are worn out.' [cmn]

(21) 上 個 月 來 看 你 的 那 個 人，
shàng gè yuè lái kàn nǐ de nà gè rén
previous CL month come see you DE that CL person
我 今天 又 見 到 他 了。
wǒ jīntiān yòu jiàndào tā le
1SG today again see 3SG SFP

'The person who came to see you last month, I saw him again today.' [cmn]

[7]The Mandarin examples presented in here, (18)-(22), (25), (26), are taken from (Chen, 2004)

In (18), the [DEM-CL-N] phrase *zhè zhī gǒu* is used anaphorically in the absence of contrast. It is also possible to have a bare noun in place of the [DEM-CL-N] phrase, as in (22). In the Cantonese counterpart, either a [DEM-CL-N] phrase or a [CL-N] phrase would be appropriate.

(22) 有 一 個 獵人 養 著 一 隻 狗。
yǒu yī gè lièrén yǎn zhe yī zhī gǒu
have one CL hunter keep PROG one CL dog
狗 很 懂事。
gǒu hěn dǒngshì.
dog very intelligent

'There was a hunter who had a dog. The dog was very intelligent.' [cmn]

The same applies to the Mandarin examples in (19) and (20). It is possible to replace the [DEM-N] phrase in (19) and (20) with just a bare noun. [DEM-N] phrases are not grammatical in Cantonese. For Cantonese, a [CL-N] phrase or a [DEM-CL-N] phrase could be used for the equivalents of (19) and (20), as shown in (23) and (24) below. [8]

(23) (lei1) di1 tin1hei3 zan1 hai6 gwai3, dou1 sap6ji3jyut6 la1, zung6
this CL weather really be strange, already December SFP, still
m4 dung3
not cold

'The weather is really strange. It is December now, but it is not cold at all.' [yue]

(24) keoi5 maai5-zo2 bou6 gau6 ce1, (go2) di1 taai1 dou1
3SG buy-PERF CL old car, that CL tire all
mo4ping4-saai3 ga3 la3
polish.flat-completely SFP SFP

'He bought an old car. All the tires are worn out.' [yue]

Based on the above Mandarin examples, Chen (2004) concludes that the Chinese demonstratives serve some of the functions that are characteristic of the definite article like *the* in English. However, Chinese demonstratives are not yet full-fledged definite articles (Chen, 2004). First of all, they still respect the visibility requirement in situational use. Consider the contrast between (25) and (26) below:

(25) 安靜 點 ， 別 把 那 孩子 吵醒 了。
ānjìng diǎn, bié bǎ nà háizi chǎo-xǐng le
quiet a.little.bit not make that baby wake-up SFP

'Be quiet. Don't wake up that baby.' [cmn]

[8] No Cantonese characters are given in these examples because some characters could not be displayed properly.

(26) 安靜 點 ，別 把 孩子 吵醒 了。
ānjìng diǎn , bié bǎ háizi chǎo-xǐng le
quiet a.little.bit , not make baby wake-up SFP

'Be quiet. Don't wake up the baby.' [cmn]

(25) is infelicitous unless the addressee can see the baby. Furthermore, the demonstratives still require contrastiveness. (27) is unnatural and a bare noun should be used, as in (28).

(27) *那 個 太陽 出 來 了。
nà gè tàiyáng chūlái le
that CL sun come.out SFP

'That sun came out.' [cmn]

(28) 太陽 出 來 了。
tàiyáng chūlái le
sun come.out SFP

'The sun came out.' [cmn]

For Cantonese, a [CL-N] phrase will be appropriate for (26) and (28). In fact, for (28), a bare noun would also be appropriate. It could be because *tàiyáng* 'sun' can be interpreted as a proper name and proper names can always appear bare in Chinese.

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]. [N] in Mandarin and [CL-N] in Cantonese are comparable to both [the x] or [a/an x] in English, except when they appear in the subject or topic position, then they can only mean [the x]. [NUME-CL-N] phrases are always indefinite. They can however still appear in the subject or topic position if they have a quantity-denoting rather than an individual-denoting reading in the sense of (Li, 1998). Chen (2004) shows that the demonstratives in Mandarin show characteristics of some of the functions of the definite articles in languages like English in allowing a non-contrastive anaphoric usage, situational usage, recognitional usage as well as can be used in contexts of shared general knowledge. Cantonese shows similar patterns. There are, however, at least two aspects showing that the Chinese demonstratives are not full-fledged definite articles. In the context of shared knowledge, the visibility requirement still applies. The demonstrative is only admissible if the referent is visible to the addressee. Furthermore, the demonstratives cannot be used with unique objects.

Adopting the framework of HPSG (Pollard & Sag, 1994) and MRS(Copestake et al., 2005), this section presents an analysis that models the different definiteness

interpretations of the four types of NPs in Mandarin and Cantonese, as well as the requirement that Chinese subjects need to be definite. Not all the observations presented earlier on can be modeled at this stage. We will leave those further research.

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 & Haugereid (2005) and Bender & 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 & Haugereid (2005) and Bender & 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 & Haugereid (2005) and Bender & Goss-Grubbs (2008) use the givenness hierarchy (Prince, 1981; Gundel et al., 1993). From right to left in Table 2, each type is exemplified in (29).

Table 2: Givenness hierarchy

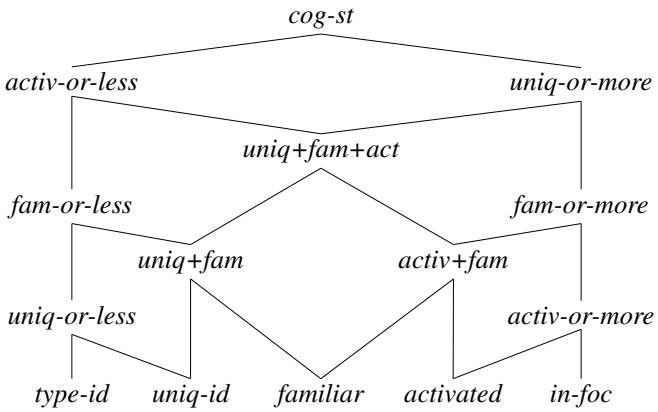
In focus	> Activated	> Familiar	> Uniq. id	> Referential	> Type id
it	*this*, *that* *this N*	*that N*	*the N*	indefinite *this N*	*a N*

(29) a. I couldn't sleep last night.

b. i. A dog (next door) kept me awake.
ii. This dog (next door) kept me awake.
iii. The dog (next door) kept me awake.
iv. That dog (next door) kept me awake.
v. That kept me awake.
vi. It kept me awake.
(Borthen & Haugereid, 2005, p. 230)

Along this line, Borthen & Haugereid (2005) provide an HPSG-based type hierarchy of cognitive status, which was then slightly refined by Bender & Goss-Grubbs (2008) as sketched out in (30).

(30)



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 (30), Table 1 is now converted into Table 3.

Table 3: Cognitive status

type	**example**	**Mandarin**	**Cantonese**
DEM-CL-N	這隻狗	*uniq-or-more*	
NUME-CL-N	三隻狗	*type-id*	
CL-N	隻狗	*type-id*	*cog-st*
N	狗	*cog-st*	*type-id*

First, if a particular construction conveys only definite meaning, the phrase places the *uniq-or-more* feature to the head noun as indicated in the second row in Table 3. Notice that in the *cog-st* hierarchy provided in (30) *uniq-or-more* excludes the leftmost item (i.e. *type-id*) that signals indefiniteness from its subtypes. In this way, *uniq-or-more* indicates that the NP can be evaluated as containing definiteness. Note also that 'Activated' and 'Familiar' in Table 2 are instantiated as NPs with demonstratives (i.e., *this N*, and *that N*). Since *uniq-or-more* includes these meanings, [DEM-CL-N] in the second row of Table 3 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 remains underspecified as *cog-st*. This means that an NP whose value of cognitive status is underspecified can be interpreted as either indefinite or definite.[9]

[9]We defer to the corpus-based findings provided in Gundel et al. (1993).

3.2 Phrase Structure Rules

In Table 3, 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.

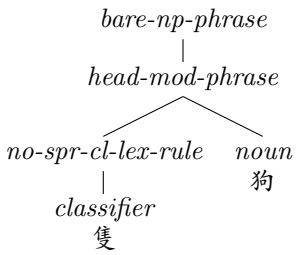
First of all, Mandarin and Cantonese share the following lexical type of classifiers, in which the element of MOD goes for the head noun, the element of SPR (i.e. specifier) goes for demonstratives and numerals. For example, in 這 隻 狗 'this CL dog', 這 and 狗 are constrained as SPR and MOD, respectively.

(31)
$$\begin{bmatrix} \textit{classifier} \\ \text{LTOP} \quad \boxed{1} \\ \text{INDEX} \quad \boxed{2} \\ \text{ARG1} \quad \boxed{2} \\ \text{MOD} \quad \left\langle \begin{bmatrix} \textit{noun} \\ \text{LTOP} \quad \boxed{1} \\ \text{INDEX} \quad \boxed{2} \\ \text{SPR} \quad \langle [\,] \rangle \\ \text{COG-ST} \quad \textit{cog-st} \end{bmatrix} \right\rangle \\ \text{SPR} \quad \left\langle \begin{bmatrix} \text{LTOP} \quad \boxed{1} \\ \text{INDEX} \quad \boxed{2} \end{bmatrix} \right\rangle \end{bmatrix}$$

Classifiers themselves impose no COG-ST constraint on the head noun, given that any types of COG-ST value can be assigned to the NPs with classifiers.

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 (32). 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.

(32)
$$\begin{bmatrix} \textit{no-spr-cl-lex-rule} \\ \text{MOD} \quad \left\langle \begin{bmatrix} \text{COG-ST} \quad \textit{type-id} \end{bmatrix} \right\rangle \\ \text{SPR} \quad \langle\,\rangle \\ \text{ARGS} \quad \left\langle \begin{bmatrix} \textit{classifier} \\ \text{SPR} \langle [\,] \rangle \end{bmatrix} \right\rangle \end{bmatrix}$$



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 (32), but it takes nouns as its daughter. The rule is Cantonese-specific.

(33)
$$\left[\begin{array}{ll} \textit{no-cl-lex-rule} & \\ \text{MOD} & \left\langle \left[\text{COG-ST} \quad \textit{type-id}\right] \right\rangle \\ \text{SPR} & \langle\ \rangle \\ \text{ARGS} & \left\langle \left[\begin{array}{l} \textit{noun} \\ \text{SPR} \left\langle [\ ] \right\rangle \end{array}\right] \right\rangle \end{array}\right] \qquad \begin{array}{c} \textit{bare-np-phrase} \\ | \\ \textit{no-cl-lex-rule} \\ | \\ \textit{noun} \\ \text{狗} \end{array}$$

Bare-np-phrase used in the parse trees of (32-33) is constrained as represented in the following AVM. This non-branching rule signals *cog-st* (i.e. underspecified) and introduces an existential quantifier (i.e. *exist_q_rel*) into the RELS list.

(34)
$$\left[\begin{array}{ll} \textit{bare-np-phrase} & \\ \text{HD} & \left[\begin{array}{ll} \textit{noun} & \\ \text{COG-ST} & \textit{cog-st} \\ \text{LTOP} & \boxed{1} \\ \text{INDEX} & \boxed{2} \end{array}\right] \\ \text{C-CONT} & \left[\begin{array}{ll} \text{RELS} & \left\langle\ !\left[\begin{array}{ll} \text{PRED} & \textit{exist\_q\_rel} \\ \text{ARG0} & \boxed{2} \\ \text{RSTR} & \boxed{3} \end{array}\right]!\ \right\rangle \\ \text{HCONS} & \left\langle\ !\left[\begin{array}{ll} \textit{qeq} & \\ \text{HARG} & \boxed{3} \\ \text{LARG} & \boxed{1} \end{array}\right]!\ \right\rangle \end{array}\right] \end{array}\right]$$

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 (32-33) are constrained as [COG-ST *type-id*]. Because *type-id* is a specific subtype of *cog-st*, 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 (35).[10]

[10]Since proper names and clausal subjects are not indefinite, this constraint does not affect other types of subjects.

(35) $\begin{bmatrix} \textit{subj-head-phrase} \\ \text{NHD} \mid \text{COG-ST} \quad \textit{uniq-or-more} \end{bmatrix}$

Note that *uniq-or-more* is mutually exclusive with *type-id*, as represented in the type hierarchy (30). For instance, the structures provided in (32-33) 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 (36) and (37).

(36) 隻 狗 走 啦。
zek3 gau2 zau2 la3
CL dog leave SFP

'The dog is leaving.' [yue]

(37) 狗 走 了。
gǒu zǒu le
dog leave SFP

'The dog is leaving.' [cmn]

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.

Figure 1 representing (36) 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 狗 'dog' is specified as *uniq-or-more* following the constraint presented in (35). Note that the NP 隻 狗 'CL-dog' itself is assigned *cog-st* as the value of COG-ST, as shown on the tree. The value becomes more hierarchically specific when the NP is used as the non-head daughter of *subj-head-phrase*: When the NP is combined with the verb 走 'leave' to form a *subj-head-phrase*, the subject is assigned [COG-ST *uniq-or-more*].

Figure 2 representing (37) 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

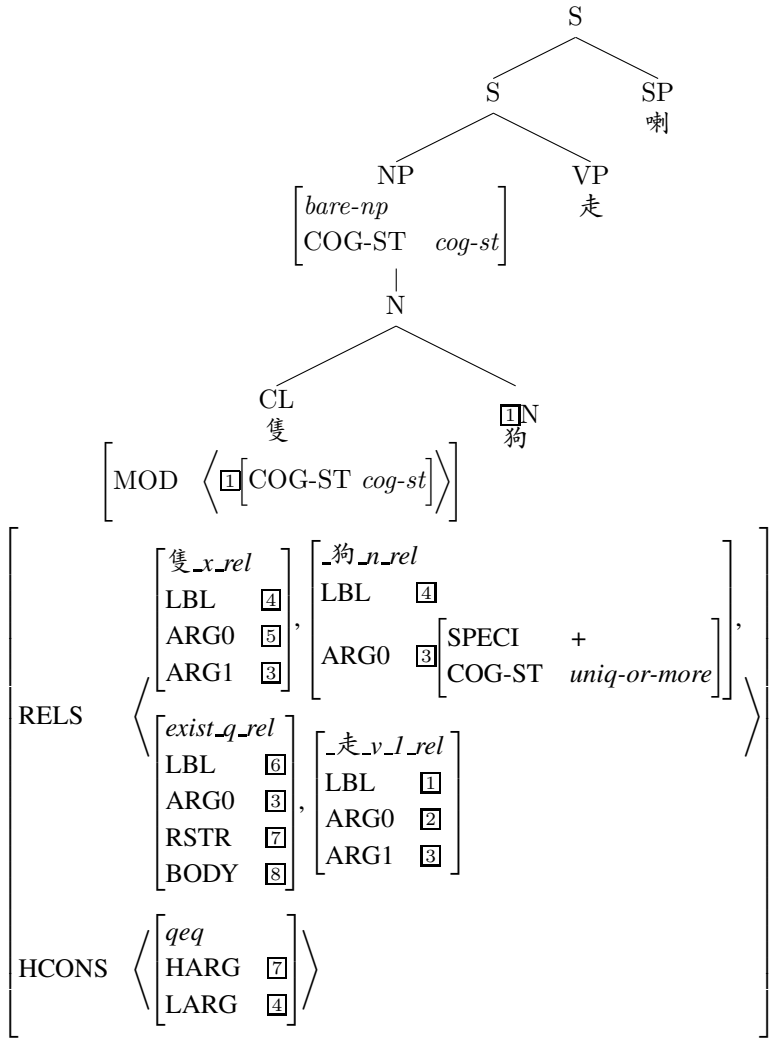


Figure 1: A sample derivation in Cantonese

subject position, it can only be interpreted as definite. The Cantonese counterpart of this sentence would be ungrammatical as [N] phrases can only be indefinite in Cantonese. The COG-ST of the subject 狗 'dog' in the MRS representation is specified as *uniq-or-more* in the same way as Figure 1.[11]

References

Bender, Emily M. & David Goss-Grubbs. 2008. Semantic Representations of Syntactically Marked Discourse Status in Crosslinguistic Perspective. In *Proceed-*

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

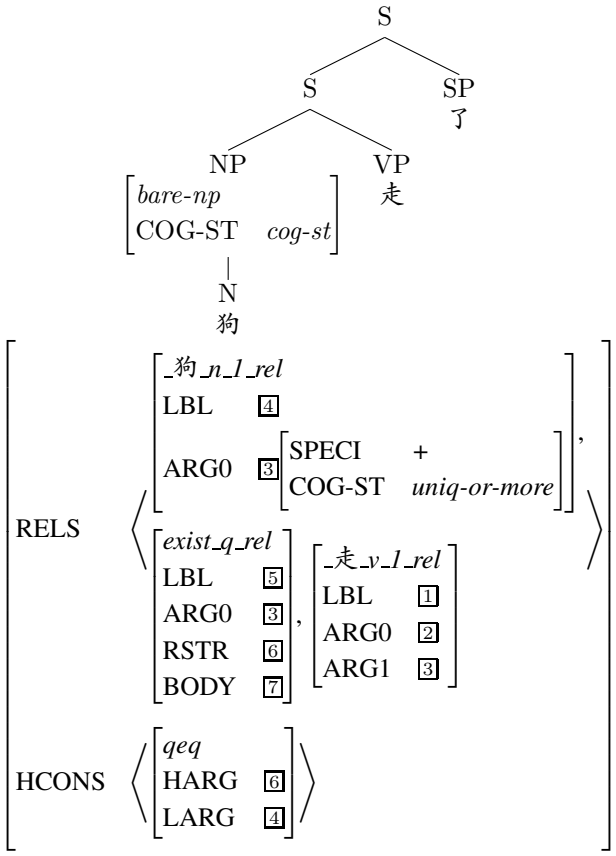


Figure 2: A sample derivation in Mandarin

ings of the 2008 conference on semantics in text processing, 17–29. Association for Computational Linguistics.

Borthen, Kaja & Petter Haugereid. 2005. Representing Referential Properties of Nominals. *Research on Language and Computation* 3(2-3). 221–246.

Chao, Yuan Ren. 1968. *Grammar of Modern Spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chen, Ping. 2004. Identifiability and Definiteness in Chinese. *Linguistics* 42(6). 1129–1184.

Cheng, Lisa Lai-Shen & Ring Sybesma. 2015. Mandarin. In Tibor Kiss & Artemis Alexiadou (eds.), *Syntax – theory and analysis. an international handbook. handbooks of linguistics and communication science*, 42.1–3. Berlin: Mouton de Gruyter.

Cheng, Lisa Lai-Shen & Rint Sybesma. 1999. Bare and not-so-bare Nouns and the Structure of NP. *Linguistic Inquiry* 30(4). 509–542.

Copestake, Ann, Dan Flickinger, Carl Pollard & Ivan A. Sag. 2005. Minimal Recursion Semantics: An Introduction. *Research on Language & Computation* 3(4). 281–332.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69(2). 274–307.

Hawkins, John A. 1978. *Definiteness and Indefiniteness: a Study in Reference and Grammaticality Prediction*. London: Croom Helm.

Himmelmann, Nikolaus P. 1996. Demonstratives in Narrative Discourse: A Taxonomy of Universal Uses. In Barbara Fox (ed.), *Studies in anaphora*, 205–254. Amsterdam: John Benjamins.

Lee, Hun-tak Thomas. 1986. *Studies on Quantification in Chinese*: University of California, Los Angeles dissertation.

Li, Charles N & Sandra A Thompson. 1989. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley and Los Angeles: University of California Press.

Li, Yen-hui Audrey. 1998. Argument Determiner Phrases and Number Phrases. *Linguistic Inquiry* 29(4). 693–702.

Lyons, John. 1977. *Semantics. Vols. I and II*. Cambridge: Cambridge University Press.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL: The University of Chicago Press.

Prince, Ellen F. 1981. Toward a Taxonomy of Given-New Information. In Peter Cole (ed.), *Radical pragmatics*, 223–256. New York: Academic Press.

Sio, Joanna Ut-Seong. 2006. *Reference and Modification in the Chinese Nominal*. the Netherlands: LOT Publication.

Yuan, Jiahua. 1983. *Hànyǔ fāngyán gàiyào [Outline of Chinese Dialects]*. Beijing: Wenzi Gaige.