



Universiteit
Leiden

The Netherlands

Classifiers in four varieties of Chinese

Cheng, L.L.; Sybesma, R.P.E.; Cinque, G.; Kayne, R.S.

Citation

Cheng, L. L., & Sybesma, R. P. E. (2005). Classifiers in four varieties of Chinese. In G. Cinque & R. S. Kayne (Eds.), *The Oxford handbook of comparative syntax* (pp. 259-292). Oxford: Oxford University Press.

doi:10.1093/oxfordhb/9780195136517.013.0007

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/2823>

Note: To cite this publication please use the final published version (if applicable).

CHAPTER 7

CLASSIFIERS IN FOUR VARIETIES OF CHINESE

LISA L.-S. CHENG AND
RINT SYBESMA

1 INTRODUCTION

This chapter deals with the use of classifiers in four varieties of Chinese: Mandarin, Wu, Min, and Cantonese (or “Yue”). The main aim is to account for the distribution and interpretation of the different forms of nominal expressions in these languages, as well as for the variation that they display in this regard. To this end, we investigate the question of what parameters play a role and attempt to find deeper reasons for some of the systematic contrasts.

It is generally assumed that there are seven major Chinese language groups (or “dialects”), four of which are looked at in this essay. The term “Mandarin” is ambiguous: it either is an alternative name for the natural variety of Chinese otherwise referred to as “the Northern-Chinese dialect,” or it is the name for the standard language, adopted as the official language on the mainland, as well as on Taiwan. Roughly speaking, as a natural (as opposed to standard) language, Mandarin is spoken north of the Yangtze River and in the southern provinces of Yunnan and Guizhou. Wu is spoken in Zhejiang and the southern tip of Jiangsu. Min is the language of Fujian and neighboring parts of Guangdong, and of Hainan and Taiwan. Cantonese is spoken in most of Guangdong and Guangxi, as well as

in Hong Kong. For an excellent introduction to several aspects of Chinese and its varieties, see Norman 1988.¹

Despite the fact that the area where Mandarin is spoken is vast, the internal variation is not very great: in essence, the Mandarin subvarieties are mutually intelligible. This does not apply to the other dialect groups (see Ramsey 1987 for a suggestion why this would be so): the respective areas in which each of these is spoken is much smaller, but the variation is enormous, typically to the point of mutual unintelligibility. This is especially true of Wu and Min, which are standardly divided into Northern and Southern Wu and Northern and Southern Min.

Unless explicitly mentioned otherwise, our Mandarin data belong to the variety spoken in the north-northeastern regions of China. For Wu, we concentrate on the Southern Wu variety of Wenzhou, but now and then we mention Shanghainese, as a representative of the northern branch. The Min data belong to the Southern Min variety spoken on Taiwan. The variety of Cantonese we use is spoken in Hong Kong.

2 THE FACTS

In this section we lay out the facts that will concern us in this chapter, concentrating on the correlations between form, interpretation, and distribution of three different types of nominal expressions in Mandarin, Min, Wu, and Cantonese: bare nouns (to be referred to as “Bare NPs”), phrases consisting of a classifier and a nominal (“Cl-NPs”) and expressions made up of a numeral, a classifier, and a nominal (“Num-Cl-NPs”). We present the facts language by language, going from north to south.

2.1 Mandarin

2.1.1 *Bare NPs*

As the following examples show, Bare NPs may receive an indefinite, a definite, or a generic interpretation. What interpretation it gets is essentially determined by the nature of the predicate; in this chapter we do not go into this matter. In sentences with an unbounded activity verb as in (1a), the Bare NP is interpreted as indefinite. In bounded events, like the one expressed in (1b), the Bare NP gets a definite reading. With unbounded states, the Bare NP is generic, as is shown in (1c). (For discussion of (1b), see Sybesma 1992, 176–178.)

- (1) a. Hufei maishu qu le.
 Hufei buy book go SFP²
 ‘Hufei went to buy a book/books.’
- b. Hufei he-wan-le tang.
 Hufei drink-finished-PRF soup
 ‘Hufei finished the soup.’
- c. Wo xihuan gou.
 I like dog
 ‘I like dogs.’

In preverbal position, however, Mandarin Bare NPs cannot be interpreted as indefinite. They get either a definite or a generic interpretation:

- (2) a. Gou yao guo malu.
 dog want cross road
 ‘The dog/the dogs want/s to cross the road.’
 not: ‘A dog/dogs want/s to cross the road.’
- b. Gou jintian tebie tinghua.
 dog today very obedient
 ‘The dog/dogs was/were very obedient today’ (not: indefinite)
- c. Gou ai chi rou.
 dog love eat meat
 ‘Dogs love to eat meat’

2.1.2 *Cl-NPs*

Interpretationally, the Cl-NP in Mandarin is limited to a nonspecific indefinite reading. As a result, it is limited distributionally, to the object position in unbounded activity predicates, as is exemplified in (3): (3a) is fine, but in the bounded predicate of (3b), which forces a strong (i.e., specific) interpretation onto indefinites (Sybesma 1992, 176–178), a Cl-NP is not possible. Cl-NPs cannot get a generic interpretation either. In preverbal position, Cl-NPs are also out:

- (3) a. Wo xiang mai ben shu.
 I want buy CL^{volume} book
 ‘I would like to buy a book.’
- b. *Ta he-wan-le wan tang.
 he drink-finished-PRF CL^{bowl} soup
 Intended: ‘He finished a (specific) bowl of soup.’

- c. *Wo xihuan wan tang.
I like CL^{bowl} soup

Intended generic reading: 'I love a bowl of soup.'

- (4) a. *Zhi gou yao guo malu.
CL dog want cross road

Intended: 'A dog wants to cross the road.'

- b. *Zhi gou xihuan chi rou.
CL dog like eat meat

Intended generic reading: 'A dog likes to eat meat.'

It is not generally acknowledged that Mandarin has Cl-NPs (but see Paris 1981); Cl-NPs like the one in (3a) are often considered to be the result of phonological reduction of the numeral *yi* 'one'; according to this reasoning, (3a) is the reduced form of the unreduced (5):

- (5) Wo xiang mai yi-ben shu.
I want buy one-CL book

'I would like to buy a book.'

However, the fact that, as we just saw, Cl-NP is limited to contexts which are definable in semantic, rather than phonological, terms suggests that the reduction view is wrong. There is no phonological reason, for instance, why *yi* could not be reduced in (3b), if (6) were the "unreduced" form of (3b).

- (6) Ta he-wan-le yi-wan tang.
he drink-finished-PRF one-CL^{bowl} soup

'He finished a/one (specific) bowl of soup.'

More arguments for the claim that Cl-NP is a real construct in Mandarin are given in Cheng and Sybesma 1999.

2.1.3 *Num-Cl-NPs*

Mandarin Num-Cl-NP shows basically the same distribution as Cl-NP, the only difference being that it can occur in the object position of bounded predicates (as we just saw in (6)), by virtue of the fact that it can be interpreted as specific. The complete paradigm is given in (7)–(8). Num-Cl-NPs in which the numeral is *yi* (i.e., *yi*-Cl-NP) may be interpreted as specific and nonspecific, comparable to indefinite articles in Germanic languages; it can, of course, also be stressed and mean *one*:

- (7) a. Wo xiang mai yi-ben shu.
I want buy one-CL book
'I would like to buy a book.'
- b. Ta he-wan-le yi-wan tang.
he drink-finished-PRF one-CL^{bowl} soup
'He finished a/one (specific) bowl of soup.'
- c. *Wo xihuan yi-wan tang.
I like one-CL^{bowl} soup
Intended generic reading: 'I love a/one bowl of soup.'
- (8) a. *Yi-zhi gou yao guo malu.
one-CL dog want cross road
Intended: 'A dog wants to cross the road.'
- b. *Yi-zhi gou xihuan chi rou.
one-CL dog like eat meat
Intended generic reading: 'A dog likes to eat meat.'

2.1.4 Summary

In Mandarin, Bare NPs can be interpreted as definite, indefinite, or generic. Num-CL-NPs and CL-NPs are invariably indefinite. A difference between the two is that Num-CL-NPs can be both specific and nonspecific and CL-NPs are limited to a nonspecific interpretation. All indefinites occur in postverbal position only.

2.2 Wu

2.2.1 Bare NPs

The distribution and interpretation of Bare NPs in Wu is the same as in Mandarin. As objects, they can be interpreted as indefinite, definite, and generic, depending on the nature of the predicate. The following sentences are Wenzhou examples:³

- (9) a. Vu² Fei¹ tsau³-khe⁵ ma⁴ si¹ fuo² ba⁴.
Vu Fei go buy book SFP SFP
'Vu Fei went to buy a book/books.'
- b. ɲ⁴ dei⁶ thu⁵ ha⁷ jy² ba⁴.
I take soup drink up SFP
'I finished the soup'

- c. ŋ⁴ si³-ɕy¹ kau³.
 I like dog
 'I like dogs.'

As to (9b), in bounded predicates Wu disallows objects in postverbal position. Instead, the object appears in preverbal position, following the element *dei*⁶ meaning 'take'.⁴

In preverbal position, Wu displays the same interpretational pattern for Bare NPs as Mandarin, as is exemplified by the following Wenzhou data. They either get a definite or a generic interpretation; an indefinite reading is excluded:

- (10) a. Kau³ i⁵ tshi⁷ niou⁸.
 dog want eat meat
 'The dog/the dogs want/s to eat meat.' (Not: indefinite)
- b. kau³ ke⁷-ne⁸ de⁸-bi⁸ teŋ¹-kuo³.
 dog today very obedient
 'The dog/the dogs was/were very obedient today.' (Not: indefinite)
- c. kau³ si³-ɕy¹ tshi⁷ niou⁸.
 dog like eat meat
 'Dogs like to eat meat.'

2.2.2 *Cl-NPs*

With respect to Cl-NPs, the situation is not the same in all Wu dialects. While in some varieties the facts are quite similar to those of Mandarin, Wenzhou is very different. As we saw, in Mandarin Cl-NPs can only be interpreted as indefinite, and their distribution is limited to postverbal position. In contrast, Wenzhou Cl-NPs may occur in both preverbal and postverbal position, and it may be interpreted as definite as well as indefinite (nonspecific). Let us look at Cl-NPs in postverbal position first:

- (11) a. ŋ⁴ ɕi³ ma⁴ paŋ³ si¹ le² tshi⁵.
 I want buy CL^{volume} book come read
 'I would like to buy a book to read.'
- b. ŋ⁴ dei⁶ y⁷/lie⁷ thu³ ha⁷ jy² ba⁴.
 I take CL^{bowl}/CL^{PL} soup drink up SFP
 'I finished the (bowl of) soup.'
- c. *ŋ⁴ si³-ɕy¹ ha⁷ y⁷/lie⁷ thu³.
 I like drink CL^{bowl}/CL^{PL} soup
 Intended generic reading: 'I like (a bowl of) soup.'

The use of the CL-NP in (11a) is similar to (3a): *paŋ³ si¹* ‘CL-book’ is interpreted as indefinite. However, the CL-NP in (11b) is interpreted as definite: whether *y⁷* ‘bowl’ or *lie⁷* ‘some’ is used makes no difference; in both cases the CL-NP combination translates as ‘the soup’ (we return to this later).

Let us look at the use of CL-NPs in preverbal position. Like Mandarin, Wenzhou excludes both an indefinite and a generic interpretation, but unlike Mandarin, preverbal CL-NPs in Wenzhou can be interpreted as definite:

- (12) a. Dyu⁸ kau⁸ i⁵ tsau³-ku⁵ ka¹-løy⁶.
 CL dog want walk-cross street
 ‘The dog wants to cross the street.’
- b. *Dyu⁸ kau³ si³-cy¹ tshi⁷ niou⁸.
 CL dog like eat meat
 Intended generic reading: ‘Dogs like to eat meat.’
 (Only possible reading: ‘The dog likes to eat meat.’)

Interestingly, when CL-NPs in Wenzhou are interpreted as definite, the tone of the classifier is affected. Wenzhou has eight tones (Norman 1988, Pan 1991), which are divided in four different classes, which for practical reasons we will call A, B, C, and D. Each of these classes, in turn is divided in two subclasses according to register, hi and lo. Thus, we have a hi-A, a lo-A, a hi-B, a lo-B, and so on, and they are numbered accordingly (see table 7.1 and the examples): 1 for hi-A, 2 for lo-A, and so on. Given a five-point scale (1 = low, 5 = high), the tonal contour values are in table 7.1. Chen’s (2000: 476) values are partly different, the biggest difference being that, according to him, the D-tones have a dipping contour. The values Chen provides are 313 and 212 for hi-D and lo-D, respectively.

What happens to the tone of the classifier when CL-NPs are interpreted as definite is that, along register-lines, it invariably is pronounced as a D-tone, whatever its original tone may have been: hi-A (tone number 1), hi-B (number 3), and hi-C (number 5), as well as hi-D (number 7), surface as hi-D (number 7), while lo-A (tone number 2), lo-B (number 4), and lo-C (number 6), as well as lo-D

Table 7.1 Tonal contour values of the eight tones of Wenzhou

Tone	1	2	3	4	5	6	7	8
	hi-A	lo-A	hi-B	lo-B	hi-C	lo-C	hi-D	lo-D
Value	44	31	45 (abrupt)	24 (abrupt)	42	11	23	12

(number 8), surface as lo-D (number 8) (You 2000; see also Pan 1991). For now, we call this process “tonal neutralization”; we discuss it more in section 4.1.2.

Let’s look at some examples. The original tone of the classifier for dogs *dyu* is the lo-A (tone number 2): *dyu*². In (12a), where the CL-NP is interpreted as definite, the tone changes to lo-D (tone number 8). If we change the tone in (12a) to its original lo-A (as we do in (13)), the sentence is ungrammatical: because of the tone, we can no longer have a definite reading and, just like in Mandarin, an indefinite reading is excluded in preverbal position:

- (13) **dyu*² *kau*⁸ *i*⁵ *tsau*³-*ku*⁵ *ka*¹-*lø*⁶
 CL dog want walk-cross street

In postverbal position, *dyu*² *kau*⁸ ‘CL dog’ would of course be grammatical—although it has an indefinite interpretation:

- (14) *ŋ*⁴ *çi*³ *ma*⁴ *dyu*² *kau*⁸.
 I want buy CL DOG
 ‘I would like to buy a dog.’

To illustrate the mechanism one more time, we present a minimal pair in (15) (cf. (11a)):

- (15) a. *ŋ*⁴ *çi*³ *ma*⁴ *paŋ*³ *si*¹.
 I want buy CL^{volume} book
 ‘I want to buy a book.’
 b. *ŋ*⁴ *çi*³ *ma*⁴ *paŋ*⁷ *si*¹.
 I want buy CL^{volume} book
 ‘I want to buy the book.’

The only difference between (15a) and (15b) is the tone on the classifier *paŋ* ‘volume’. Its original tone is hi-B, (number 3). Keeping the original tone, the CL-NP in (15a) is interpreted as indefinite. In (15b), *paŋ* appears in the hi-D-tone (number 7), and the CL-NP receives a definite interpretation.

Some classifiers inherently have a D-tone. In the appropriate context, CL-NPs with these classifiers are ambiguous:

- (16) *ŋ*⁴ *çi*³ *ha*⁷ *y*⁷/*lie*⁷ *thu*⁵¹.
 I want drink CL^{bowl}/CL^{PL} soup
 ‘I would like to drink a bowl of soup/some soup.’
 or: ‘I would like to drink the soup.’

In short, Wenzhou CI-NPs can be interpreted as definite, as long as the classifier is pronounced with a D-tone. For all classifiers that do not inherently have a D-tone, this means a tone change, to be referred to as tonal neutralization (for discussion, see section 4.1.2).⁵

2.2.3 *Num-Cl-NPs*

With respect to Num-Cl-NPs, Wu is the same as Mandarin. They are interpreted as indefinite, either specific or nonspecific. Distributionally, they are limited to postverbal position. Here are some Wenzhou examples. The counterparts of (7c) and (8a,b) are also ungrammatical in Wu/Wenzhou. It should be noted, that the classifier preceded by *i*⁷ ‘one’ keeps its inherent tone:

- (17) a. ɲ⁴ ɕi³ ma⁴ i¹ paŋ³ si¹ le² tshi⁵.
 I want buy one CL^{volume} book come read
 ‘I would like to buy a book to read.’ (nonspecific)
- b. ɲ⁴ dei⁶ i⁷ y⁷ thuɔ¹ ha⁷ jy² ba⁴.
 I take one CL^{bowl} soup drink up SFP
 ‘I finished a bowl of soup.’ (specific)

2.2.4 *Summary*

In Wu, Bare NPs can be interpreted as definite, indefinite, or generic. Cl-NPs are interpreted as indefinite in all varieties of Wu (nonspecific only); in some, like Wenzhou, they can also be interpreted as definite, in which case the tone of the classifier is affected in the sense that some neutralization takes place. Num-Cl-NPs are invariably indefinite (specific or nonspecific). All indefinites occur in postverbal position only.

2.3 *Min*

The situation in Southern Min is the same as Mandarin, with one exception: it has no Cl-NPs, not even in an indefinite reading (Chen 1958, Zhou 1991).

2.3.1 *Bare NPs*

Thus, just like in Mandarin (as was the case in Wu), Bare NPs can be interpreted as definite, indefinite, and generic, depending, partly, on the predicative context and the position vis-à-vis the verb (preverbal position excluding an indefinite interpretation). The sentences in (18) and (19) are parallel to the Mandarin ex-

amples in (1) and (2); as to (18b), like Wu, Min disallows postverbal objects in bounded predicates, and, like Wu (see (9b)), it marks the object in preverbal position, using an element meaning ‘take’:

- (18) a. I be bue zhu.
 he want buy book
 ‘He would like to buy a book/books.’
- b. I ga teN lim liao a.
 he take soup drink PRF SFP
 ‘He finished the soup.’
- c. I ai hue.
 he like flower
 ‘He likes flowers.’
- (19) a. Gau be lim zhui.
 dog want drink water
 ‘The dog/dogs want/s to drink water.’ (definite only)
- b. Gau ai lim zhui.
 dog like drink water
 ‘Dogs like to drink water.’ (definite possible; indefinite excluded)

2.3.2 *Cl-NPs*

It is explicitly mentioned in the literature (e.g., Chen 1958, Zhou 1991) and confirmed by the informants that Southern Min has no Cl-NPs: the classifier can never occur without being preceded by either a numeral or a demonstrative.⁶ Thus, the counterparts of Mandarin (3a) and Wenzhou (12a) are ungrammatical in Southern Min:

- (20) a. *Ua siuN bue bun zhu.
 I want buy CL^{volume} book
 Intended: ‘I would like to buy a book.’
- b. *Jia gau be lim zhui.
 CL dog want drink water
 Intended: ‘The dog wants to drink water.’

2.3.3 *Num-Cl-NPs*

Just like in Mandarin and Wu, Southern Min Num-Cl-NPs receive a specific or nonspecific indefinite interpretation:

- (21) a. Ua siuN bue jit-bun zhu.
 I want buy one-CL^{volume} book
 'I would like to buy a book.' (nonspecific)
- b. I ga jit-waN teN lim liao a.
 he take one-CL^{bowl} soup drink PRF SFP
 'He finished a bowl of soup.' (specific)

2.3.4 Summary

In Southern Min, Bare NPs can be interpreted as definite, indefinite, or generic. Num-CL-NPs are invariably indefinite, either specific or nonspecific. CL-NPs do not occur in this language in either an indefinite or in a definite interpretation. As before, all indefinites occur in postverbal position only.

2.4 Cantonese

2.4.1 Bare NPs

Cantonese is different from all three varieties of Chinese previously described in that Bare NPs cannot be interpreted as definite: they can receive only a generic and an indefinite reading:

- (22) a. Wufei heoi maai syu.
 Wufei go buy book
 'Wufei went to buy a book/books.'
- b. *Wufei jam-jyun tong la.
 Wufei drink-finish soup SFP
 Intended: 'Wufei finished drinking the soup.'
- c. Ngo zung-ji gau.
 I like dog
 'I like dogs.'
- (23) a. *Gau soeng gwo maalou.
 dog want cross road
 Intended: 'The dog wants to cross the road.'
- b. Gau zung-ji sek juk.
 dog like eat meat
 'Dogs love to eat meat.' (only interpretation possible)

2.4.2 *Cl-NPs*

With respect to Cl-NPs, Cantonese is similar to Wenzhou: aside from an indefinite interpretation, Cl-NPs can also be interpreted as definite. In Cantonese, however, there are no tonal effects. The indefinite use is illustrated in (24a); the definite interpretation is exemplified in (24b,c) (cf. (22b) and (23a); also, see the Wenzhou sentences in (11b) and (12a)):

- (24) a. Ngo soeng maai bun syu lei taai.
 I want buy CL^{volume} book come read
 ‘I want to buy a book to read.’
- b. Wufei jam-jyun wun/di tong la.
 Wufei drink-finish CL^{bowl}/CL^{pl} soup SFP
 ‘Wufei finished the soup.’
- c. Zek gau soeng gwo maalou.
 CL dog want cross road
 ‘The dog wants to cross the road.’

2.4.3 *Num-Cl-NPs*

Num-Cl-NPs have the same distribution and interpretation in Cantonese as they do in the other varieties of Chinese, discussed previously.

2.4.4 *Summary*

In Cantonese, Bare NPs can be interpreted as indefinite or generic, but not as definite. Cl-NPs can be either definite or indefinite (nonspecific), with no tonal repercussions. Num-Cl-NPs are invariably indefinite (specific or nonspecific). As in all the other varieties, all indefinites occur in postverbal position only.

2.5 Summary of Form-Interpretation Correlations

The form–interpretation correlations are summarized in table 7.2. Depending on the perspective chosen, these facts can be described in different ways. From the perspective of the formal properties, we can summarize them as follows:

Formal Properties

1. Num-Cl-NPs are the same in all languages: they allow for only an indefinite reading. Later we present a proposal to explain why that is the case.
2. Cl-NPs show more variation. In Min, they are not possible at all. In Mandarin, Cantonese, and Wu, they occur, but while in Mandarin they

Table 7.2 Correlations between form, interpretation, and distribution of nominal expression in four Chinese languages

Interpretation	Form		
	Num-Cl-NP	Cl-NP	Bare NP
Indefinite	Mandarin	Mandarin	Mandarin
	Wu	Wu	Wu
	Min	*Min	Min
	Cantonese	Cantonese	Cantonese
Definite	*Mandarin	*Mandarin	Mandarin
	*Wu	Wu—w. tone neutr.	Wu
	*Min	*Min	Min
	*Cantonese	Cantonese	*Cantonese
Generic	*Mandarin	*Mandarin	Mandarin
	*Wu	*Wu	Wu
	*Min	*Min	Min
	*Cantonese	*Cantonese	Cantonese

*Not found.

can receive only an indefinite reading, in Cantonese and Wenzhou they can be interpreted as definite as well. Wenzhou has the extra feature of tonal neutralization affecting the classifier in definite Cl-NPs, while the indefinite counterpart does not involve such tonal neutralization. In Cantonese, the tones are unaffected. Later, we propose an explanation as to why Cl-NPs can be interpreted as indefinite and definite, why Cl-NPs cannot be definite in Mandarin while they can in Cantonese and Wenzhou, and what the significance of the tonal neutralization in Wenzhou definite Cl-NPs is.

3. Bare NPs can be interpreted as indefinite and generic in all four languages. As far as the definite interpretation is concerned, Bare NPs can be so interpreted in Mandarin, Wenzhou and Min—that is, in the two languages in which Cl-NPs cannot be interpreted as definite (Mandarin and Min) and in the language in which the definite Cl-NP involves tonal neutralization for the classifier. In Cantonese Bare NPs cannot be interpreted as definite. Later, we investigate the question as to how it is possible that Bare NPs can be interpreted as indefinite, definite, and generic, as well as the question why Cantonese is different.

From the perspective of the interpretative properties, the facts in table 7.1 may be summarized as follows:

Interpretive Properties

1. To express an indefinite reading, the four languages are basically the same: Num-Cl-NPs, Cl-NPs, and Bare NPs can be interpreted indefinitely in all cases, the only exception being Min, in which Cl-NP-phrases do not occur at all. Indefinites are only found in postverbal position.
2. The generic interpretation is even more simple: all four languages use only Bare NPs to express genericity.
3. The way to express the definite reading is more complicated. In Mandarin and Min, definiteness is expressed with Bare NPs only. Cantonese employs Cl-NPs to express definiteness. Wenzhou does both, but in the definite Cl-NPs the tone of the classifier is affected.

Later we discuss the form, interpretation, and distribution of the different nominal expressions in the four languages discussed here. We will focus on definite and indefinite phrases, because they show interesting variation.⁷ Before investigating the variation, however, we want to take a closer look at the nature and use of classifiers.

3 CLASSIFIERS

3.1 Classifiers and the Functional Domain of N

3.1.1 *The Function of Classifiers*

So far we have been using the term “classifier” loosely, as a cover term for two types of elements, traditionally referred to as “classifiers” and “measure words” and dubbed “count-classifiers” and “mass-classifiers” in Cheng and Sybesma (1998) (“count-classifiers” and “mass-classifiers” were subsequently abbreviated to “classifiers” and “massifiers,” respectively). The difference between the two elements is the following (for some discussion, see Croft 1994 and Tai and Wang 1990). Measure words are used to make masses countable; this is the case in Chinese as well as in other languages:

- (25) a. a glass of water
 b. een glas water (Dutch)
 a glass water
 c. yat bui seoi (Cantonese)
 one cup water

As these examples show, if you want to count a mass like water, you need a counter, a unit to count it with. Measure words make masses countable by creating a unit by which they can be counted. These units have to be created because masses, by their very nature, do not come naturally in units by which they can be counted. WATER does not naturally come in countable units, so we force it into glasses or bottles (or liters or barrels).

Aside from measure words, languages like Chinese have classifiers. In such languages, count nouns look like mass nouns in that they, too, need a counter when you want to count; the examples in (26) show that counting without a counter, in this case the classifier for books, leads to ungrammaticality:

- (26) a. yi *(ben) shu (Mandarin)
 b. i⁷ *(paŋ³) si¹ (Wenzhou)
 c. jit *(bun) zhu (Southern Min)
 d. yat *(bun) syu (Cantonese)
 one CL^{volume} book

The difference with measure words is that classifiers do not *create* any unit to count by. Unlike mass nouns, count nouns have a built-in semantic partitioning; they come in naturally countable units. Classifiers, then, simply *name* the unit that the semantic representation of the noun naturally provides.

In other words, just like other languages, Chinese languages have count nouns and mass nouns in the sense that they have nouns whose semantic representation does not have a built-in partitioning in natural units and nouns whose semantic representation does have such partitioning. Languages like Chinese, however, need a counter for both noun categories; it simply is a property of these languages that they cannot count anything without the intervention of a counter. In the case of mass nouns, these counters (the measure words, or mass-classifiers) create their unit of counting, in the case of count nouns the counters (the classifiers, or count-classifiers) simply name the unit that the semantic representation of the noun provides.

They are called “classifier” because different nouns have different count-classifiers, depending on the shape or any other property of the individual units that come with the natural partitioning. In Mandarin, for instance, long, tall things like humans are counted with a classifier, *ge*, that goes back to a word meaning ‘bamboo’. In this sense they can be seen to classify.

In the following discussion, we will use the term “classifier” either as a cover term for measure words and classifiers or as a shorthand for count-classifier. Only when it really matters do we strictly distinguish between count-classifiers and mass-classifiers.

3.1.2 *Why Count-Classifiers?*

The question that comes up immediately is why we need count-classifiers. If the semantics of the nouns in question already involves a partitioning in natural units, why do we need elements to name them when we start counting?

Doetjes (1996) argues that, in general, for count nouns to be countable, the partitioning that is part of their semantic denotation must be (made) syntactically visible; numerals require the presence of a syntactic marker of countability. She further argues that this countability marker can be realized in different ways: in some languages, number morphology performs this function, while other languages use count-classifiers. In other words, count-classifiers and number morphology have the same function of serving as a syntactic countability marker.⁸

The idea that number and count-classifiers are associated is not new, as Doetjes points out. Greenberg (1963) notes that there is a tendency for languages without grammatical number to make use of count-classifiers. Interesting evidence also comes from language development. Ikoru (1994) describes the development of Kana, a Nigerian Cross River language, which lost its noun class-related number morphology and replaced it with a count-classifier system, while other languages of the same family kept the original system and did not develop count-classifiers. Peyraube (1997) suggests that the development of count-classifiers in Chinese is related to the loss of an element that may have been a plurality marker.

3.1.3 *Association with Number*

Aside from the considerations given previously, there are other indications that classifiers have something to do with number (for discussion, see Paris 1981, 1989). First of all, in the Wenzhou and Cantonese definite Cl-NP phrases, it is the classifier that determines whether we are dealing with plural or singular.

- (27) a. bun syu (Cantonese)
 CL^v book
 ‘the book(*s)’

- b. di syu
 CL^{PL} book
 ‘the book*(s)’

- (28) a. paŋ⁷ si¹ (Wenzhou)
 CL^v book
 ‘the book’

- b. liɛ⁷ si¹
 CL^{PL} book
 ‘the book*(s)’

All Chinese languages have a general classifier for the unspecified plural; the same form is used for all nouns—count-nouns, and mass-nouns alike. In Cantonese it is *di*, in Wenzhou *liɛʔ*, in Mandarin *xie*. A “canonical” classifier like Cantonese *bun* ‘volume’ in (27a) signals singular, whereas the plural classifier *di* in (27b) marks the whole phrase as plural. The Wenzhou examples in (28) illustrate the same thing. Mandarin can be shown to work in exactly the same way, with indefinite CL-NP phrases:

- (29) a. Wo xiang mai ben shu. (Mandarin)
 I want buy CL^{volume} book
 ‘I would like to buy a book.’
 b. Wo xiang mai xie shu.
 I want buy CL^{volume} book
 ‘I would like to buy some books.’

The only difference between these sentences lies in the choice of the classifier. These facts strongly suggest that the classifier is the locus of Number in Chinese.

Second, Iljic (1994) turns the argument around by showing that, in any case, the noun itself is unmarked for number. The following example (somewhat adapted) shows that the phrase *zhe xin* ‘DEM letter/s’ can be referred back to with a singular, as well as a plural, noun phrase:

- (30) a. Ni zhe xin dei cheng yi-xia ...
 your this/ese letter/s must weigh a-bit
 ‘This/These letter(s) of yours must be weighed ...’
 b. ... ta chao-zhong-le / liang-feng dou chao-zhong-le.
 it overweight / two-CL all overweight-PRF
 ... it is/they are both overweight.’

If one considers universal quantification as a kind of pluralization, reduplication of the classifier may be seen as constituting a third indication of the connection between number and the classifier, since reduplication of the classifier yields a universal quantification reading (Paris 1981: 69). The flexibility in this respect is not the same in all Chinese languages, but they all allow for classifier reduplication at least to a very basic degree. Here are examples, one from Mandarin and one from Cantonese:

- (31) a. ge-ge xuesheng (Mandarin)
 CL-CL student
 ‘every student’

- b. zek-zek gau (Cantonese)
 CL-CL dog
 ‘every dog’

In short, the classifier is the locus for grammatical number in Chinese.⁹

3.1.4 *Individuation*

Closely related to the countability function is the “individualizing function” (Croft 1994: 162) of classifiers, a function that is also noticed by Iljic (1994) and Paris (1981), among others. As we just saw in (27) and (28), the canonical classifiers are always singular.

We may rephrase this and state that the classifier singles out one entity from the plurality of entities provided by the semantic representation of the noun in the lexicon; it picks out one instance of what is denoted by N. This is also represented in the Chinese way of saying ‘three books’: it uses the individuating classifier as shown in (32), not the plural one, as the ungrammaticality of the phrases in (33) asserts:

- (32) san ben shu (Mandarin)
 saN bun zhu (Southern Min)
 saam bun syu (Cantonese)
 three CL^{volum} book

- (33) *san xie shu (Mandarin)
 *saN se zhu (Southern Min)
 *saam di syu (Cantonese)
 three CL^{PL} book

Thus, the literal translation of ‘three books’ in Chinese is ‘three singular units / instances of book’.

The relevance of this discussion will become clear shortly, when we discuss more generally the functional superstructure of the noun in the following section.

3.2 The Functional Superstructure of N

3.2.1 *What It Does*

The discussion in the previous sections is limited to classifiers in Chinese. In this section, we relate this discussion to the issue of the functional superstructure of the noun more generally, especially because we assume that classifiers and D serve similar functions.

Several notions come up in general discussions on the functional domain of the noun phrase. Two of these notions, which seem to be expressed in the nominal domain in all languages of the world, are definiteness and number. All languages have ways of discriminating between definite and indefinite reference. Likewise, they all have ways of distinguishing singular from plural.

Before going into definiteness and number, we prefer to change the perspective and concretely look at what basic functions D, the head of the functional category DP, has been associated with. We single out two. First, it has been associated with what one could call a “subordinating” function (Szabolcsi 1987, 1994; Abney 1987; Stowell 1989). In semantic terms, NPs are predicates, which can be turned into arguments by D (type-shifting; Partee 1987). It is only thanks to being embedded in DP that an NP can be used as an argument at all. Second, Longobardi (1994) argues that the individuating function we associated with the classifier in Chinese is also a typical D-function. He states (1994: 634) that D has the ability to pick out a single instance of whatever is described by N (see also Higginbotham 1985).

We think that these two functions which D is supposed to perform (individuation, syntactic subordination) are closely related to, or even different manifestations of, a more fundamental property of the DP domain: its deictic property—the property to be able to refer at all.

Generally speaking, in language, there is a division of labor between the lexical domain and the functional domain. The division of labor can be summarized as “lexical units describe, functional units refer.” It can also be described as “lexical units refer to a concept, functional units refer to actual instantiations of that concept in the real world.” The lexical entity *dog*, for instance, “describes” in the sense that it refers to a concept—“dogness.” It is only thanks to functional elements like *a* and *the* that we can use lexical items like *dog* to refer to actual instantiations of “dogness” in the real world.

The same division of labor is found in the verbal domain. *Bake cookies* describes a certain type of event. Embedding it in a deictic category like Tense Phrase (TP) enables one to use it to refer to an actual event in the real world that can be described as “bake cookies”: (*John*) *baked* cookies; T explicitly links the lexical phrase *bake cookies* with a specific event associated to the time axis of the real world.

This division of labor between functional and lexical categories is such a fundamental property of language that it must be part of Universal Grammar: all languages have it.

From the literature it may be deduced that this deictic function is a function mainly associated with D. D individuates and, possibly, by doing so, subordinates. The question that arises is this: If this division of labor is such a fundamental property of language, and the deictic function is performed by D, what is going on in languages that lack D?

Before answering this question, we take a look at the nominal phrase in English and French.

3.2.2 *The D in English and French*

Consider the following phrases from English and French. In all these phrases, the determiners (*the*, *a*, and $\emptyset$ in English; *le*, *les*, *un*, and *des* in French) presumably perform the deictic function. What else do they do?

- (34) a. the boy, the boys
 b. a boy, boys
- (35) a. le garçon, les garçons
 them:SG boy thePL boys
 b. un garçon, des garçons
 am:SG boy aPL boys

In English, the determiner *the* signals definiteness. It does so in both plural and singular phrases. In other words, *the* does not give us any information with respect to Number. In English definites, Number is marked on the N. *A* signals indefiniteness, but also singularity. The plural indefinite does not have an overt determiner in English.

In French, the determiner is responsible for expressing both number (Delfitto and Schroten 1991) and definiteness. *Garçon* and *garçons* sound exactly the same: whether we are dealing with plural or singular, definiteness or indefiniteness is marked by the determiner. The singularizing function apparent from *a* in English seems to be part of *le*'s function in French as well.¹⁰

When we look at (36), we see that *le* does even more:

- (36) a. le garçon
 them:SG boy
 b. la fille
 theF:SG girl

As Croft (1994) notes, grammatical gender must be seen as a kind of noun classification, and in languages such as French, gender is expressed through the determiner.

In short, D in French expresses definiteness, number, and noun classification. Besides, it has the deictic function we mentioned earlier; it individuates in the sense developed previously and makes it possible for an NP to be used as an argument.

In English, D performs the deictic function, but it only expresses (in)definiteness; it does not express Number, and noun classification does not play a role in English. Number in English is expressed on the noun.

3.2.3 *Cl in Cantonese*

Earlier we posed the question: If Universal Grammar incorporates a describing-referring dichotomy and D takes care of the referring deictic function in the nominal domain, what happens in languages that don't have determiners?

The answer is, of course, that some other functional head will perform that function. When we turn our attention to Cantonese against the backdrop of our discussion of English and French in the previous section, we find that in Chinese the classifier does many things that determiners do in other languages. Consider the basic Cantonese data in (37):

- (37) a. bun syu; di syu
 CL^{volume} book; CL^{PL} book
 'the/a book; the/ø books'
- b. yat-bun syu; syu
 one-CL^{volume} book; book
 'a book; books'
- c. go jan; zek gau
 CL person; CL dog
 'a/the person; a/the dog'

What do these examples tell us? First, the fact that these phrases "refer" means that some element performs the deictic function. Second, the fact that, as was shown in the example sentences in the first section, these phrases can be arguments of predicates indicates that some element functions as the subordinator. Third, these clusters of facts show very clearly that the classifier is the locus of number (see the a-examples) as well as classification (compare the classifiers used in the a-examples with the ones used in the c-examples); the individuation function was elaborated on previously. The definiteness-indefiniteness contrast is less straightforward, at least in the singular (we discuss this fact later); the plural shows a clear presence-absence contrast, as was the case in English.

In short, all the functions the determiner takes care of in French are performed by the classifier in Cantonese: definiteness, number, individuation, noun classification, subordination, and deictism.

This leads to the conclusion that the classifier is the head of a functional projection, CIP, which is in all relevant ways comparable to DP in languages like French and English. (For earlier proposals of CIP, see, for instance, Tang 1990.)

3.3 Conclusion of Classifiers

In this section we have investigated the function of the classifier, in its own right as well as to the background of the more general discussion on the function of the superstructure in the nominal domain. We found that there is one fundamental function to be performed by a functional head in the nominal domain: the deictic function. The lexical NP describes; the functional head deictically refers. We also found that this function is intricately related to the expression of number, definiteness, and even noun classification. It is also linked to the more purely syntactic function of making it possible for the lexical NP to be embedded in a predicate as an argument.

In languages with determiners, the deictic function is performed by D. In Chinese, at least in Cantonese, we concluded, it is taken care of by the classifier. The classifier heads its own functional projection, ClP. This leads to the postulation of the following structure (to be qualified later):



We will see that this is the structure we find in all Chinese languages discussed here. The question is, of course, how we account for the variation we found in section 2.

4 INTERPRETATION AND DISTRIBUTION

In section 2 we saw that both Bare NPs and Cl-NP phrases can be interpreted as either definite or indefinite. (As noted, the position in the sentence plays a role in determining the interpretation.) Among the languages we discuss in this chapter, Wenzhou is unique in allowing both interpretations for both. In Mandarin, Bare NPs can have both interpretations, while Cl-NPs are exclusively indefinite. In contrast, in Cantonese, Cl-NPs can have both interpretations, while Bare NPs are restricted to indefiniteness. In Min, which disallows Cl-NPs altogether, Bare NPs can get both readings.

In section 3 we preliminarily postulated the structure in (38) as the structure for nominal phrases in Chinese. Now in section 4 we investigate the question

what determines the variation we found: Why can some phrases be definite in one position but not in the other? Why can some types of NP be definite in the one language but not in the other?

4.1 Definites

4.1.1 *Cantonese, Mandarin, and Min*

Two different types of phrases can be interpreted as definite: CI-NPs and Bare NPs. Before we turn to Chinese, let us first determine how, in general, definiteness is supposed to arise.

Essentially following the literature, we assume that definiteness arises as the result of the insertion of a certain operator, called the ι operator, into D (Partee 1987). We further assume that the ι operator may be realized as a definite article, as is the case in languages such as English, or may be nonovert, for languages without a definite article. Insertion of the ι operator in D most likely subsequently triggers N-to-D-movement, overtly or nonovertly (see, e.g., Ritter 1989, Longobardi 1994); for enlightening discussion on these matters, see Chierchia (1998).

Let us now turn to Cantonese. In section 3 we reached the conclusion that CI has many properties in common with D—the singularizing, deictic function being one of them—and we subsequently likened CIP in Chinese-type languages to DP in English and French-type languages. We propose that definite CI-NPs in Cantonese have the structure in (38): the lexical NP is embedded in a functional structure, CIP, the head of which is filled by an ι operator realized as an overt classifier. The overt classifier has exactly the same function as determiners like English *the*: thanks to the ι operator, insertion in CI^0 makes the phrase definite.

As to Mandarin, we suggest that it takes the other option we mentioned: definite Bare NPs also have the structure in (38), but now the ι operator in the head of CIP is nonovert (presumably triggering covert N-to-CI movement).

This means that all definite nominals in Cantonese and Mandarin have the underlying structure in (38). Cantonese fills the CI^0 -position with an overt classifier, yielding definite CI-NPs. Mandarin derives definite Bare NPs by inserting the nonovert ι operator in CI^0 (followed by covert N-to-CI).

Southern Min, of course, is the same as Mandarin. Its definite Bare NPs are derived in exactly the same way: insertion of the empty ι operator in CI^0 , followed by covert N-to-CI.

4.1.2 *Definites in Wenzhou*

Among the languages we discuss here, Wenzhou is exceptional because it allows a definite interpretation for both Bare NPs and CI-NPs. The definite Bare NPs

should be unproblematic: their derivation will be the same as in Mandarin and Min: Cl^0 in their underlying structure (38) hosts an empty ι operator, yielding the definiteness.¹¹

The definite Cl -NP, in contrast, may be more problematic because of its extra feature, which we have been referring to as tonal neutralization. In definite Cl -NPs, all classifiers with tones belonging to the hi-register subcategory end up being pronounced as a hi-D-tone; all lo-register tones end up being pronounced as lo-D.

In actual fact, what we observe is a general lowering. If we describe the tones of Wenzhou in terms of a three-point scale (H,M,L) instead of a five-point scale, as was done in table 7.1, hi-A, hi-B, and hi-C can be characterized as H, and hi-D as M or L. At the same time, lo-A, lo-B, and lo-C can be characterized as M or L, while lo-D is L.¹² In other words, H goes to M or L; both M and L end up as L.

We propose that in Wenzhou, the ι operator, in addition to being nonovert, can also be realized as a low tone, a nonsegmental unit. When the low tone ι operator is inserted in Cl^0 , it must be supported by an element with a richer phonological (i.e., segmental) matrix, like a classifier. As a result of the fusion between the low tone ι operator and the classifier, the classifiers with a tone that is characterizable as H are lowered to M (or L), and the ones with an M or L tone are lowered to L.¹³ This proposal is in line with the principle that all tones must be borne by a full syllable (see the work by John Goldsmith—e.g., Goldsmith 1976), as well as the principle that morphemes must always be realized; otherwise we wouldn't know it is there (Lin 1993). In the case at hand, the ι operator in Cl^0 has the form of a low tone, and nothing else; as a consequence, in order for the definiteness to be expressed, the tone has to be realized—on the classifier. Interestingly, the original tone of the classifier is reflected in the register: hi-register tones come out as hi-D, lo-register tones as lo-D (except, of course, if the two D-tones are not distinguished; see note 12).

If this proposal is correct, it means that the ι operator can have a phonological reflex that is anywhere between zero and a lexical, fully segmental element like a determiner or a classifier: it can be realized as something nonsegmental as a tone.

Also, the proposal entails that the formation of definites in Wenzhou involves a certain degree of optionality: it can pick either of two manifestations of the ι operator (discussed further in section 5.2).

4.1.3 *Summary*

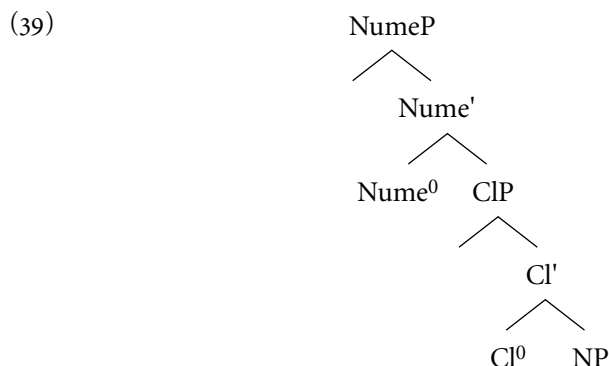
To conclude this section on definites, we have seen that all definite nominals in Chinese we discussed have (38) as their underlying structure. The derivation of definites involves the insertion of an ι operator into the head of ClP . The ι operator has different manifestations. It may be nonovert, as is the case in Mandarin and Min and is an option in Wenzhou; or it may be overt, in which case

it is either realized as a full lexical element, a classifier, as is the case in Cantonese, or it may be realized as a low tone, as we saw is an option in Wenzhou. In the latter case, the tone has to be provided with a segmental matrix, like a classifier, which has lowering effects on the tone of the classifier involved.

At least two questions arise: (a) What determines the difference between the different languages? For instance, is there any reason that, say, Mandarin cannot be like Cantonese? And where does the optionality in Wenzhou come from? (b) If Cl-NPs and Bare NPs have the underlying structure in (38), how can they also be interpreted as indefinite? We answer the first question in section 5, after having answered the second question in the next section.

4.2 Indefinites

The only type of phrase that is invariably indefinite in all languages dealt with in this study is the Num-Cl-NP phrase, the one with an overt numeral. In view of the discussion on definites in the previous section, we assume that the indefinite reading is due to the presence of the Numeral; we return to this later. Let us postulate the following structure for indefinite nominal phrases in Chinese, where NumeP stands for NumeralP (instead of NumP, to avoid confusion with NumberP).



Is there any evidence that (39) is the underlying structure for all different types of indefinite noun phrases, including Cl-NPs and Bare NPs? We think there is.

Longobardi (1994) develops a proposal with respect to the distribution and interpretation of bare nouns in Germanic and Romance. Distributionally, bare nouns in Germanic and Romance are limited to lexically governed positions (in effect, object positions) and interpretationally, they are restricted to an indefinite reading. Longobardi's proposal comes down to the following. Bare nouns are not

really bare: they are, of course, embedded in a full-fledged DP projection, the head of which is empty. The assumption of the empty D helps Longobardi explain both the restricted distribution and the limitations on the interpretation in the following way. First, as the empty D is just like any other empty category in having to be lexically governed, the distribution of bare nouns is limited to lexically governed positions. Second, Longobardi argues that an empty D is associated with an existential reading (see also Chierchia 1998).

Turning our attention to indefinite CI-NP phrases, if they have the same structure as Num-CI-NPs—that is, the structure in (39), in which the Numeral is empty—we predict that they are limited to lexically governed positions. As the facts in section 2 show, this is the case: CI-NPs can have an indefinite interpretation only in object position.¹⁴

For indefinite Bare NPs, the reasoning and the conclusion are the same. The fact that the Bare NPs with an indefinite reading occur only in lexically governed positions suggests that they involve an empty category that has to be governed. If it is the case that in Chinese indefiniteness arises as a result of the presence of a numeral, the Bare NP must have a numeral. In short, we assume that indefinite Bare NPs also have the structure in (39); the difference with indefinite CI-NPs is that in the case of Bare NPs, not only the Numeral head Nume^0 is empty but also the CI^0 is empty.¹⁵

In short, assuming the structure in (39) for indefinite CI-NPs and Bare NPs explains their limited distribution.

Comparing (39) with (38) we see that (39) actually consists of (38) embedded in a NumeP . Because the structure in (38) is the structure, that underlies all definite nominals in Chinese suggests that the indefinite interpretation arises as a result of the presence of the Numeral. The NumeP , being a quantificational expression, is interpreted as indefinite. “A book” in Chinese languages is literally “[one [the book]].” To get an indefinite reading, a NumeP is required.

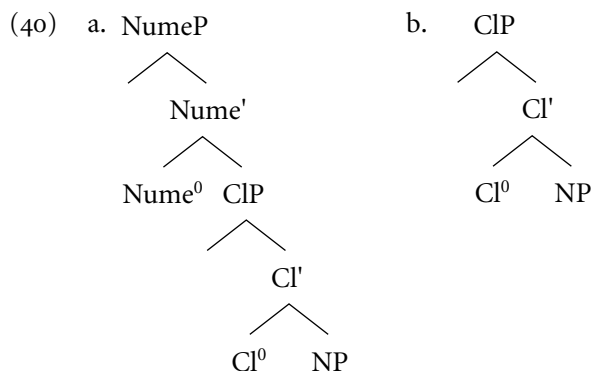
In this respect, there seems to be a difference between languages with articles and those without. As was suggested by Teun Hoekstra (class lectures, fall 1996), nominals in Germanic and Romance languages are quantificational expressions—that is, indefinite—unless they are embedded in a DP. Thus, “the book” in English is in fact “[the [(a) book]].” Articled languages have indefinite articles (or, in their absence, the numeral ONE) to pick out singularities, while article-less languages have only classifiers for that purpose. Classifiers, however, are very similar to definite determiners, as we saw earlier.

Before closing off this section, we would like to mention two things. First, definite CI-NPs and definite Bare NPs do not involve an empty category: in case of the definite CI-NP, there is no numeral ((38) does not involve a NumeP); in the case of the Bare NP, aside from there being no NumeP , CI^0 is occupied by the ι operator and N^0 as the result of N-to-CI. Since they do not involve an empty category, they should not be restricted distributionally, and they aren’t.

Second, in section 2 we noted that indefinite nouns with an overt numeral can be interpreted as both specific and nonspecific, while indefinite Bare NPs and CI-NPs can only be nonspecific. In Cheng and Sybesma (1999: 530–532), we supply ample evidence that indefinite CI-NPs involve an empty numeral and that they, and indefinite Bare NPs, are excluded from contexts in which a specific interpretation is the only possible option for an indefinite; we have seen some such contexts before, such as (3b). The fact that indefinite CI-NPs and indefinite Bare NPs may only be nonspecific, while full-fledged Num-CI-NPs may be both specific and nonspecific, is likely to be related to the empty Numeral in the former: an empty Numeral leads to a nonspecific interpretation. A possible explanation for the difference between an empty Numeral and an overt Numeral is that the latter is a full-fledged quantifier and as such can undergo quantifier raising (QR), yielding a specific reading. In contrast, an empty Numeral lacks the QR option. Instead, it relies on the presence of existential closure to supply the existential quantification, yielding a narrow scope nonspecific reading (cf. Diesing 1992).

4.3 Summary

These are the structures postulated for definite and indefinite nominal phrases in Chinese: all indefinites (Num-CI-NPs, CI-NPs, Bare NPs) have the structure in (40a) (= (39)); all definite structures (CI-NPs, Bare NPs) have the structure in (40b) (= (38)). In the indefinite case of (40a), Nume^0 and CI^0 may be left empty; in the definite (40b), CI^0 is filled by the ι operator (possibly realized as an overt classifier) and the covertly moved N^0 :



5 HOW TO ACCOUNT FOR THE VARIATION

The four varieties of Chinese dealt with here differ from one another in the following ways:

1. Mandarin and Min do not have definite CI-NPs, they have definite Bare NPs; for Cantonese, exactly the opposite is true.
2. Mandarin has indefinite CI-NPs, like Cantonese and Wenzhou; Min does not: it is the only variety with no CI-NPs at all.
3. Cantonese and Wenzhou are the only varieties with definite CI-NPs, but they differ in that in Wenzhou the tone of the classifier is affected.
4. Wenzhou differs from all others in that it allows a definite reading for both CI-NPs and Bare NPs; all other languages allow a definite reading for just one of the two.

5.1 Mandarin and Min versus Cantonese

We saw that the derivation of definite nominals involves the ι operator, the difference between the languages discussed in this chapter resulting from the different manifestations the ι operator may take on. Disregarding Wenzhou for the moment, we see that definite nominals are either CI-NPs (as in Cantonese) or Bare NPs (as in Mandarin and Min): Cantonese involves the insertion of an overt classifier, just as English inserts its overt determiner, while Mandarin and Min have no overt reflex of the ι operator.

From a general point of view, it seems that the option favored by Cantonese is the preferred one (the default). That is to say, in his discussion on noun phrases in English and other languages, Chierchia (1998) suggests that, as far as definiteness is concerned, expressing it overtly is favored over doing it covertly. In other words, if, in principle, there is a choice, languages opt for expressing definiteness overtly, using a definite article (or, we may add now, a classifier). Since the Cantonese option involves insertion of a classifier in CI⁰, this embodies the preferred procedure. This explains why in Cantonese this is the only option. At the same time, it raises the question what is wrong with Mandarin and Min: What bars them from opting for the preferred choice?

In this context we would like to point at two facts. First, in Min, it is impossible to have an overt classifier without a numeral: as we have mentioned several times, Min has no CI-NPs at all. Second, in Mandarin, CI-NPs are always indefinite. If our assumption that indefinites have the underlying structure in (40a) is correct, this means that Mandarin overt classifiers also cannot occur

without a numeral, the difference between Mandarin and Min being that Min cannot have empty numerals.

Let us phrase this in the form of a restriction on the appearance of overt classifiers in these varieties of Chinese:

- (41) In Mandarin and Min, overt classifiers are always accompanied by a numeral. (In Min, the numeral must be overt; in Mandarin it may be overt or nonovert.)

Cantonese does not have this property. Although we have no deeper explanation for this difference between Cantonese on the one hand and Mandarin and Min on the other, the restriction explains why it is impossible for Mandarin and Min to get definite nominals in the preferred way. After all, the preferred way involves the insertion of an overt classifier in the head of CIP. For Mandarin and Min, overt classifiers always come with a numeral, and numerals lead to indefiniteness.

5.2 Optionality in Wenzhou

It is not clear how Wenzhou, having both definite Bare NPs (like Mandarin and Min) and definite CI-NPs (like Cantonese), fits into this picture. Since it does not have the restriction in (41) (as is clear from the data in section 2.2), there is no reason why it should not be like Cantonese.

To be sure, Wenzhou does have definite CI-NPs, but the tone of the classifier is lowered. Before, we claimed that this is the result of a fusion between a non-segmental, strictly tonal ι operator and the classifier. In other words: we analyzed definite CI-NPs in Cantonese and Wenzhou as entirely different. In definite CI-NPs in Cantonese, the CI is the classifier in its full form—a segmental ι operator. In the Wenzhou definite CI-NPs, in contrast, the classifier is there only to provide the nonsegmental ι operator with segmental material.

In section 4.1.2, we suggested that the ι operator can have a phonological manifestation that is anywhere between zero and a fully segmental lexical element like a determiner or a classifier; after all, it can also manifest itself as something nonsegmental, (a low tone, for instance). Let us be less liberal and state that there are three options: segmental, nonsegmental, and empty.

In view of the facts of Cantonese and Wenzhou, we may conclude that only the segmental manifestation counts as fully overt for the purposes of expressing definiteness as discussed previously in the context of Chierchia's hypothesis that definiteness is preferably expressed overtly. Syntactically, the nonsegmental tone does not count as "overt."

This means that the optionality in expressing definiteness displayed by Wenzhou does not involve the preferred, overt option; it is between two nonovert options. That is why it can have optionality at all. As Chierchia states, *if* there is a choice, and the overt option is one of the alternatives, then that is, really, the only option, the default. Then there can be no optionality.¹⁶

The answer why Wenzhou is not like Cantonese, then, comes down to the fact that the choice it has for expressing definiteness does not involve the overt option. We don't know why this is the case.

5.3 Parameters Discussed in This Chapter

The first parameter to be mentioned is the one that distinguishes “languages with articles” from “languages with classifiers.” Several differences between these two types of languages seem to fall out from one main difference—the fact discussed in section 4.2 that whereas in article languages definites are indefinites-turned-definites, languages with classifiers derive indefinites on the basis of definites. This difference is the consequence of the latter type of language having no indefinite article.

Within the domain of Chinese, the variation may be reducible to the following parameters. The first is whether or not definiteness is expressed by a segmental ι operator in Cl^0 —that is, in the form of a full-fledged classifier—which would set Cantonese apart from the other varieties. A second is whether or not it is possible to have an overt classifier without a numeral (setting apart Mandarin and Min from Wenzhou and Cantonese). Third, Mandarin, Wenzhou, and Cantonese can have empty numerals, whereas Min cannot. The parameters can be charted out as in table 7.3. The claim that the nonovert ι operators involve a (phonologically speaking) truly empty one and one in the manifestation of a low tone is not considered in the table.

It is conceivable that some of these parameters are related to the status of the classifier at a deeper level. As such, they may possibly be connected to phenomena other than the ones discussed in this chapter. There may be a deeper difference, distinguishing “weak classifiers” from “strong classifiers.” Aside from the question as to whether they can occur with or without a numeral, we should consider the freedom of reduplication of the classifier (which seems, at a first impression, freer in Cantonese and Wenzhou than in Mandarin and Min), as well as the question whether they are used as the element that links the modifier to its noun (which is the case in Cantonese and some varieties of Wu, but not in Mandarin and Min). More research in this area is necessary to see whether there is such a “deep” parameter.

Table 7.3 Parameters for variation within Chinese

	Classifier or article	Parameter		
		ι operator realized segmentally	Overt classifier obligatory with numeral	Can numeral be empty?
Mandarin	Classifier	No	Yes	Yes
Wenzhou	Classifier	No	No	Yes
Southern Min	Classifier	No	Yes	No
Cantonese	Classifier	Yes	No	Yes

NOTES

This chapter incorporates most of the arguments and conclusions in Cheng and Sybesma (1999). In gathering and sorting the data, we received help and advice from many people, which we gladly and gratefully acknowledge. In particular we acknowledge the input from the following colleagues:

Wu: We thank You Rujie for sharing with us his Wu expertise in general and for providing and discussing the Wenzhou data. Thanks a lot! For Shanghainese we are especially indebted to Hua Dongfan. We also discussed Shanghainese with You Rujie, Eric Zee, Lu Bingfu, and Duanmu San.

Min: The data were provided (and checked with other informants) by C.-C. Jane Tang. Thanks are also due to Tang Ting-chi.

Cantonese: Thomas Lee and Sze-Wing Tang helpfully provided judgments.

We furthermore thank the following colleagues for more general comments and discussion: Thomas Lee, Jane Tang, Dylan Tsai, and especially Richard Kayne, Guglielmo Cinque, and Moira Yip.

Sybesma's research was partly supported by P. Muysken's Spinoza Research Program "Lexicon and syntax" and partly by his own "Vernieuwingsimpuls"-project on syntactic variation in southern China, cofunded by the Dutch Organization for Scientific Research NWO, Universiteit Leiden (main sponsors), and the International Institute for Asian Studies (IIAS).

1. The three varieties of Chinese not mentioned are Hakka, spoken in the area where the provinces of Jiangxi, Guangdong, and Fujian border; Gan, spoken in Jiangxi; and Xiang, the language of Hunan.

2. We use the following abbreviations in the glosses: CL = classifier; PRF = perfectivity/boundedness marker; SFP = sentence-final particle. In case it seems useful to provide the meaning of the classifier, it is added in superscript. Mandarin is transcribed using the pinyin system. For Cantonese we used the Romanization system developed by

the Linguistic Society of Hong Kong. Transcribing Wenzhou, we use the IPA. The Min transcription was provided by Jane Tang (for Min, there does not seem to be a standard); the capital N stands for nasalization. Tones have been left unmarked except in the Wenzhou data, where they are marked using superscript numbers.

3. Without the help of You Rujie we would not have been able to present a complete set of Wenzhou data; especially the presentation of the tone facts would have been much more sketchy.

4. For discussion of the Mandarin counterpart, *ba*, see Sybesma (1999).

5. The situation in Shanghainese is not as clearcut as in Wenzhou. You Rujie (pers. comm.) confirms that the Shanghainese language situation is unstable, or at least that a lot of variation exists among speakers. Qian (1997: 98–99) claims that Shanghai is like Wenzhou the way we described it in the text, in that Cl-NPs can be interpreted as definite, in which case the classifier undergoes a tone change (the precise mechanics of the tonal changes being different from what happens in Wenzhou). Qian says that the younger generation uses it rarely. One of our Shanghainese informants agrees that Cl-NP can be used with a definite reading but is not aware of any tonal change. This is basically the picture sketched by Pan (1991). Three of our informants do not accept Cl-NPs in a definite reading; in an indefinite reading, they are judged fine. The Shanghainese of these speakers is exactly like Mandarin.

6. It seems that the younger generation of Taiwan's Min speakers who are fluent in Mandarin as well do not reject Cl-NPs in Min entirely, in which case they get a nonspecific indefinite interpretation; this may be due to influence from Mandarin.

7. We have included the generic interpretation of noun phrases in this description for the sake of completeness only. Mainly because the languages discussed do not show any variation here (as well as because of space limitations), we do not discuss the generic interpretation of Bare NPs in this chapter. For the same reasons, we leave the derivation of the proper name interpretation of common nouns undiscussed here. For some discussion of generics in Chinese, see Cheng and Sybesma (1999: 532–534) and references cited therein; for proper names, see Cheng and Sybesma (1999: 523–524).

8. We are aware of the fact that languages such as Finnish and Hungarian do not have plural markings when combined with a numeral. However, these languages would still count as “Number” languages. The absence of the plural marking may be a matter of agreement.

9. As a consequence, *men* cannot be a plurality marker—contra Li (1999). We disagree with Li for many reasons, one being that her approach does not explain the lack of productivity of *men*; another reason is that it does no justice to the meaning of *men*, which is much more a collective marker than a plural marker. See Iljic (1994) and Cheng and Sybesma (1999) for discussion.

10. In some cases French nouns do make a plural-singular distinction (*cheval*, *chevaux*), and in combination with the preposition *à* the determiner loses its number function: *au* and *aux*. Similarly, nouns may express gender, and the gender function of the determiner may be neutralized; both are illustrated by *l'étudiant* and *l'étudiante*. These exceptions do not undermine the system of French as described in the text.

11. We thank Moira Yip for discussion of the phonological side of the discussion in this section.

12. Chen (2000: 476–477) reports that in his informant's speech, the two D-tones are not distinguished. In Chen's tone system of Wenzhou, the D tone is characterized as

L. (Recall that Chen works with partly different values; see the comment under table 7.1 in the text.)

13. The lowering rule apparently does not apply to the D-tones themselves.

14. Indefinites with a numeral can also not occur in subject position; this is due to reasons related to existential closure: see Diesing (1992), Cheng (1991), and Tsai (1994), and, for some discussion, Cheng and Sybesma (1999).

15. We assume that movement of N to Cl (or Nume, for that matter) would only take place if it is triggered.

16. The restriction in (41) bars Mandarin and Min opting for both the segmental and the nonsegmental-only ι operators. At the same time, the reasoning in the text leaves open the possibility that they pick the empty one for reasons not related to (41). After all, we do not have a clear idea as to what factors have a role in determining the choice for any of the three ι operators.

REFERENCES

- Abney, Stephen (1987) "The English Noun Phrase in Its Sentential Aspect." Ph.D. diss., MIT, Cambridge, Mass.
- Chen, Chuimin (1958) "Minnanhua he Putonghua changyong liangci de bijiao [A comparison of frequently used classifiers in Southern-Min and Standard Mandarin]." *Zhongguo Yuwen* 1958/12: 591–593.
- Chen, Matthew Y. (2000) *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge University Press, Cambridge.
- Cheng, Lisa L.-S. 1991. On the Typology of wh-Questions. Ph.D. diss., MIT, Cambridge, Mass.
- Cheng, Lisa L.-S., and Rint Sybesma (1998) "Yi-wan tang, yi-ge Tang: Classifiers and massifiers." *Tsing-Hua Journal of Chinese Studies*. New Series 28/3: 385–412.
- Cheng, Lisa L.-S., and Rint Sybesma (1999) "Bare and Not-So-Bare Nouns and the Structure of NP." *Linguistic Inquiry* 30/4: 509–542.
- Chierchia, Gennaro (1998) "Reference to Kinds across Languages." *Natural Language Semantics* 6: 339–405.
- Croft, William (1994) "Semantic Universals in Classifier Systems." *Word* 45/2: 145–171.
- Delfitto, Denis, and Jan Schrotten (1991) Bare Plurals and the Number Affix in DP." *Probus* 3: 155–185.
- Diesing, Molly (1992) *Indefinites*. MIT Press, Cambridge, Mass.
- Doetjes, Jenny (1996) Mass and count: syntax or semantics? *Meaning on the HIL*. [Occasional Papers in Linguistics.] HIL/Leiden University.
- Goldsmith, John (1976) "Autosegmental Phonology." Ph.D. diss., MIT.
- Greenberg, Joseph (1963) *Universals of Language*. MIT Press, Cambridge, Mass.
- Higginbotham, James (1985) "On Semantics." *Linguistic Inquiry* 16: 547–593.
- Ikoro, Suanu (1994) "Numeral Classifiers in Kana." *Journal of African Languages and Linguistics* 15: 7–28.
- Iljic, Robert (1994) "Quantification in Mandarin Chinese: Two Markers of Plurality." *Linguistics* 32: 91–116.

- Li, Y.-H. Audrey (1999) "Plurality in a Classifier Language." *Journal of East-Asian Linguistics* 8: 75–99.
- Lin, Yen-hwei (1993) "Degenerate Affixes and Templatic Constraints: Rhyme Change in Chinese." *Language* 69/4: 649–682.
- Longobardi, Giuseppe (1994) "Reference and Proper Names." *Linguistic Inquiry* 25: 609–666.
- Norman, Jerry (1988) *Chinese*. Cambridge University Press, Cambridge.
- Pan, Wuyun (1991) "An Introduction to the Wu Dialects." William S.-Y. Wang (ed.) *Languages and dialects of China*. Journal of Chinese Linguistics Monograph Series 3: 237–293.
- Paris, Marie-Claude (1981) *Problèmes de syntaxe et de sémantique en linguistique chinoise*. Collège de France, Paris.
- Paris, Marie-Claude (1989) *Linguistique générale et linguistique chinoise: Quelques exemples d'argumentation*. UFR, Université Paris 7, Paris.
- Partee, Barbara (1987) "Noun Phrase Interpretation and Type-Shifting Principles," in J. Groenendijk and M. Stokhof (eds.) *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*. Foris, Dordrecht, 115–143.
- Peyraube, Alain (1997) "On the History of Classifiers in Archaic and Medieval Chinese," B. T'sou, A. Peyraube, and L. Xu, (eds.) *Studia Linguistica Sinica*. Hong Kong City University, Hong Kong.
- Qian, Nairong (1997) *Shanghaihua yufa* [Shanghainese grammar]. Renmin, Shanghai.
- Ramsey, S. Robert (1987) *The Languages of China*. Princeton University Press, Princeton.
- Ritter, Elizabeth (1989) "A Head-Movement Approach to Construct-State Noun Phrases." *Linguistics* 26: 909–929.
- Stowell, Tim (1989) "Subjects, Specifiers, and X-bar Theory." in M. Baltin and A. Kroch (eds.) *Alternative Conceptions of Phrase Structure*. University of Chicago Press, Chicago.
- Sybesma, Rint (1992) "Causatives and accomplishments. The Case of Chinese *ba*." Ph.D. diss., HIL/Leiden University.
- Sybesma, Rint (1999) *The Mandarin VP*. Kluwer, Dordrecht.
- Szabolsci, Anna (1987) "Functional Categories in the Noun Phrase," I. Kenesei (ed.) *Approaches to Hungarian* 2, JATE, Szeged, 167–190.
- Szabolsci, Anna (1994) "The Noun Phrase," in the Ferenc Kiefer and Katalin Kiss (eds.) *Syntactic Structure of Hungarian*, Syntax and semantics, 27. Academic Press, New York, 179–274.
- Tai, J. and L. Q. Wang. (1990) A Semantic Study of the Classifier *tiao*. *Journal of the Chinese Language Teachers Association* 25/1: 35–56.
- Tang, C.-C. Jane (1990) "Chinese Phrase Structure and the Extended X-theory." Ph.D. diss., Cornell University.
- Tsai, Dylan W.-T. (1994) "On Economizing the Theory of A-bar Dependencies." Ph.D. diss., MIT, Cambridge, Mass.
- You, Rujie (2000) *Wenzhou fangyan yufa gaiyao* [Outline of Wenzhou dialect grammar]. *You Rujie zixuanji* [Selected writings of You Rujie]. Anhui Jiaoyu, Hefei.
- Zhou, Changji (1991) *Minnanhua yu Putonghua* [Southern Min and Standard Mandarin]. Yuwen, Peking.