

# John Benjamins Publishing Company



This is a contribution from *New Explorations in Chinese Theoretical Syntax. Studies in honor of Yen-Hui Audrey Li.*

Edited by Andrew Simpson.

© 2022. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: [www.copyright.com](http://www.copyright.com)).

Please contact [rights@benjamins.nl](mailto:rights@benjamins.nl) or consult our website: [www.benjamins.com](http://www.benjamins.com)

Tables of Contents, abstracts and guidelines are available at [www.benjamins.com](http://www.benjamins.com)

# Constraints on the representation of anaphoric definiteness in Mandarin Chinese

## A reassessment

Andrew Simpson and Zoe Wu

University of Southern California / Pasadena City College

There is increasing evidence that languages may use distinct forms to encode definiteness in instances of anaphoric and non-anaphoric definiteness, for example, distinct determiners, or bare classifier vs. bare noun patterns. This chapter considers a recent, prominent claim in Jenks (2018) that such distinctions are systematically encoded in Mandarin Chinese via the presence vs. absence of demonstratives with a noun, and regulated by a principle *Index!*, which requires the use of overt marking of referential indices. The chapter argues that there is greater optionality in the forms used to represent definite NPs in Mandarin than assumed in Jenks' characterization, and that alternations between definite bare nouns and demonstrative-marked nouns are affected by issues of discourse coherence and parsing expectations.

### 1. Introduction

Much interest has recently been focused on the ways that languages may represent definite reference in different ways, depending on the nature of definiteness that is highlighted in specific contexts. In contrast to English, French and Spanish, which employ a single determiner for all typical instances of definite reference, languages such as Fering (Schwarz 2009, 2013) have been noted to have two distinct definite determiners, and these are used in different discourse situations. Fering's 'D-Article' (*di, det, dön*) is used in instances of anaphoric reference and when a referent is visible, while its 'A-article' (*a, at*) occurs with referents which are unique in a particular situation or a broader context, also with generic noun phrases, and in instances of bridging reference (where an element is identified by means of a connection to some other discourse referent). Related patterns have been reported in other varieties of Germanic (standard and regional forms of German, Dutch, Scandinavian (Ebert 1971a/b, Schwarz 2013, Löbner 2011, Ortman 2014), Hausa (Afroasiatic; Jaggarr

1985), Akan (Niger-Congo, Arkoh, and Matthewson 2013) and Lakhota (Siouan; Lyons 1999), as well as other languages. A broad generalization which emerges from these works is that one mode of representation is typically used in instances of anaphoric reference, when an element refers back to another individual/object previously introduced into the discourse – ‘anaphoric definites’ – while a second representational form generally occurs in instances of non-anaphoric definiteness when a noun phrase refers to a unique individual in a particular situation/context – ‘unique (non-anaphoric) definites’. In Chinese, such distinctions have also been noted to occur, and speakers of varieties of Wu Chinese have been found to selectively use either bare nouns or nouns combined with a classifier in ways that mirror the unique vs. anaphoric division in languages with definite determiners (Li and Bisang 2012, Simpson 2017).

Extending the scope of such studies in Chinese, a recent prominent work focused heavily on Mandarin, Jenks (2018), has suggested that a parallel division in the representation of definiteness may be performed by the presence vs. absence of *demonstratives* with a noun. Jenks argues that anaphoric definites in Mandarin require the presence of a demonstrative as the result of a principle *Index!* which necessitates that referential indices are projected in anaphoric relations. Unique (non-anaphoric) definites, by way of contrast, are noted to occur as bare nouns, not requiring any referential index. Such differences are illustrated in (1) and (2). In example (1), the anaphoric linking of *nansheng* ‘male student’ to its antecedent in the preceding sentence requires the use of a demonstrative, whereas the interpretation of *shichang* ‘market’ as a unique non-anaphoric definite is encoded without a demonstrative, by means of a bare noun.

- (1) jiaoshi li zuo-zhe yi ge nansheng he yi ge nüsheng.  
 classroom inside sit-PROG one CL boy and one CL girl  
 ‘There are a boy and a girl sitting in the classroom.’  
 wo zuotian yudao #(na ge) nansheng.  
 I yesterday meet that CL boy  
 ‘I met the boy yesterday.’ (Jenks 2018: 510)
- (2) wo xianzai qu shichang mai dongxi.  
 I now go market buy thing  
 ‘I’m going to the market now to buy some things.’

Such a difference in the way that definiteness is represented in Mandarin Chinese is noted to exhibit a potentially confounding complication, however, relating to the syntactic position occupied by the definite noun phrase. Jenks suggests that ‘Mandarin shows a general requirement for demonstratives with anaphoric definite noun phrases with the exception of subject positions, which also allow a bare

noun.’ (Jenks 2018: 511). In such cases, it is proposed that the subjects are in fact topics and that the status of a noun phrase as a topic pragmatically neutralizes the effect of *Index!* The present chapter takes Jenks’ study of definiteness in Mandarin as its starting point and investigates further the issue of optionality in the way that anaphoric definite reference can be encoded in Mandarin – either by means of a demonstrative or with a bare noun. Considering a wider body of empirical patterns, it is noted that bare nouns may actually be used in instances of anaphoric definite reference in a much broader range of syntactic positions than assumed in Jenks’ characterization, which calls into question the status of the principle *Index!* as an across-the-board constraint in Mandarin Chinese and leads to a reassessment of the way Mandarin should typologically be categorized in terms of the strategies it employs to encode anaphoric definiteness. The chapter compares Mandarin with other languages and varieties which more rigidly require the overt representation of definiteness, such as English, French, and Cantonese, and argues that Mandarin, by comparison, is still at a much earlier stage of development in the grammaticalization of definiteness-marking in which genuine optionality is grammatically permitted in the encoding of NPs interpreted as anaphoric definites, although such optionality may be conditioned by certain, additional discourse factors.

The structure of the chapter is as follows. Section 2 introduces the critical oppositions examined in the chapter, and how two representational forms may be used in various languages to encode anaphoric and unique, non-anaphoric definiteness. Section 2 also presents Jenks’ Mandarin Chinese data set and his characterizations of this data, along with the key ideas of the analysis in Jenks (2018). Section 3 then adds a more extended set of patterns from Mandarin probing the ways that anaphoric noun phrases may be encoded in both argument and adjunct positions, and concludes that the use of bare nouns in such positions is in principal significantly more available than suggested by the data reviewed in Section 2, leading to questions about the generalizations drawn from Jenks’ more restricted data set. Section 4 compares Mandarin with three other languages/varieties which similarly lack definite determiners – Korean, Hindi, and Cantonese – and reports on similarities and differences in the representation of definite anaphoric noun phrases across these languages, with the goal of establishing how *Index!* may (or may not) apply as a cross-linguistic constraint. Section 5 summarizes what has been concluded from earlier sections and offers a different assessment of optionality attested in the representation of anaphoric definiteness.

## 2. Two ways to encode definiteness

Studies of languages which allow for definiteness to be represented in different ways have repeatedly observed that one form is regularly used in instances of anaphoric definite reference, and a second pattern occurs with situationally or globally unique definites, whose identity is not established by any discourse-anaphoric relation. The pioneering work in this area was carried out in Ebert (1971a, b) and Schwarz (2009), where Fering and German were noted to use different determiner forms to encode anaphoric and non-anaphoric unique definite NPs, as illustrated in the Fering examples (3) and (4) from Ebert (1971b, reproduced in Schwarz 2013)

- (3) Oki hee an hingst keeft. \*A/Di hingst haaltet.  
 Oki has a horse bought the/the horse limps  
*anaphoric definite reference*  
*'D-article' used*  
 'Oki has bought a horse. The horse limps.'
- (4) Ik skal deel tu a/\*di kuupmaan. *non-anaphoric unique definite*  
 I must down to the/the grocer *'A-article' used*  
 'I have to go down to the grocer.'

Following similar observations in other languages with overt determiners (Löbner 2011, Arkoh, and Matthewson 2013, Ortman 2014), work on various numeral classifier languages has established that languages which do not have definite articles may also make use of two distinct overt forms to represent anaphoric and non-anaphoric definite NPs. Simpson and Biswas (2016) document a definiteness split Bangla, where the combination of a classifier and noun (with no accompanying numeral) – a 'bare classifier pattern' – is used for anaphoric definite reference, as illustrated in (5), while a bare noun occurs in instances of non-anaphoric definiteness, as seen in (6).

- (5) kalke ram Ek Ta kalo Tupi ar Ek Ta Sada Tupi kinechhe. kalo  
 yesterday Ram 1 CL black hat and 1 CL white hat bought black  
 Tupi \*(Ta) or bORo hoechhe.  
 hat CL his big be  
 'Yesterday, Ram bought a black hat and a white hat. The black hat is too big for him.'
- (6) Context: Every day on planet Varg, a security vehicle delivers one new prisoner to guards from the camp. Today the vehicle drives up, the back doors open out and no-one is inside.  
 The guard says:  
 kOedi kothay?  
 prisoner where  
 'Where is the prisoner?'

In different varieties of Wu Chinese, both Li and Bisang (2012) and Simpson (2017) note that bare noun and classifier + noun patterns are used in the same way, the former for situational/globally-unique definites, and the latter for anaphoric definiteness, as shown in (7) and (8) from Simpson (2017) with Jinyun Chinese:

- (7) dzɿɔsɔ zōniɛi ji-dzai njiəm mɔ-də. guɽɛ gō dzai njiəm ha  
 Zhangsan yesterday 1-CL cow sell-ASP they say CL COW NEG  
 dɔ dziɛkō.  
 be healthy  
 ‘Zhangsan sold a cow yesterday. They say the cow was not healthy.’
- (8) Context: A firefighter at the scene of a car crash:  
 dzɿɔsiyɛ lei tɕieikɛi a? ηə niɔ-a-dzyo.  
 driver at where PRT I cannot-see  
 ‘Hey, where’s the driver? I can’t see him.’

These works make the point that the referential distinctions encoded via the selective use of determiners in Germanic, Akan and other languages may be reproduced by other means in languages where determiners have not developed, highlighting the fundamental nature of such distinctions. Simpson and Biswas (2016: 28) conclude their study of Bangla with the comment:

This demonstrates that a classifier language such as Bangla which has neither definite or indefinite determiners is interestingly able to adapt another aspect of its functional inventory (the presence/absence of a classifier with a noun) to achieve highly similar results in the representation of different components of definiteness, suggesting that such referential complexity is significantly shared across languages at an underlying level, and simply realized by different overt mechanisms.

(Simpson and Biswas 2016: 28)

Given such findings in (certain) languages where bare classifier patterns are present and used in frequent opposition to bare nouns to represent definite NPs, it would not be unexpected to attest the deployment of other elements to encode anaphoric vs. non-anaphoric definiteness in languages without determiners, and this is precisely what is suggested in Jenks (2018) for Mandarin Chinese, in its use of demonstratives with nouns vs. simple bare noun forms, as described and illustrated in 2.1.

## 2.1 Jenks (2018) and Mandarin Chinese

Jenks (2018) first observes that non-anaphoric unique definites occur as bare nouns (when not referenced with pronouns/pro), both in the case of immediate-situation definites, where an entity/individual is the only referent of its type in a particular situation, and with larger situation definites, where there is reference to an entity/individual that is unique in a broader context. Examples such as (9) and (10) illustrate

this patterning, drawn from other works on the structure and interpretation of noun phrases in Chinese. (9a/b) from Cheng and Sybesma (1999: 510) are examples of immediate-situation definites, while (10) an instance of a larger situation definite.

- (9) a. Hufei he-wan-le tang.  
 Hufei drink-finish-ASP soup  
 'Hufei finished the soup.'  
 b. gou yao guo malu.  
 dog want cross road  
 'The dog wants to cross the road.' (Cheng and Sybesma 1999: 510)
- (10) yueliang sheng shang lai le.  
 moon rise up come ASP  
 'The moon has risen.' (Chen 2004: 1165)

Jenks then presents a further range of elicited data suggesting that anaphoric definites are regularly coded in Mandarin by means of a demonstrative combined with a classifier and a noun, and may not occur as bare nouns, except when such referents occupy the (matrix) subject position. Jenks' examples demonstrating this distribution are reproduced in (11a–e). (11a) establishes a context in which two new referents are introduced in the context. (11b,c,e) indicate ways in which one of these two new referents can be referred back to by means of a demonstrative combined with a noun, but not with just a bare noun in direct object, indirect object and embedded subject position. (11d) shows that either a demonstrative-noun combination or a bare noun is, however, possible when the anaphoric definite occurs in matrix subject position.

- (11) a. jiaoshi li zuo-zhe yi ge nansheng he yi ge nüsheng.  
 classroom inside sit-PROG one CL boy and one CL girl  
 'There are a boy and a girl sitting in the classroom.'  
 b. wo zuotian yudao #(na ge) nansheng.  
 I yesterday meet that CL boy  
 'I met the boy yesterday.'  
 c. wo dai gei #(na ge) nansheng yi ge liwu.  
 I bring give that CL boy one CL present  
 'I'm bringing a gift for the boy.'  
 d. (na ge) nansheng kanqilai you er-shi sui zuoyou.  
 that CL boy look have two-ten year or.so  
 'The male student looks twenty years old or so.'  
 e. wo bu renwei?(na ge) nansheng hen youqu.  
 I NEG think that CL boy very interesting  
 'I don't think that boy is very interesting.'

Jenks' analysis of this paradigm builds on work in Chierchia (1998), Yang (2001), Dayal (2004, 2011), and Jiang (2012), and assumes that Mandarin employs type-shifting to effect the definite interpretation of bare nouns, but only in unique definite environments, resulting in the use of bare nouns to represent immediate and larger situation (non-anaphoric) definites (examples 9, 10). To account for the argued general absence of bare nouns in anaphoric definite reference, Jenks suggests that the relevant type-shifting operation available with nouns is unable to introduce a referential index necessary for anaphoric linking and identification. Demonstratives, however, are possible in such contexts, it is suggested, because such elements can introduce a referential index, licensing anaphoric construal with an appropriate antecedent, as in (the demonstrative variants of) examples (11b, c, and e). Demonstratives furthermore must occur in such cases of definite anaphoric reference, it is argued, because all referential indices must be represented overtly due to a principle which Jenks (2018: 524) dubs *Index!*

(12) *Index!*

Represent and bind all possible indices.

Finally, with regard to the exceptional case of subjects, Jenks proposes that bare nouns in subject positions are actually (continuing) topics and the 'pragmatic function of topic marking overrides and neutralizes the effect of *Index!* in such environments...: topics do not need to be indexed because they are salient members of the question under discussion' (Jenks 2018: 525). Anaphoric definites may therefore be realized in subject position either as demonstrative-classifier-noun combinations, or as bare nouns.

The patterns reported in Jenks characterize Mandarin as an interesting extension of the paradigms found in other languages in which anaphoric and unique definites are systematically coded in different ways by means of determiners or classifiers in opposition to bare nouns or nouns marked with a second, distinct determiner. Jenks' analysis is theoretically well-worked out and opens up new avenues for the cross-linguistic study and analysis of definiteness splits, bringing demonstratives into consideration as a marking strategy which may contrast with the use of bare nouns in languages without definite articles. However, with regard to Mandarin, Section 3, will now show that the descriptive conclusions in Jenks (2018) are actually not well-supported when a wider body of data is examined. This subsequently calls into question the formal status of *Index!* as a general principle regulating the representation of definite reference.



### 3. Probing the potential use of bare nouns as anaphoric definites in Mandarin

The Mandarin examples considered in Jenks (2018) involve a comparison of demonstrative and bare noun patterns of anaphoric definiteness in subject, object, and indirect object positions (11b–e). Here we will first look at the potential use of both forms of representation when an anaphoric definite occurs in other, pre-verbal positions, introduced by elements such as *ba* 把, *ti* 替, *gen* 跟 and *zai* 在, before also reconsidering the acceptability of bare nouns as anaphoric definites when occurring as direct and indirect objects. Where semantically plausible, we present examples in which the animacy of the target noun is varied in order to ensure that human vs. non-human distinctions are not an interfering factor.<sup>1</sup>

#### 3.1 NPs introduced by *ba* 把

When an anaphoric definite occurs as the object of *ba* 把 in pre-verbal position, as illustrated in examples (12–14), speakers consulted by the authors uniformly confirmed that either a demonstrative or a bare noun pattern is fully acceptable, hence bare nouns are not excluded from occurring in such a position when referring anaphorically to an element in a preceding sentence, and there is genuine optionality in the use of either demonstrative or bare noun pattern.

- (12) a. jiaoshi-li zuo-zhe yi-ge nan-sheng he yi-ge  
classroom-inside sit-ASP one-CL male-student and one-CL  
nǚ-sheng.  
female-student  
'In the classroom were sitting a male student and a female student.'
- b. wo ba (na-ge) nan-sheng ma-le yi dun.  
I BA (that-CL) male-student scold-ASP one session  
'I scolded the boy.'
- (13) a. wo zuotian shouyang-le yi-zhi xiaogou he yi-zhi xiaomao.  
I yesterday adopt-ASP one-CL puppy and one-CL kitten  
'Yesterday I adopted a puppy and a kitten.'
- b. wo ba (na zhi) xiaogou guan-zai huayuan-li.  
I BA (that-CL) puppy enclose-at garden-inside  
'I shut the puppy in the garden.'

1. In Bangla, for example, similar patterns are affected by animacy restrictions (Simpson and Biswas 2016) and animacy-related factors need to be carefully controlled for.

- (14) a. wo zuotian mai-le yi-tai diannaoh he yi-tai dianshi.  
I yesterday buy-ASP one-CL computer and one-CL TV  
'I bought a computer and a TV set yesterday.'
- b. wo ba (na-zhi) diannaoh fang-zai shufang-li.  
I BA (that-CL) computer put-at study-inside  
'I put the computer in the study.'

### 3.2 NPs introduced by *ti* 替

When an anaphoric definite is introduced with *ti* 替 in pre-verbal position, the same result as with *ba* 把 was obtained. Native speakers indicated that there is optionality in the way that the anaphoric definite may be represented, and either a bare noun or a demonstrative-marked noun may occur:

- (15) a. wo you yi-zhi xiaogou he yi-zhi xiaomao.  
I have one-CL puppy and one-CL kitten  
'I have a puppy and a kitten.'
- b. xiaogou changchang ti (na-zhi) xiaomao zhua laosu.  
puppy often for (that-CL) kitten catch mouse  
'The puppy often catches mice for the kitten.'
- (16) a. women gongsi you yi-ge kuaiji he yi-ge gongguan.  
our company have one-CL accountant and one-CL PR  
'Our company has an accountant and a PR (Public Relations Officer).'
- b. wo ti (na-ge) kuaiji zuo-le yi-ge baobiao.  
I for (that-CL) accountant make-ASP one-CL report  
'I made a report for the accountant.'

### 3.3 NPs introduced by *gen* 跟

Optionality in the use of demonstrative and bare noun patterns as anaphoric definites was similarly found with NPs that are introduced by *gen* 跟 in pre-verbal position:

- (17) a. wo you yi-zhi xiaogou he yi-zhi xiaomao.  
I have one-CL puppy and one-CL kitten  
'I have a puppy and a kitten.'
- b. wo changchang gen (na-zhi) xiaogou qu sanbu.  
I often with (that-CL) puppy go stroll  
'I often go walking with the puppy.'

- (18) a. women gongsi you yi-ge kuaiji he yi-ge gongguan.  
 our company have one-CL accountant and one-CL PR  
 ‘Our company has an accountant and a PR (Public Relations Officer).’
- b. wo gen (na-ge) kuaiji qu chuchai.  
 I with (that-CL) accountant go on-business-trip  
 ‘I went on a business trip with the accountant.’

### 3.4 NPs introduced by *zai* 在

Similar to the patterning with *ba*, *ti*, and *gen*, (inanimate) anaphoric definites combined with *zai* 在 in pre-verbal position are also free to occur either as bare nouns or in demonstrative-classifier-noun combinations:

- (19) a. wo mai-le yi-ge fangzi he yi-jian gongyu.  
 I buy-ASP one-CL house and one-CL apartment  
 ‘I bought a house and an apartment.’
- b. wo zai (nei-ge) gongyu-li fang-le henduo  
 I at (that-CL) apartment-inside put-ASP many  
 piaoliang-de dongxi.  
 pretty-DE things  
 ‘I put a lot of pretty things in the apartment.’

Across the four positional categories examined here which were not considered in Jenks (2018), it is therefore found that bare nouns may indeed occur as anaphoric definites and there is no complementary distribution between the demonstrative and bare noun pattern. Rather, there is optionality in the way that anaphoric definites may be encoded in Mandarin, contra the expectations of *Index!* which requires the use of demonstratives (or some other overt morpheme such as a determiner) when anaphoric definiteness can be represented with such an element.

### 3.5 Revisiting direct and indirect objects

In Jenks (2018), it is conceded that subjects may occur as bare nouns in contexts of anaphoric definiteness, following similar observations made in Jiang (2012). However, the legitimate use of bare nouns as anaphoric definites in subject position is argued to contrast with their potential occurrence in direct object and indirect object position, where it is suggested that only the demonstrative-classifier-noun pattern is acceptable. A consideration of further data relating to direct and indirect objects indicates that this is actually not the case, and bare nouns are in fact widely accepted by speakers as anaphoric definites in both such post-verbal positions, as shown in (20–24).

- (20) a. wo zuotian shouyang-le yi-zhi xiaogou he yi-zhi xiaomao.  
I yesterday adopt-ASP one-CL puppy and one-CL kitten  
'Yesterday I adopted a puppy and a kitten.'
- b. wo bijiao xihuan (na-zhi) xiaomao.  
I comparatively like (that-CL) kitten  
'I like the kitten more.'
- (21) a. wo zuotian mai-le yi-tai diannao he yi-tai dianshi.  
I yesterday buy-ASP one-CL computer and one-CL TV  
'I bought a computer and a TV set yesterday.'
- b. wo bijiao xihuan (na-tai) diannao.  
I comparatively like (that-CL) computer  
'I like the computer more.'
- (22) a. jiaoshi-li zuo-zhe yi-ge nan-sheng he yi-ge  
classroom-inside sit-ASP one-CL male-student and one-CL  
nü-sheng.  
female-student  
'In the classroom were sitting a male student and a female student.'
- b. wo gei (na-ge) nan-sheng yi-ge pingguo, gei (na-ge)  
I give (that-CL) male-student 1-CL apple give (that-CL)  
nü-sheng yi-ge juzi.  
female-student 1-CL orange  
'I gave the male student an apple and the female student an orange.'
- (23) a. fangjian-li you yi-zhi xiaogou he yi-zhi xiaomao.  
room-inside have one-CL puppy and one-CL kitten  
'There was a puppy and a kitten in the room.'
- b. wo gei (na-zhi) xiaogou yi-jian maoyi, gei (na-zhi) xiaomao  
I give (that-CL) puppy one-CL sweater give (that-CL) kitten  
yi-ge wanju.  
one-CL toy  
'I gave the puppy a sweater and the kitten a toy.'
- (24) you yi-ge xuesheng baifang-le Piao jiaoshou de bangongshi. Meixiangdao  
have one-CL student visit-ASP Park professor DE office unexpectedly  
Piao jiaoshou zhi tongma-le xuesheng yi-dun, ranhou jiu ba ta gan chu qu  
Piao professor only scold-ASP student 1 CL then just BA 3 chase out go  
'A student visited Prof. Park's office. However, Prof. Park unexpectedly just  
scolded him and chased him out of the office.'

(Roger Liao, personal communication)<sup>2</sup>

2. Roger Liao (personal communication) also notes that bare nouns can also be used as donkey pronouns in object position in Mandarin, contra what is suggested in Jenks (2018):

### 3.6 Conclusions, consequences and extensions

The observations made in 3.2–3.5 above indicate that the empirical generalizations concerning bare nouns and their potential use as anaphoric definites offered in Jenks (2018) do not seem to be correct once a broader array of data is considered. Whereas Jenks (2018: 501) suggests that ‘anaphoric definites are realized with a demonstrative, except in subject position’, it is not the case that the subject position is a single exception to the enforced use of demonstratives, and bare nouns are judged to be acceptable in a full range of syntactic positions. This then clearly calls into question the status of the principle *Index!* which requires that indices be represented overtly in anaphoric relations, wherever this is possible. As things stand, from the fuller paradigm presented in 3.2–3.5, it may well appear that there is no substantial justification for *Index!* as a principle governing the realization of anaphoric definite reference in Mandarin, and no reason to compare and equate the use of demonstratives in Mandarin in anaphoric dependencies with the forced use of definite determiners in languages such as English, French etc (as discussed in Chierchia 1998). This, we feel, is indeed the right general conclusion once Mandarin patterns are examined more holistically and a wider set of data is investigated with native speakers. The use of bare nouns as anaphoric definites is not restricted in the way suggested in Jenks (2018) and no principle barring their occurrence in non-subject positions is appropriate for Mandarin. Optionality in the representation of anaphoric definites either by overt means, employing a demonstrative, or without such overt coding, as simple bare nouns, is in principle permitted in the language in all sentential positions, suggesting that Mandarin has not grammaticalized any overtness requirement on the realization of anaphoric definiteness, unlike the obligatory occurrence of overt definite articles for all instances of definiteness in west European languages.

At such a point, one might conclude that there is nothing more to investigate in the domain of anaphoric definiteness and its representation in Mandarin, characterizing the language as one in which speakers freely use either demonstratives or

- 
- (i) mei-ge shouyang gou de ren dou yao dai gou qu kan shouyi.  
 every-CL adopt dog DE person all need take dog go see veterinarian  
 ‘Everyone who has adopted a dog needs to take the dog to see a vet.’

Example (ii) additionally shows a bare noun used as a donkey pronoun in pre-verbal object-of-*gei* position. The context for (ii) would be a festival where people dress the buffaloes they own:

- (ii) mei-ge yang shuiniu de renjia dou gei shuiniu chuan-shang-le yifu.  
 every-CL keep buffalo DE family all give buffalo wear-up-ASP clothing  
 ‘Every family who has a buffalo put clothes on it.’

bare nouns to encode referential noun phrases linked to some linguistically-present antecedent. However, there is, in fact, more to consider here and certain additional confounding complexity which must at least be recognized, and ideally should (in some measure) be explained. The issue is that the acceptability of the bare noun pattern in instances of anaphoric definite reference which is under discussion here does appear to vary somewhat according to context, ranging from full acceptability (leading to the conclusion that such patterns are in principle well-formed), to hesitation among speakers and uncertainty about whether such forms should be classed as acceptable or not. In the latter cases, speakers are typically unwilling to class examples as ungrammatical (unlike instances of the omission of articles in English-type languages), but it is suggested that the use of bare nouns as anaphoric definites in certain sentences may sometimes sound rather unnatural and would not typically be used, there being a preference in such cases for use of the demonstrative pattern. While the use of bare nouns in examples presented in 3.2–3.5 above were all readily accepted by the speakers we consulted, we experienced mixed reactions to Jenks' original data, in particular to example (11b), repeated below. Informants frequently indicated that they were not sure how to class (11b) when a bare noun was used and suggested that they themselves would not use a bare noun in such a context, although it did not seem ungrammatical. This reaction was clearly different to the way that the data (12–23) was judged, which speakers were regularly quick to judge as acceptable, raising questions both about why the bare noun strategy seems unnatural in certain contexts but is broadly available in other instances (in all syntactic positions), and what kinds of lesson can be drawn with regard to the construction and comparison of data within a single language and across different languages. We will attempt an answer to the first, challenging question in 3.7 below, and then add some suggestions about researching this area of language.

- (11) a. jiaoshi li zuo-zhe yi ge nansheng he yi ge nüsheng.  
 classroom inside sit-PROG one CL boy and one CL girl  
 'There are a boy and a girl sitting in the classroom.'
- b. wo zuotian yudao #(na ge) nansheng.  
 I yesterday meet that CL boy  
 'I met the boy yesterday.'

### 3.7 Discourse constraints on the bare noun pattern in anaphoric definiteness

We believe that speakers' preference for a demonstrative form in certain instances and their suggestions that a bare noun may not sound fully natural are actually due to two factors, which may conspire with each other in various constructed examples. The first of these is a preference found elsewhere in anaphoric reference in Chinese and other languages for the use of more explicit representational forms when there are major or minor breaks in the coherence of sentences comprising a particular discourse (discussed below). The second factor we believe to be relevant is the tendency and expectation for (post-verbal) objects in Chinese to be interpreted by default as indefinite in reference, unless otherwise marked as definite or construed as specific/definite due to the specificity of the event described in a sentence. In what follows, we elaborate on these two factors and the ways they may lead to preferences in the use of either bare nouns or demonstrative patterns in the representation of anaphoric definites.

It can first be noted that the demonstrative-classifier-noun pattern is clearly more explicit in its coding of definiteness than the use of a bare noun, as the former is unambiguously definite (and singular) in specification, while bare nouns may be interpreted as either definite or indefinite (and singular or plural). The occurrence of a demonstrative with a noun provides listeners with an instruction to search for a local antecedent in the discourse (if manual pointing is not involved, causing a hearer to look for a local reference in his/her visual area). Such a parsing instruction is not present with bare nouns, which can be interpreted in various ways, and this difference introduces a hierarchical relation in the pair of forms potentially available for definite anaphoric reference – where a more explicit referential form is felt necessary in a particular context, the expectation is that a demonstrative pattern will naturally be preferred, whereas bare nouns may be used as anaphoric definites in discourse contexts (shortly to be described) when there is less call for an explicitly marked definite form. Approaching the bare noun vs. demonstrative patterning from such a perspective, we suggest that there is an important comparison that can be made with other sets of referential elements and the preferences for their use which relate to the availability of a range of forms which are referentially more vs. less explicit. Specifically, in studies of anaphoric reference in Chinese and other pro-drop languages (Givón 1983, Chen 1986, Christensen 2000, and Pu 2011), it has frequently been argued that a hierarchy of representational forms exists and governs speakers' selection of null subjects/objects (pro, 'zero anaphora'), overt pronouns, or full NPs in instances of anaphoric reference, there being a preference for less explicit forms (pro, or overt pronouns) for discourse referents that are highly activated and salient, and more explicit forms (full NPs) when a referent becomes less activated/salient. In attempting to account for the relative acceptability/

preference for use of *pro* vs. overt pronouns vs. full NPs as less vs. more explicit forms of reference, it has been proposed that this can be attributed to aspects of *discourse coherence* and the ways that the semantic-pragmatic relation of sentences to each other affects the activation level of referents within a particular discourse. The more tightly ‘connected’ that sentences are to each other in terms of shared event time, location, action, speaker perspective and other properties, the greater probability that speakers will opt for use of less explicit referential forms, as discourse referents remain activated and salient through closely-bound sequences of sentences. By way of contrast, where there are breaks in discourse coherence caused by switches in event time, place and perspective-shift, it is argued that speakers will select a more explicit means of reference (e.g. full NPs) to compensate for the lowering of individual referents’ activation level. In previous work carried out by the authors of the current chapter (investigating pronoun preference strategies in Mandarin Chinese – Simpson, Li and Wu 2016: 3), we described the line of research pursued in Chen (1986), Tai (1978), Pu (1995, 2011) and Li and Thompson (1979), in following way:

These works suggest that major and minor discontinuities in discourse structure cause alternations in the different anaphoric forms of reference that speakers regularly use. Zero anaphora/zero pronouns are described as typically being used when a topical referent remains in focal attention across many sentences in succession in a ‘topic chain’, in which the sentences all ‘cohere’ well and describe a closely related sequence of events (Chen 1986), or are otherwise semantically closely linked with each other (Tai 1978). Overt pronouns are argued to occur when anaphor-antecedent relations are structurally interrupted by minor breaks and discontinuities in discourse structure, caused by changes in temporal, spatial or action continuity, from transitions in description of a referent’s physical activities to his/her mental activity, and from switches in narration from background to foreground information (Pu 2011). Such disruptions are suggested to affect speakers’ attention on a referent, lowering its activation level and favoring the use of a more explicit form of reference – an overt pronoun rather than zero anaphora. More major breaks in the episodic structure of a discourse/narrative, often corresponding with paragraph breaks in writing, are described as resulting in the use of repeated full NPs to refer back to topical referents in a story line. (Simpson, Li and Wu 2016: 3)

Two examples from Tai (1978) illustrate how preferences in the choice of anaphoric forms may be linked to discourse coherence. In (25), speakers are reported to (often) prefer the use of a *pro* subject to an overt pronoun (which is felt to be less natural, though not ungrammatical), as a tight semantic connection is perceived to hold between the two clauses, whereas in (26), there is less of a (perceived) close connection between the two clauses and an overt pronoun is frequently noted to be preferred over the use of *pro*:



- (25) xiao Mei jie-le hun, pro/?ta sheng-le liang-ge haizi.  
 little Mei connect-ASP wedding pro/3 bear-ASP 2-CL child  
 ‘Little Mei got married and had two children.’
- (26) lao Zhang zuotian lai kan wo, ta/?pro dai-le yi-ben zazhi  
 old Zhang yesterday come see me 3/pro bring-ASP 1-CL magazine  
 gei wo.  
 for me  
 ‘Yesterday old Zhang came to see me and he brought me a magazine.’

Returning to the alternation between bare nouns and demonstratives which are the focus of the present chapter, we suggest that similar features of discourse coherence to those highlighted above are also at play in guiding speaker preference for one form over the other in various contexts, and may be responsible for differences in speaker reaction to examples which are otherwise not syntactically distinguished. Where discourse coherence (gauged approximately in terms of shared time, location, and speaker perspective) is high, we posit that bare nouns will be judged fully acceptable in instances of anaphoric definite reference to salient individuals, while the more explicit demonstrative pattern will be preferred when there are breaks in coherence and continuity, and bare nouns will be viewed as less natural and dispreferred. Taking a specific example to illustrate this, Jenks’ pair of sentences (11a/b) repeated again below involve a shift in time and (presumably) place between the situation described in (11a) and the event in (11b), which took place on a preceding day (and most probably in a different location). Speakers report that they feel little close connection between the content of (11a) and (11b) and indicate that they would prefer the use of the demonstrative pattern in such cases, so as to make it very clear that the ‘boy’ mentioned in (11b) is intended to refer to the boy in (11a). The shift in time and location between these juxtaposed sentences arguably constitutes a break in the coherence of the discourse that is significant enough to render a less explicit form of reference less than optimal, and speakers typically avoid the bare noun pattern in such a context.

- (11) a. jiaoshi li zuo-zhe yi ge nansheng he yi ge nüsheng.  
 classroom inside sit-PROG one CL boy and one CL girl  
 ‘There are a boy and a girl sitting in the classroom.’
- b. wo zuotian yudao #(na ge) nansheng.  
 I yesterday meet that CL boy  
 ‘I met the boy yesterday.’

Further examples can be created where bare nouns are felt to be less natural than the use of a demonstrative pattern and where there are shifts in time or place or perspective between two paired sentences, with the first sentence containing an

indefinite noun phrase and the second an anaphoric definite intended to refer back to the indefinite NP.

- (27) a. wo jintian zai gongyuan renshi-le yi-ge yisheng he yi-ge kuaiji  
 I today in park meet-ASP 1-CL doctor and 1-CL accountant  
 ‘Today in the park I met a doctor and an accountant.’  
 b. zuotian #(na-ge) yisheng zai women jia-fujin mai-le  
 yesterday DEM-CL doctor at 1.PL home-nearby buy-ASP  
 yi-suo fangzi.  
 1-CL house  
 ‘Yesterday the doctor bought a house in our neighborhood.’
- (28) a. wo zai xuexiao ganggang kandao-le yi-zhi mao.  
 I at school just-now see-ASP 1-CL cat  
 ‘I just saw a cat in school.’  
 b. Zhangsan zuotian shuo #(na-zhi) mao tou-le hen duo  
 Zhangsan yesterday say DEM-CL cat steal-ASP very many  
 dongxi chi.  
 thing eat  
 ‘Yesterday Zhangsan said that the cat stole many things to eat.’
- (29) a. wo jintian mai-le yi-tai dianshi.  
 I yesterday buy-ASP one-CL TV  
 ‘Today I bought a computer and a TV set.’  
 b. zuotian baihuogongsi ba #(na-tai) dianshi jiang-le liangbai  
 yesterday department store BA DEM-CL television lower-ASP  
 kuai.  
 200 dollar  
 ‘Yesterday the department store reduced the price of the computer by \$200.’

Calculating precisely how much of a time/location/perspective shift may result in a discontinuity significant enough to make bare nouns feel less natural as anaphoric definites and less preferred than demonstrative patterns will not be attempted here and is a task that would require considerable data analysis, complicated further by the fact that there is also certain speaker variability in the judgement of relevant examples. However, given that discourse coherence effects of the type alluded to here have been argued to constrain preferences in the choice of other anaphoric elements such as *pro*, pronouns and full NPs in a variety of studies, we believe such an approach is likely to be on the right track and is the kind of explanation that can make sense of variable speaker reactions to the use of bare nouns and demonstratives in patterns of anaphoric definiteness.

A second factor we believe also plays a potentially significant role in constraining the natural use of bare nouns as anaphoric definites is the tendency for

object positions in Chinese (and other languages) to be the locus of new information and indefinite noun phrases (Chao 1968, Li and Thompson 1981, Givón 1983, Lambrecht 1994, Erteschik-Shir 2007). In some varieties of Chinese from the Wu and Min areas, it is reported that definite NPs may either not occur in post-verbal positions (e.g. Li and Bisang on Fuyang Wu) or are ordinarily preposed to some position preceding the verb (as, for example, in Taiwanese), and for Mandarin it has regularly been observed that objects which are definite in reference are very frequently displaced from post-verbal position to initial or preverbal topic or focus positions, or introduced with pre-verbal *ba* constructions (Chao 1968, Li and Thompson 1981). NPs in post-verbal position are therefore commonly indefinite in reference.<sup>3</sup> This general aspect of information structure can be assumed to establish a default parsing strategy, where NPs which are not explicitly marked as definite in reference will be interpreted by hearers as indefinite and discourse-new, unless other aspects of the containing discourse strongly favor a definite interpretation. As bare nouns in Mandarin can be interpreted as definite, indefinite or generic in reference, it may be expected that speakers will tend to avoid the use of a bare noun as an anaphoric definite in object position and instead prefer the use of a demonstrative as a way to override a default indefinite interpretation – *unless* a definite interpretation is otherwise naturally made salient by properties of the context. In charting speaker reaction to the use of bare nouns and demonstrative-marked nouns in various sentence-internal positions, the current investigation observed that speakers exhibited hesitation in judging sentences to be acceptable more frequently when bare nouns occurred (as anaphoric definites) in post-verbal object position, rather than other positions, though typically speakers concluded that such examples were in fact acceptable and grammatical after certain reflection. We take this slight hesitation as indication that the *tendency* to construe objects as indefinite,

3. In one particular construction, an NP which can be interpreted as definite in subject position may *not* be given a definite interpretation in object position. Sequences of numeral+classifier+noun are generally interpreted as indefinite in reference, but it is also possible for such constituents to be construed as definite in situations of anaphoric definite reference, as in (i)

- (i) liang ben shu zai nali?  
 2 CL book be where  
 ‘Where are the two books?’

However, it is not possible for such a definite interpretation to occur in object position, and numeral+classifier+noun sequences can only be interpreted as indefinite in post-verbal positions, as illustrated in (ii):

- (ii) ni mai-le liang ben shu ma  
 you buy-ASP 2 CL book Q  
 Only: ‘Did you buy two books?’  
 Not: ‘Did you buy the two books?’

new information and the expectation that they will have such an interpretation may sometimes interfere with their construal as anaphoric definites, although such an interpretation is grammatically licensed and available as an interpretation when also contextually natural within a sufficiently coherent segment of discourse.<sup>4</sup>

We therefore posit that aspects of discourse coherence and also expectations relating to the positioning of in/definite information may affect speaker reactions to example sentences in instances of anaphoric definiteness and skew judgments in ways that do not reflect the grammaticality of such forms. The potential effect of such confounds emphasizes the need for researchers to consider a wide range of data when examining patterns such as the representation of discourse referents, to help avoid extra-grammatical interference in the characterization of language-particular paradigms. Additionally, the *cross-linguistic comparison* of such phenomena will benefit greatly, we believe, from care being taken to investigate example sentences which are as similar as possible in different languages rather than the use of data that is not obviously equivalent (where distortions may arise as the result of variable discourse factors), as has sometimes occurred in previous studies. Section 4 of the chapter now reports on a comparison of Mandarin with three other languages/varieties, Korean, Hindi and Cantonese, with replications of the same data presented in sections 2 and 3. We use this investigation to establish a preliminary impression of the ways that other languages without articles may represent anaphoric definites and specifically whether bare noun patterns are available as options alongside other more explicit modes of representation – hence whether the optionality identified here in Mandarin may also perhaps occur in other languages.

#### 4. The representation of anaphoric definites in Korean, Hindi, and Cantonese

Korean, Hindi, and Cantonese resemble Mandarin in not having any definite article. This raises the question of how anaphoric definites may be represented in such language varieties and whether there is any optionality in the use of demonstrative and bare noun patterns (or the ‘bare classifier’ pattern found in Cantonese). Sections 4.1 and 4.2 report that both Korean and Hindi appear to pattern like Mandarin in regularly permitting either a demonstrative+noun or a bare noun to be used in instances of anaphoric definite reference, hence the same (or a very similar) kind

---

4. Jenks example (11b) as a follow-on to (11a) therefore resists an easy anaphoric definite interpretation when a bare noun occurs due to both of the factors mentioned here, which conspire to reduce the acceptability of such an interpretation in such examples. The bare noun in (11b) occurs in post-verbal object position, and the sequence of (11a–b) lacks the discourse coherence of other more acceptable examples, with a significant backward shift in both time and location.

of optionality as that noted in Mandarin in Section 3. It can consequently be concluded that Mandarin, as characterized in the current study, is not exceptional in allowing for either of two representational strategies for anaphoric definites, subject to appropriate discourse licensing. Cantonese, by way of contrast, does not appear to tolerate optionality in the representation of anaphoric definites, and enforces the exclusive use of its bare classifier pattern and no occurrence of bare nouns in contexts of anaphoric definiteness. In Cantonese, the representation of definiteness has therefore been fully grammaticalized (with the bare classifier pattern), as in languages with definite determiners such as English, whereas Mandarin, Korean and Hindi remain in a more fluid state, with alternative modes of representation not constrained by any hard principles of syntax.

#### 4.1 Korean

Considering Korean, an investigation of the same patterns of data examined in Section 3 for Mandarin results in the conclusion that either a demonstrative+noun pattern or a bare noun is widely acceptable in instances of anaphoric definiteness. In examples (15–21) below, we present the acceptable occurrence of a bare noun as anaphoric definite in a range of syntactic positions, including subject, direct object, indirect object, and different adjunct positions. In all such cases, native speaker consultants reported that a demonstrative+noun would also be acceptable. Korean therefore appears to pattern like Mandarin Chinese (under further scrutiny), allowing for optionality in the use of a demonstrative and bare noun pattern in contexts of anaphoric definiteness, and such alternations are not restricted by syntactic position.<sup>5 6</sup>

##### *Target (anaphoric definite) in subject position*

- (30) a. kyosil-ey            nam-haksayng-kwa ye-haksayng-i  
           classroom-LOC male-student-and female-student-NOM  
           anc-aissessta.  
           sit-were            sitting  
           ‘In the classroom were sitting a male student and a female student.’

5. The potential use of bare nouns as anaphoric definites has also very recently been reported in Kang and Park (2020), though just with a single example (Example 5, p. 454).

6. We can also report that our brief survey of Korean with ten speakers from Pusan University did not detect any very obvious effects of discourse coherence on the use of bare nouns as anaphoric definites. However, we would not be surprised if such factors were to condition other speakers’ judgments, and one additional speaker we consulted was less automatic in her acceptance of bare nouns in all of the cases tested. Further investigation is therefore warranted, we suggest. Many thanks indeed to Soyoung Park for her help in facilitating this short study of Korean.

- b. nam-haksayng-un sumu-sal cengto toyé pointa.  
 male-student-TOP 20-year or so become look  
 'The male student looks about 20 years old.'

*Target (anaphoric definite) in direct object position*

- (31) a. kyosil-ey nam-haksayng-kwa ye-haksayng-i  
 classroom-LOC male-student-and female-student-NOM  
 anc-aissessta.  
 sit-were sitting  
 'In the classroom were sitting a male student and a female student.'
- b. na-nun nam-haksayng-ul honnayssta.  
 I-TOP male-student-ACC scolded  
 'I scolded the male student.'
- (32) a. na-nun ecey kangaci-wa koyangi-lul ipyanghayssta.  
 I-TOP yesterday puppy-and kitten-ACC adopted  
 'Yesterday I adopted a puppy and a kitten.'
- b. na-nun kangaci-lul cengwen-ey katwuessta.  
 I-TOP puppy-ACC garden-LOC shut  
 'I shut the puppy in the garden.'
- (33) a. na-nun ecey khempyuthe-wa thibi seyту-lul sassta.  
 I-TOP yesterday computer-and TV set-ACC bought  
 'I bought a computer and a TV set yesterday.'
- b. na-nun khempyuthe-lul kongpu pang-ey nohatwuessta.  
 I-TOP computer-ACC study room-LOC put  
 'I put the computer in the study.'

*Target (anaphoric definite) in indirect object position*

- (34) a. kyosil-ey nam-haksayng-kwa ye-haksayng-i  
 classroom-LOC male-student-and female-student-NOM  
 anc-aissessta.  
 sit-were sitting  
 'In the classroom were sitting a male student and a female student.'
- b. na-nun nam-haksayng-eykey sakwa-lul, ye-haksayng-eykey  
 I-TOP male-student-DAT apple-ACC female-student-DAT  
 kyul-ul cwuessta.  
 tangerine-ACC gave  
 'I gave the male student an apple and the female student a tangerine.'

*Target (anaphoric definite) in PP benefactive position*

- (35) a. na-eykey-nun kangaci-wa koyangi-ka issta.  
 I-DAT-TOP puppy-and kitten-NOM be  
 'I have a puppy and a kitten.'

- b. kangaci-nun congcong kongyangi-lul-wyhay cwyi-lul capa-cwunta.  
 puppy-TOP often kitten-for mice-ACC catch-give  
 'The puppy often catches mice for the kitten.'

*Target (anaphoric definite) in PP comitative position*

- (36) a. na-eykey-nun kangaci-wa koyangi-ka issta.  
 I-DAT-TOP puppy-and kitten-NOM be  
 'I have a puppy and a kitten.'
- b. na-nun congcong kangaci-wahamkkey sanchayk-ul hanta.  
 I-TOP often puppy-with walking-ACC do  
 'I often go walking with the puppy.'

## 4.2 Hindi

For Hindi, we also tested the same set of data examined in Mandarin and Korean to see how anaphoric definites would be encoded, either via a demonstrative+noun pattern or with bare nouns, or with either such forms of representation. As with Korean and Mandarin, the target anaphoric definite was placed in a full range of syntactic positions, to check whether the syntactic position of the target might affect its mode of representation. Illustrative examples with bare nouns are given in (37–41). The broad generalization found in the data investigated with ten native speakers was that bare nouns are widely acceptable as anaphoric definites in Hindi.<sup>7</sup> Speakers regularly accepted either a bare noun or a demonstrative pattern, with a common preference typically being expressed for the bare noun pattern, or, in certain instances (examples 25 and 26), a strong majority of speakers accepting only the bare noun pattern. The responses gathered in this study suggest that the use of a bare noun for anaphoric definites occurs as a broadly acceptable pattern, and for the majority of speakers consulted was furthermore a preferred alternative to the use of a demonstrative. There was one curious exception to this clear patterning found in our group of speakers. For example (41), our consultants oddly reversed their preferences and indicated that either they preferred a demonstrative over a bare noun pattern, or they only found a demonstrative acceptable/natural in this context. Interestingly, example (41) corresponds in meaning to Jenks' (2018) Mandarin example numbered (11a/b) in the current chapter, which our Mandarin consultants also hesitated to accept with use of a bare noun. Above in Section 3 we suggested that this may be due to a reduced level of discourse coherence and a time/location continuity break between the two conjoined sentences, favoring the

7. Many thanks to Bhamati Dash for helping facilitate this study of Hindi with other native speakers.

use of a more explicit demonstrative in place of a bare noun. The observation that such a discontinuity also triggers a different, reversed judgment from speakers of Hindi who otherwise find bare nouns very acceptable (or even preferable to the use of a demonstrative) suggests to us that similar aspects of discourse coherence may condition speaker acceptance of bare nouns vs. demonstratives in Hindi as well as Mandarin. When discourse discontinuities were avoided in examples (37–40), speakers readily accepted or preferred bare nouns as anaphoric definites, whereas they expressed an opposite reaction to the sentence pairing in (41), adapted from Jenks' data set, where a more significant break in discourse continuity occurs.

*Target (anaphoric definite) in subject position*

- (37) a. kaksha-me ek laDkaa aur ek laDkii betihe the  
classroom-in one boy and one girl sitting were  
'In the classroom were sitting a boy and a girl.'  
b. laDka takriban bees saal ka hai  
boy approximately 20 year LNK be  
'The boy looks twenty years old or so.'

*Target (anaphoric definite) in direct object position*

- (38) a. kal mein ek kutte aur ek billi-ko ghar lekar aaya  
yesterday I one dog and one cat-ACC home bring came  
'Yesterday I adopted a dog and a cat (lit. 'brought a dog and cat home').'  
b. mei-ne kutte-ko bagiche-me band kar diyaa  
I-ERG dog-ACC garden-LOC close do gave  
'I shut the dog in the garden.'

*Target (anaphoric definite) in indirect object position*

- (39) a. kaksha-me ek laDkaa aur ek laDkii betihe the  
classroom-in one boy and one girl sitting were  
'In the classroom were sitting a boy and a girl.'  
b. mei-ne laDke-ko ek sev diyaa  
I-ERG boy-DAT I apple gave  
'I gave the male student an apple.'

*Target (anaphoric definite) in PP locative position*

- (40) a. mei-ne kal ek computer aur ek tv kharidaa  
I-ERG yesterday one computer and one tv bought  
'I bought a computer and a TV set yesterday.'  
b. mei-ne computer studyroom-me rakh diyaa  
I-ERG computer study-LOC put gave  
'I put the computer in the study.'



*Target (anaphoric definite) in postpositional object position*

- (41) a. kaksha-me ek laDkaa aur ek laDkii betihe the  
 classroom-in one boy and one girl sitting were  
 'In the classroom were sitting a boy and a girl.'
- b. mein kal laDke-se milii thii  
 I-ERG yesterday boy-with met did  
 'Yesterday I met the boy.'

## 4.3 Cantonese

Having identified a much greater optionality in the use of bare nouns as anaphoric definites in Mandarin than suggested in recent literature, and also found similar optionality to be present in both Korean and Hindi, we decided to revisit Cantonese, a variety in which it has long been reported that definiteness is regularly represented with the combination of a classifier and a noun – the bare classifier pattern described in Section 2 for Bangla and Jinyun Wu Chinese (examples 5–8). Works such as Cheng and Sybesma (1999, 2005) have well documented the use of this pattern in Cantonese and contrasted it with the occurrence of bare noun patterns in Mandarin, pointing out that Cantonese makes use of the combination of a classifier and noun (with no accompanying numeral) to encode definite reference in positions where Mandarin would use a bare noun, as in (42) and (43):

- (42) \*(zek) gau soeng gwo maalou. (Cantonese)  
 CL dog want cross road  
 'The dog wants to cross the road.' (Cheng and Sybesma 1999)
- (43) gou xiang guo malu. (Mandarin)  
 dog want cross road  
 'The dog wants to cross the road.' (Cheng and Sybesma 1999)

Cheng and Sybesma's seminal studies of Cantonese did not set out to distinguish contexts of anaphoric vs. situational definiteness and the potential effects such contexts might have on the ways that definites are encoded in Cantonese. The question we wished to reach an answer to as an extension of the chapter's current investigation is whether any variation in the use of a bare classifier pattern might occur in contexts of anaphoric definiteness, and whether speakers might perhaps also accept the use of bare nouns in such contexts if data parallel to that used in Section 3 for Mandarin were to be tested with speakers of Cantonese. The results of probing such patterns in Cantonese confirm the default assumption present in Cheng and Sybesma (1999, 2005) that anaphoric definites, like unique definites, are regularly encoded with the bare classifier pattern, and bare nouns are not acceptable.

This is illustrated with a subset of the data tested with informants in (44) – (49). The brief study of Cantonese carried out here also asked consultants whether a demonstrative-classifier-noun pattern could be used in instances of anaphoric definiteness, and the results indicated that this was in many cases not felt to be natural. The broad generalizations emerging from the data are that: (a) the most natural pattern used to represent anaphoric definites in Cantonese is indeed the bare classifier pattern,<sup>8</sup> (b) bare nouns are not acceptable in such contexts, and (c) demonstratives are often not acceptable substitutes for the bare classifier pattern.<sup>9</sup> Cantonese is consequently much more rigid in its representation of anaphoric definiteness than Mandarin, Korean and Hindi, and the required use of the bare classifier pattern blocks any potential use of bare nouns in contexts of anaphoric (and non-anaphoric unique) definiteness. This is represented in the following examples modeled on the Mandarin patterns examined in Section 3.

*Target (anaphoric definites) in subject and benefactive positions*

- (44) a. ngo5 jau5 jat1-zek3 gau2-zai2 tung4 jat1-zek3 maau1zai2  
 I have one-CL puppy and one-CL kitten  
 ‘I have a puppy and a kitten.’  
 b. {<sup>?</sup>go2-zek3/\*ø} gau2zai2 sing4jat6 bong1 {<sup>?</sup>go2-zek3/zek3/\*ø} maau1zai2  
 DEM-CL/CL/ø puppy often for DEM-CL/CL/ø kitten  
 zuk1 lou5syu2  
 catch mouse  
 ‘The puppy often catches mice for the kitten.’

*Target (anaphoric definite) in direct object position*

- (45) a. ngo5 cam4jat6 maai5-zo2 jat1-bou6 din6nou5 tung4 jat1-bou6 din6si6  
 I yesterday buy-ASP one-CL computer and one-CL TV  
 ‘I bought a computer and a TV set yesterday.’  
 b. ngo5 bei2gaau3 zung1ji3 {<sup>?</sup>go2-bou6/bou6/\*ø} din6nou5  
 I comparatively like DEM-CL/CL/ø computer  
 ‘I like the computer more.’

8. Jenks (2018) also reports that the bare classifier pattern is used for anaphoric definites in Cantonese, though with a more restricted data set than examined here.

9. More specifically, as indicated in examples (44–49), demonstratives were felt to be acceptable when the target anaphoric definite occurred as indirect object (46) and pre-verbal object of *zoeng3* (47) (equivalent to Mandarin *ba*), but not natural when an anaphoric definite occurred as subject (44), direct object (45), benefactive object (44), or as the pre-verbal objects of *tung4* ‘with’ (48) or *hai2* ‘in’ (49). It is currently not clear how to make good sense of this uneven distribution in acceptability.

*Target (anaphoric definite) in indirect object position*

- (46) a. fo3sat1-leoi5min6 co5-zo2 jat1-go3 naam4zai2 tung4  
classroom-inside sit-ASP one-CL boy and  
jat1-go3 nei5zai2  
one-CL girl  
'In the classroom were sitting a boy and a girl.'
- b. ngo5 bei2-zo2 jat1go3 ping4gwo2 {go2-go3/go3/\*ø} naam4zai2,  
I give-ASP 1-CL apple DEM-CL/CL/ø boy  
bei2-zo2 jat1-go3 caang2 {go2-go3/go3/\*ø} nei5zai2  
give-ASP 1-CL tangerine DEM-CL/CL/ø girl  
'I gave the boy an apple and the girl a tangerine.'

*Target (anaphoric definite) in pre-verbal zoeng3 object position<sup>10</sup>*

- (47) a. fo3sat1-leoi5min6 co5-zo2 jat1-go3 naam4zai2 tung4  
classroom-inside sit-ASP one-CL boy and  
jat1-go3 nei5zai2  
one-CL girl  
'In the classroom were sitting a boy student and a girl.'
- b. ngo5 zoeng3 {go2-go3/go3/\*ø} naam4zai2 naau6-zo2 jat1 caan1  
I BA DEM-CL/CL/ø boy scold-ASP one session  
'I scolded the boy.'

*Target (anaphoric definite) as pre-verbal object of tung4 'with'*

- (48) a. ngo5dei6 gung1si1 jau5 jat1-go3 wui6gai3 tung4  
our company have one-CL accountant and  
jat1-go3 gung1gwaan1  
one-CL PR  
'Our company has an accountant and a Public Relations Officer.'
- b. ngo5 soeng5-go3-jyut6 tung4 {go2-go3/go3/\*ø} wui6gai3 heoi3  
I last-CL-month with DEM-CL/CL/ø accountant go  
gung1gon3  
on-business-trip  
'Last month I went on a business trip with the accountant.'

*Target (anaphoric definite) as pre-verbal object of hai2 'in'*

- (49) a. ngo5 maai5-zo2 jat1gaan1 nguk1 tung4 jat1-go3 daan1wai2  
I buy-ASP 1-CL house and 1-CL apartment  
'I bought a house and an apartment.'

---

10. Equivalent to the Mandarin *ba* construction.

- b. ngo5 hai2 {<sup>?</sup>go2-gaan1/gaan1/\*ø} nguk1-jap6min6 fong3-zo2 hou2do1  
 I in DEM-CL/CL/ø house-inside put-ASP many  
 hou2leng3 ge3 je5  
 pretty-DE things PRT  
 ‘I put a lot of pretty things in the house.’

Having extended our Mandarin-based study into other languages/varieties without definite articles and seen clear similarities and differences between Mandarin, Korean, Hindi and Cantonese, we now bring the findings of the chapter together in Section 5 and summarize the conclusions that can be drawn about overt/covert representational forms of anaphoric definiteness.

## 5. Summary and conclusions

This chapter set out to re-examine the descriptive generalization offered in Jenks (2018) that bare nouns in Mandarin Chinese may only be used as anaphoric definites in subject and topic positions, as a partial reflex of the principle *Index!*, which requires that all referential indices be overtly represented, in Mandarin this being satisfied by the use of a demonstrative. Considering a broader range of data than presented in Jenks’ study, Section 3 established that bare nouns may actually be judged to be acceptable as anaphoric definites in all syntactic positions in Mandarin Chinese, contra the expectations of *Index!* Examining the contrasts found between judgments elicited from our consultants and those reported in Jenks (2018), it was suggested that the naturalness of bare nouns in contexts of anaphoric definiteness in Mandarin may be affected by two additional factors – the degree of discourse coherence existing between sentences which contain the antecedent and anaphoric definite expression, and default expectations that bare nouns in object positions introduce new information and so should be interpreted as indefinite. The effects of discourse coherence on speakers’ choices of representational forms have elsewhere been invoked as an explanation of preferences in the use anaphoric elements such as *pro*, overt pronouns, and full NPs, and may lead speakers to prefer more explicit forms of reference where there are minor or major breaks in discourse continuity. Here we have suggested that similar aspects of discourse structure may affect the choice of representational forms in other instances of anaphoric definiteness when pronouns/*pro* are not possible choices (due to reasons of ambiguity when there are two potential referents in a preceding sentence) and sometimes lead speakers to a preference for a more explicit, demonstrative form, over the use of a bare noun, although the latter is a choice formally permitted by the grammar. The selection of an explicitly definite form, signaled by the presence of a demonstrative, may also,

for certain speakers, be preferred as a way to over-ride default expectations that nouns in object positions in Mandarin are indefinite in reference. We believe that a combination of these factors may account for variability in speaker reaction to data involving the use of bare nouns as anaphoric definites, and when such factors are effectively controlled for, the correct observation is that both representational forms (bare nouns and demonstrative patterns) are *grammatically* licensed to occur in all syntactic positions in Mandarin. The principle *Index!* and its requirement of the use of overt markers of indices (such as demonstratives) is consequently too rigid as a constraint on Mandarin Chinese, and genuine optionality regularly occurs in the ways that anaphoric definites may be encoded – either by means of bare nouns or with demonstrative patterns.

The chapter also extended its study of Mandarin further to consider related patterns in three other language varieties, Korean, Hindi, and Cantonese. In doing so, we hoped to be able to situate Mandarin relative to other languages without articles and the way they may represent anaphoric definites. In 4.3, Cantonese was confirmed as a variety in which anaphoric definites are rigidly encoded with a bare classifier pattern (classifier + noun) and not with bare nouns. Because situational definites in Cantonese are also commonly represented with the bare classifier pattern (as noted in Jenks 2018), Cantonese appears to require a rule blocking the use of bare nouns in (all) contexts of definiteness, as suggested in Chierchia (1989) for languages which do have definite articles, such as English, French and German. Such a blocking principle justified for Cantonese will necessarily have to be language-specific, however, as bare nouns are readily available as definites in other languages such as Mandarin, Korean, and Hindi, and so should not be blocked by any grammatical principle requiring that definiteness be expressed overtly wherever possible. In Mandarin, as noted above, there is optional alternation between bare nouns and more ‘explicit’ demonstrative patterns in contexts of anaphoric definiteness, and similar patterns were observed in Korean and Hindi, hence no absolute blocking principle is warranted for these languages and *Index!* cannot be assumed to have a universal, cross-linguistic application. Languages appear to vary, genuinely, in allowing optionality in the representation of definiteness, with certain language having grammaticalized a formal, strict requirement that definiteness be expressed by overt means (English, French, Cantonese etc), and others allowing for alternations which are syntactically licensed but may be conditioned by additional factors relating to discourse structure (Mandarin, Korean, and Hindi). As work in this area continues in the future, inspired by Jenks (2018) paper, we believe it will be interesting to see what further variation may be found in other languages without articles when comparable patterns of data are examined, and whether bare nouns might even turn out to be cross-linguistically more common than demonstrative patterns in the representation of anaphoricity.

## References

- Arkoh, Ruby & Matthewson, Lisa. 2013. A familiar definite article in Akan. *Lingua* 123: 1–30. <https://doi.org/10.1016/j.lingua.2012.09.012>
- Bisang, Walter. 1993. Classifiers, quantifiers and class nouns in Hmong. *Studies in Language* 17. 1–51. <https://doi.org/10.1075/sl.17.1.02bis>
- Chao, Yuen-Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Chen, Ping. 1986. Referent introducing and tracking in Chinese narratives. PhD dissertation, UCLA.
- Chen, Ping. 2004. Identifiability and definiteness in Chinese. *Linguistics* 42:1129–1184. <https://doi.org/10.1515/ling.2004.42.6.1129>
- Cheng, Lisa L-S. & Sybesma, Rint. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30: 509–542. <https://doi.org/10.1162/002438999554192>
- Cheng, Lisa L-S. & Sybesma, Rint. 2005. Classifiers in four varieties of Chinese. In *The Oxford Handbook of Comparative Syntax*, Richard Kayne and Guglielmo Cinque (eds), 259–292. Oxford: Oxford University Press.
- Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6:339–405. <https://doi.org/10.1023/A:1008324218506>
- Christensen, Matthew. 2000. Anaphoric reference in spoken and written Chinese narrative discourse. *Book of Chinese Linguistics* 28:2:303–336.
- Dayal, Veneeta. 2004. Number marking and (in)definiteness in kind terms. *Linguistics and Philosophy* 27:393–450. <https://doi.org/10.1023/B:LING.0000024420.80324.67>
- Dayal, Veneeta. 2011. Bare noun phrases. In *Semantics: An international handbook of natural language meaning*, Claudia Maienborn, Klaus von Heusinger, and Paul Portner (eds), 1088–1109. Berlin: Mouton de Gruyter.
- Diesing, Molly. 1992. *Indefinites*. Cambridge, Mass: MIT Press.
- Ebert, Karen. 1971a. Referenz, Sprechsituation und die bestimmten Artikel in einem Nordfriesischen Dialekt (Fering). [Reference, speech situation, and the definite article in a north Frisian dialect (Fering)] *Studien und Materielen* 4, Bredstedt: Nordfriisk Institut.
- Ebert, Karen. 1971b. Zwei Formen des bestimmten Artikels. [Two forms of the definite article] In *Probleme und Fortschritte der Transformationsgrammatik*, [Problems and progress in transformational grammar] Dieter Wunderlich (ed), 159–74, Munich: Hueber.
- Erteschik-Shir, Nomi. 2007. *Information structure: The syntax-discourse interface*. Oxford: Oxford University Press.
- Givón, Talmy. 1983. *Topic continuity in discourse: quantitative cross-linguistic studies*. New York: John Benjamins. <https://doi.org/10.1075/tsl.3>
- Givón, Talmy. 1984. *Syntax: a functional-typological introduction*. Amsterdam: John Benjamins. <https://doi.org/10.1075/z.17>
- Jaggar, Philip. 1985. Factors governing the morphological coding of referents in Hausa narrative discourse. PhD dissertation, UCLA.
- Jenks, Peter. 2018. Articulated definiteness without articles. *Linguistic Inquiry*. 49:501–536. [https://doi.org/10.1162/ling\\_a\\_00280](https://doi.org/10.1162/ling_a_00280)
- Jiang, Li. 2012. Nominal arguments and language variation. PhD dissertation, Harvard University.
- Jiang, Li. 2014. A parametric analysis of nominal arguments. In *Chinese syntax in a cross-linguistic perspective*, Audrey Li, Andrew Simpson and Dylan Tsai (eds), 51–72. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199945658.003.0002>

- Kang, Arum, & Park, Myung-Kwan. 2020. (In)definiteness in Korean without an article. *Korean Book of Linguistics* 45, 451–475.
- Lambrecht, Knud. 1994. *Information structure and sentence form*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511620607>
- Li, Charles & Thompson, Sandra. 1979. Third person pronouns and zero anaphora in Chinese discourse. In *Syntax and semantics 12: Discourse and syntax*, Talmy Givon (ed), 311–335. New York: Academic Press.
- Li, Charles & Thompson, Sandra. 1981. *Mandarin Chinese: a functional reference grammar*. Berkeley: University of California Press.
- Li, Xuping & Bisang, Walter. 2012. Classifiers in Sinitic languages: from individuation to definiteness-marking. *Lingua* 122:335–55. <https://doi.org/10.1016/j.lingua.2011.12.002>
- Löbner, Sebastian. 2011. Concept types and determination. *Book of Semantics* 28: 279–333. <https://doi.org/10.1093/jos/ffq022>
- Ortmann, Albert. 2014. Definite article asymmetries and concept types: semantic and pragmatic uniqueness. In *Frames and concept types*, Thomas Gamerschlag, Doris Gerland, Rainer Osswald, & Wiebke Petersen (eds), 293–321. New York: Springer.0. [https://doi.org/10.1007/978-3-319-01541-5\\_13](https://doi.org/10.1007/978-3-319-01541-5_13)
- Pu, Ming-Ming. 1995. Anaphoric patterning in English and Mandarin narrative production. *Discourse processes* 19:279–300. <https://doi.org/10.1080/01638539509544918>
- Pu, Ming-Ming. 2011. *Discourse anaphora: a cognitive functional approach*. Munich: Lincom Europa.
- Schwarz, Florian. 2009. *Two types of definites in natural language*. PhD dissertation, University of Massachusetts, Amherst.
- Schwarz, Florian. 2013. Two types of definites cross-linguistically. *Language and Linguistics Compass*, 534–558.
- Simpson, Andrew. 2017. Bare classifier/noun alternations in the Jinyun (Wu) variety of Chinese and the encoding of definiteness. *Linguistics* 55:2: 305–332.. <https://doi.org/10.1515/ling-2016-0041>
- Simpson, Andrew, Soh, Hooi Ling, & Nomoto, Hiroki. 2011. Bare classifiers and definiteness: a cross-linguistic investigation. *Studies in Language* 35:168–93. <https://doi.org/10.1075/sl.35.1.10sim>
- Simpson, Andrew & Biswas, Priyanka. 2016. Bare nominals, classifiers, and the representation of definiteness in Bangla. *Linguistic Analysis* 40:3–4:167–198.
- Simpson, Andrew, Li, Yan, & Wu, Zoe. 2016. Grammatical roles, Coherence Relations, and the interpretation of pronouns in Chinese. *Lingua Sinica* 2:2: 1–20. <https://doi.org/10.1186/s40655-016-0011-2>
- Tai, James H.-Y. 1978. Anaphoric restraints in Mandarin Chinese narrative discourse. In *Anaphora in discourse*, John Hinds (ed), 279–338. Edmonton, Alberta: Linguistic Research.
- Yang, Rong. 2001. *Common nouns, classifiers, and quantification in Chinese*. PhD dissertation, Rutgers University.