

# Updating the typology of definiteness: Evidence from bare nouns in Shan

Mary Moroney  
Cornell University  
mrm366@cornell.edu

Accepted to *Glossa* on October 30, 2020  
Word count: 14694

**Abstract** This paper demonstrates that bare nouns in Shan (Tai-Kadai) can express both unique and anaphoric definiteness, a distinction first noted by [Schwarz \(2009\)](#). This pattern of data as well as similar patterns found in Serbian and Kannada motivate adding a category to the typology of definiteness marking described by [Jenks \(2018\)](#) to include languages that allow bare nouns to express anaphoric definiteness. An extended type-shifting analysis offers an account of the availability of bare noun anaphora, as well as the other bare noun interpretations, such as the indefinite, generic, and kind interpretations. Variation in the use of bare nouns versus more marked anaphoric expressions is tied to pragmatic factors such as what anaphoric definite expressions are available in the language, contrast, information structure, and potential ambiguity of nominal expressions. This account proposes that a constrained semantic account of the interpretations of nominal expressions combined with a pragmatic account of their use can model much of the cross-linguistic data on definiteness.

**Keywords:** bare nouns; definiteness; uniqueness; familiarity; anaphora; type shifting

## 1 Introduction

Bare nouns in languages without overt articles can be interpreted as indefinite, definite, generic, or as a kind (see, e.g., Mandarin; [Yang 2001](#)). Therefore, it is not surprising that bare nouns in Shan, a Southwestern Tai (Tai-Kadai) language of Myanmar which lacks overt articles, can have an indefinite, definite, generic, and kind interpretation, as shown in (1).<sup>1</sup> The bare noun in (1a) can have either a definite or indefinite interpretation

---

<sup>1</sup>Data for this paper comes from the author's fieldwork unless otherwise noted. This data has been collected by working primarily with two Shan speakers from Southern Shan State, Myanmar. One comes from Keng Tawng in Shan State. She has lived in Thailand for over 10 years, and speaks Shan, Thai, and English. The other speaker is from Langkho who has lived in the U.S. in Jacksonville, Florida for approximately 7 years. He speaks Shan, Thai, Burmese, and English. Data was collected either in Thailand or by working remotely. The elicitation methods used include translation of stories based on storyboards and felicity judgments on grammatical sentences in specific contexts, following techniques described in [Bochnak & Matthewson 2015](#). Elicitation sessions were conducted in English, with Thai used to clarify vocabulary.

depending on whether there is already an established dog (or dogs) in the context. (1b) is identical to (1a) except that (1b) lacks the imperfective aspect marker that makes an object-level definite or indefinite reading more salient. (1b) is more likely to get a generic interpretation of the bare noun. (1c) has a kind-level predicate and so the bare noun is only compatible with a kind-level reading. This paper focuses on the expression of definiteness in Shan, though the proposed analysis will allow for all these interpretations.

### (1) SHAN BARE NOUN INTERPRETATIONS

- |  |  |
|--|--|
| <p>a. <b>mǎa</b> hàw jù.<br/>dog bark IPFV<br/>‘Dogs are barking.’ <i>indefinite</i></p> | <p>b. <b>mǎa</b> hàw.<br/>dog bark<br/>‘Dogs bark.’ <i>generic</i></p>                                       |
| <p>‘The dog(s) is/are barking.’ <i>definite</i></p>                                      | <p>c. <b>mǎa</b> mət.wáaj.hǎaj kwàa jâw.<br/>dog disappear go finish<br/>‘Dogs are extinct.’ <i>kind</i></p> |

Based on data from German, Schwarz (2009) proposed splitting ‘definiteness’ into two types: uniqueness and familiarity. Uniqueness means uniqueness within a non-linguistic context, and familiarity refers to discourse anaphora. Building on this proposal, Jenks (2015), (2018) argued that bare nouns in Thai and Mandarin only express one kind of definiteness: unique definiteness. He offers the typology of definiteness in Table 1.

**Table 1:** Typology of Definiteness Marking (adapted from Jenks 2018).

	Both marked		One marked	
	same	different	unique	anaphoric
Unique ( $l$ )	Def	Def <sub>weak</sub>	Def <sub>weak</sub>	$\emptyset$
Anaphoric ( $l^x$ )	Def	Def <sub>strong</sub>	$\emptyset$	Def <sub>strong</sub>
Languages	Cantonese, English	German, Lakhota	(unattested)	Mandarin, Akan, Wu

I propose the following definitions of the categories in the typology. The original names of the categories by Jenks (2018) are given in parentheses.

### (2) THE TYPOLOGY OF DEFINITENESS MARKING

- i. Both marked, same (GENERALLY MARKED): The primary strategy of definiteness marking is used in both unique and anaphoric contexts.
- ii. Both marked, different (BIPARTITE): Unique and anaphoric definiteness are marked using different definiteness marking in at least some contexts.
- iii. One marked, uniqueness (MARKED UNIQUE): Unique definiteness is obligatorily marked but anaphoric definiteness is not. (*unattested*)
- iv. One marked, anaphora (MARKED ANAPHORIC): Unique definiteness is not marked, but anaphoric definiteness marking is obligatory.

In this typology some languages, like German and Lakhota, differentially mark unique and anaphoric definiteness. Some languages, like Mandarin (Jenks 2018), Akan (Arkoh

& Matthewson 2013), and Wu (Li & Bisang 2012; Simpson 2017), only mark anaphoric definiteness. English marks both types of definiteness using the same morpheme, meaning that *the* is compatible with both unique and anaphoric definiteness. Jenks (2018) claimed that there are no languages attested where only unique definiteness is morphologically marked.

This paper uses fieldwork with Shan, an under-studied Southwestern Tai language of Myanmar, to explore the expressions of definiteness using bare nouns and makes the claim that Shan bare nouns can express both unique and anaphoric definiteness. This means that there should be one more category in Jenks's (2018) typology of definiteness, a category with 'unmarked' unique and anaphoric definiteness. It is expected that other languages would fall in this unmarked category and use bare nouns to express both unique and anaphoric definiteness. This paper provides evidence that Shan falls into this category and that other languages, such as Serbian and Kannada, do as well.

Section 2 discusses the unique and anaphoric types of definiteness identified by Schwarz (2009), primarily looking at data from Thai as discussed by Jenks (2015). Section 3 introduces data demonstrating that bare nouns in Shan, a language related to Thai, can express both unique and anaphoric definiteness. Section 4 discusses the typology of definiteness marking adding data from Serbian and Kannada. Section 5 presents a type shifting analysis for bare nouns in Shan that uses two  $\iota$  type-shifting operators and identifies a problem with using the Consistency test introduced by Dayal (2004) to decide whether a word counts as a determiner for the Blocking Principle. Additionally, this section proposes to use the economy principle *Don't Overdeterminate!* from Ahn (2019) to explain the variation in choosing bare nouns or other anaphoric expressions within a language. Section 6 discusses some remaining issues connected to the typology of definiteness marking, including the role of contrast and ambiguity, and Section 7 concludes.

## 2 Two kinds of definiteness: Unique and anaphoric

### 2.1 Background

Analyses of definiteness have tried to represent the semantics in terms of uniqueness (Frege 1892; Russell 1905; Strawson 1950) and familiarity (Heim 1982). Schwarz (2009) proposed that instead there are two types of definiteness, unique and anaphoric, as had been suggested by, for example Kadmon (1990) and Roberts (2003). This can be seen overtly in how definiteness is expressed in German. In German, certain preposition-definite article constructions can either appear as two words in a full form or as one word in a reduced form, combining the preposition and definite article. For example, *von* is the reduced form and *von dem* is the full form of the preposition and determiner combination meaning 'by the'. Schwarz (2009) claimed that the reduced form, called the *weak* form, and the un-reduced form with the preposition + definite article combination, called the *strong* form, overtly represent unique and anaphoric definiteness, respectively.

The difference between these two forms can be seen in example (3). This example involves a unique definite context since it is common ground that there is only one mayor in this context. The uniqueness of the mayor in the situation triggers the use of the weak

definite article form, *vom* ('by the'), and the strong form, *von dem*, is infelicitous. In contrast, the strong form is obligatorily used in familiar/anaphoric contexts.

(3) WEAK VERSUS STRONG ARTICLES IN GERMAN

Der Empfang wurde **vom** / #**von dem** Bürgermeister eröffnet.  
 the reception was by-the<sub>weak</sub> / by the<sub>strong</sub> mayor opened  
 'The reception was opened by the mayor.' (Schwarz 2009: (42))

Building on the categories of definiteness described by Hawkins (1978), Schwarz (2009) identified several contexts for definite expressions: immediate situation (local non-linguistic context), larger situation (global non-linguistic context), anaphoric/familiar, and bridging (associative anaphora). To these, Schwarz (2009) added a category for donkey anaphora. According to Schwarz (2009), German uses the weak form of the definite article in the contexts of situational uniqueness, which includes uniqueness in an immediate situation or larger situation, as well as in a type of bridging called part-whole bridging. The strong form of the definite article is used in anaphoric contexts, producer-product bridging, and donkey anaphora—which Schwarz (2009) argued all involve a kind of anaphora. Table 2 gives examples of these categories and the article form used for German. Most of these will be discussed more in section 2.2. This paper will not include the evidence from bridging anaphora other than to note in the tables that it follows the predictions of Schwarz (2009).

**Table 2:** Types of definiteness described by Schwarz (2009), citing Hawkins (1978).

Type of Definite Use	Example	German
Unique in immediate situation	the desk (uttered in a room with exactly one desk)	weak
Unique in larger situation	the prime minister (uttered in the UK)	weak
Anaphoric	John bought a book and a magazine. The book was expensive.	strong
Bridging: Producer-product	John bought a book today. The author is French.	strong
Bridging: Part-whole	John was driving down the street. The steering wheel was cold.	weak
Donkey anaphora	Every farmer who owns a donkey hits the donkey	strong

## 2.2 Uniqueness versus familiarity/anaphoricity

Schwarz (2009) argued that the weak definite article expresses *unique definiteness*. This means that the intended referent is unique in an immediate situation, as in (4), or in a larger or global context. In (4), there is only one glass cabinet in the immediate context, so it is expressed using the weak form of the definite article, *im*.

## (4) GERMAN: UNIQUE IN IMMEDIATE SITUATION

Das Buch, das du suchst, steht **im** / **#in dem** **Glasschrank**.  
 the book that you look-for stands in-the<sub>weak</sub> / in the<sub>strong</sub> glass-cabinet

‘The book that you are looking for is in the glass-cabinet.’ (Schwarz 2009: (40))

Schwarz (2009) proposed that the strong definite article, in contrast, expresses *familiarity* or *anaphoricity*. This includes either being perceptually or generally familiar, or part of the preceding discourse. (5) provides an example with discourse anaphora. The second sentence must use the strong form of the definite article, *von dem Politiker* (‘from the politician’), to refer back to the politician that was introduced in the first sentence.

## (5) GERMAN: ANAPHORA

Hans hat einen Schriftsteller und **einen Politiker** interviewt. Er hat  
 Hans has a writer and a politician interviewed He has  
**#vom** / **von dem** **Politiker** keine interessanten Antworten  
 from-the<sub>weak</sub> / from the<sub>strong</sub> politician no interesting answers  
 bekommen.  
 gotten

‘Hans interviewed a writer and a politician. He didn’t get any interesting answers from the politician.’ (Schwarz 2009: (23))

This phenomenon is not limited to German. Schwarz (2013) found that the strong/weak contrast is apparent in many languages. Jenks (2015), (2018) showed that Mandarin and Thai use bare nouns in the same places where German would use the weak definite article and phrases with a noun, classifier, and demonstrative (N Clf Dem) where German would use the strong definite article. I will be referring to expressions that include a noun and demonstrative, with or without a classifier, as ‘demonstrative-noun phrases’.

Examples (6) and (7) show the use of the bare noun in situations of unique definiteness in Mandarin and Thai, respectively. *Gou* ‘dog’ in (6) and *mǎa* ‘dog’ in (7) must refer to the unique dog or dogs in the context.

## (6) MANDARIN: UNIQUE IN IMMEDIATE SITUATION

**Gou** yao guo malu.  
 dog want cross road

‘The dog(s) want to cross the road.’ (Jenks 2018: (31), Cheng & Sybesma 1999)

## (7) THAI: UNIQUE IN IMMEDIATE SITUATION

**mǎa** kamlan hǎw.  
 dog PROG bark

‘The dog is barking.’ (Jenks 2015: (2))

(8) and (9) show how the demonstrative-noun phrase is used to express discourse anaphora in Mandarin and Thai. In the Mandarin example in (8a), a boy and a girl are first introduced into the discourse context. Then in (8b) and (8c), the demonstrative-noun

phrase *na ge nasheng* ‘the/that boy’) refers back to the boy. According to Jenks (2018), bare nouns in Mandarin can express anaphoric definiteness in certain contexts, namely in the subject position. The classifier and demonstrative are optional in subject position, but not in object position, as shown in (8b) and (8c). Jenks (2018) claimed that this is because the Mandarin subject is a topic, and topic marking negates the effect of *Index!*, the requirement that an indexical expression be used whenever possible. This will be discussed further in section 5.2. The connection between topics and using bare nouns or weak definites to express anaphora has been noted before. For example, German can use a weak definite article to refer back to a referent that is a topic (Schwarz 2009: 47).

(8) MANDARIN: NARRATIVE SEQUENCE (ANAPHORIC)

- a. Jiaoshi li zuo-zhe **yi ge nansheng** he **yi ge nüsheng**,  
classroom inside sit-PROG one CLF boy and one CLF girl

‘There is a boy and a girl sitting in the classroom...’

- b. Wo zuotian yudao **na ge nansheng**  
I yesterday meet that CLF boy

‘I met the boy yesterday.’

- c. **na ge nansheng** kanqilai you er-shi sui zuoyou.  
that CLF boy look have two-ten year or-so

‘The boy looks twenty-years-old or so.’

(Jenks 2018: (15a,b,d))

Example (9) from Thai demonstrates anaphoric reference across two sentences. The first sentence introduces a student into the discourse context using *nákrian khon nīj* ‘one student’. In (9a), the demonstrative-noun phrase *nákrian khon nán* ‘that boy’) refers back to that student. In order to have the anaphoric reading in Thai, Jenks (2015) says the demonstrative construction is required even in subject position. In this way, Mandarin and Thai differ in their choices of anaphoric expressions.

(9) THAI: NARRATIVE SEQUENCE (ANAPHORIC)

mâawaan phǒm cəə kàp **nákrian khon nīj**.  
yesterday 1ST meet with student CLF INDEF

‘Yesterday I met a student’

- a. **nákrian** **na khon nán** chalàat mâak.  
student CLF that clever very

‘That student was very clever.’

(Jenks 2015: (17))

## 2.3 Donkey anaphora

Previous work has shown that when a quantified nominal expression is referred back to anaphorically, languages employ the strong definite form—a strong definite article in German (Schwarz 2009) or a demonstrative-noun phrase in Thai (Jenks 2015) and Mandarin (Jenks 2018). Donkey anaphora provides a case of quantificational anaphora. Typically, a discourse referent is introduced in a relative clause with a universally quantified head or in an if-clause. Then, the discourse referent is referred to again in the matrix clause.



In (10), a referent is introduced by *khwaai tua niŋ* ‘one buffalo’ in the universally quantified relative clause, and a demonstrative-noun expression must be used to refer back to the buffalo that each farmer has. Using a bare noun to refer back to the buffalo gives the sentence a generic meaning: ‘Every farmer that has a buffalo hits buffalo’.

(10) THAI: DONKEY ANAPHORA

chaawnaa thúk khon thîi mii **khwaai tua niŋ** tii **khwaai tua nán**  
farmer every CLF that have buffalo CLF INDEF hit buffalo CLF that

‘Every farmer that has a buffalo hits [that buffalo].’ (Jenks 2015: (23))

As demonstrated here, there is a contrast between unique and anaphoric definiteness in Thai and Mandarin, where unique definites are expressed with bare nouns and anaphoric definites are expressed with demonstrative-noun phrases, with the exception of Mandarin subjects. In Table 3 is a summary of the required definite expressions in certain contexts in German, Thai, and Mandarin. Examples of all the contexts described by Schwarz (2009) for all three languages can be found in the cited sources.

**Table 3:** Expressions of definiteness in German, Thai, and Mandarin.

Type of Definite Use	German (Schwarz 2009)	Thai (Jenks 2015)	Mandarin (Jenks 2018)
Immediate situation	weak	bare	bare
Larger situation	weak	bare	bare
Anaphoric	strong	dem.	dem.
Bridging: Producer-product	strong	dem.	dem.
Bridging: Part-whole	weak	bare	bare
Donkey anaphora	strong	dem.	dem.

### 3 Two kinds of definiteness with Shan bare nouns

This section will demonstrate that while expressions of definiteness in Shan are sensitive to the unique-anaphoric definiteness distinction, as described by Schwarz (2009) and Jenks (2015), (2018), Shan bare nouns can express both unique and anaphoric definiteness.

#### 3.1 Background on Shan

Shan is a Southwestern Tai language in the Tai-Kadai family. The language is spoken in Myanmar and surrounding countries by approximately 3 million speakers (Eberhard et al. 2019). Shan is an analytic, rigid-SVO language, as shown in (11). In existential contexts, such as (12), the bare noun is underspecified for plurality, suggesting Shan is a number neutral language. As has been similarly noted in Thai (Jenks 2011), bare nouns in Shan can refer to either singular or plural entities.

- (11) háw hǎn mǎa  
1 see dog  
'I see a dog/dogs.'

- (12) tinaj mí mǎa  
here have dog  
'Here there is a dog/are dogs.'

When a numeral combines with a noun, a classifier must also appear with the phrase, as in (13)–(14). The same is true for mass nouns: when there is a numeral combining with the noun, there must also be a measure word, as in (15)–(16). Classifiers vary based on properties of the noun they combine with. The classifier for animals is *tǒ*, as in (13)–(14), and the classifier for humans is *kô*.

- (13) mǎa nuŋ \*(tǒ)  
dog one CLF.ANML  
'one dog'

- (14) mǎa sǎam \*(tǒ)  
dog three CLF.ANML  
'three dogs'

- (15) nâm nuŋ \*(kók)  
water one cup  
'one cup of water'

- (16) nâm sǎam \*(kók)  
water three cup  
'three cups of water'

Although some languages allow the classifier to appear alone with the noun (see [Simpson et al. 2011](#)), Shan does not allow this, as shown in (17). Classifiers in Shan seem to be derived from nouns. For example, as (18) shows, *tǒ*, the classifier for animals, is also the word for 'body'.

- (17) N CLF  
\*mǎa tǒ  
dog CLF.ANML  
*intended*: 'the dog'

- (18) CLF N  
tǒ mǎa  
CLF.ANML dog  
'dog body', *not* 'the dog'

The demonstrative in Shan can either appear directly with the noun, as in (19), or with a classifier, as in (20). The difference between (19) and (20) in meaning is that (19) can refer to singular and plural definite objects, and (20) can only refer to a singular definite object.

- (19) N DEM  
mǎa nân  
dog that  
  
'that dog/those dogs'

- (20) N CLF DEM  
mǎa tǒ nân  
dog CLF.ANML that  
  
'that dog'

(21) gives the plural version of (20). As (22) shows, the plural classifier *tsʰ* means 'group'.

- (21) N PL DEM  
mǎa tsʰ nân  
dog CLF.PL that  
  
'those dogs'

- (22) PL N  
tsʰ mǎa  
CLF.PL dog  
  
'the group of dogs'



Shan patterns along the lines of many classifier languages in that it can have bare noun arguments with a variety of interpretations. Demonstrative phrases are overtly marked definite expressions and may include a classifier or not. Numerals do not seem to have a definite interpretation. The following section examines the available definite interpretations of bare nouns.

### 3.2 Uniqueness versus familiarity/anaphoricity in Shan

Shan patterns like Mandarin and Thai in that in unique situations, such as are shown in (23) and (24), a bare noun must be used. In (23), there is only one teacher in the context, so that teacher must be referred to using a bare noun. For (24), we know that there is only one sun in our global context. Consequently, a bare noun must be used to refer to the sun. The demonstrative is not allowed in either case. In these examples, the entity described by the bare noun (i.e., the teacher or sun) has not been mentioned before in the linguistic context, so these examples do not involve discourse anaphora.

(23) SHAN: UNIQUE IN IMMEDIATE SITUATION

(Context: classroom with just one teacher)

Náaŋ Lǎn ʔám tsaan kwàa hǎa **khúsǎn** (#kô      nân).

Ms. Lun NEG able go find teacher CLF.HUM that

‘Ms. Lun cannot find the teacher.’

(24) SHAN: UNIQUE IN LARGER SITUATION

**kǎaŋwán** (#hòj      nân) lǒŋ hɿ sòŋ.

sun CLF.RND that very bright glitter

‘The sun is very bright.’

(Speaker comment on the demonstrative: there is more than one sun)

In contrast to Mandarin and Thai, Shan can use a bare noun in anaphoric contexts. The narrative sequence in (25) demonstrates this. In the first sentence a man is introduced into the discourse context. Another individual, the store owner, is also introduced in the story in order to make *phu-tsáaj* ‘man’ a more natural anaphoric expression to use. Schwarz (2009) and Simpson et al. (2011) use a similar strategy when looking into anaphoric examples. This is because if a speaker introduces only one individual, she is much more likely to refer back to that individual using a pronoun, rather than a nominal expression. This will be discussed more in section 5.4. In (25), since at least two discourse referents are available, anaphoric reference to the man is made either with a bare noun, *phu-tsáaj* ‘man’, or with a demonstrative-noun phrase (N Clf Dem), *phu-tsáaj kô nân* ‘that man’.

## (25) SHAN: NARRATIVE SEQUENCE (ANAPHORA)

**phu-tsáaj kô nuŋ** kwàa ti hân khăaj mǎa tàa sũ mǎa ?òn  
 person-man CLF.HUM one go at store sell dog for buy dog small  
 tǒ nuŋ pǎn lukjín mán-tsáaj... **phu-tsáaj (kô nân)** khúŋ  
 CLF.ANML one give daughter 3-man person-man CLF.HUM that back  
 tǒp waa,  
 respond that

‘A man went to a dog store to buy a puppy for his daughter... The/that man replied, ...’

An example with inanimate referents can be seen in (26). In the first clause, three inanimate things are introduced: *kók kòfi* ‘a cup of coffee’, *phŷn* ‘a table’, and *pâplik* ‘a book’. In the second clause, the cup of coffee and the book can be referred back to using a bare noun or a demonstrative-noun phrase.<sup>2</sup>

## (26) SHAN: CROSS-CLAUSAL ANAPHORA

jôn **kók kòfi** mí nŷ phŷn mí hímtsǎm **pâplik** le, háw laj ?ǎw **kók**  
 because cup coffee exist on desk exist near book and 1 get take cup  
**kòfi (nân)** he sàj **pâplik (nân)**  
 coffee that spill in book that

‘Since a cup of coffee was on the table near a book, I spilled the/that coffee on the/that book.’

Previous work has proposed that contrast can license bare noun anaphora (Jiang 2012; Jenks 2018). In particular, Jenks (2018) discusses that contrastive topics (Büiring 2003) license bare noun anaphora in Mandarin. These contrastive contexts occur when there is ‘a topical set of alternatives relevant to a particular QUD’ (Jenks 2018: 526). This is different from Ahn’s (2019) discussion of contrast in connection with the use of bare nouns in languages where a morphologically simplex pronoun otherwise blocks bare noun anaphora. In such languages, the presence of multiple salient entities that would be referenced using the same pronoun, creates a situation where a bare noun can be used since a pronoun would not unambiguously identify the intended referent. Some might think that (26) only allows bare noun anaphora because there are two contrastive salient entities. The notebook and cup of coffee are not contrastive topics in the sense of Büiring (2003) because Shan uses specific morphemes to indicate contrastive topics.

It is possible to see bare noun anaphora in contexts where there is only one salient, animate entity, as in (27). This example further involves an anaphoric expression in object position. In (27), the classifier and demonstrative are optional. In this example, the squirrel is the only significant individual introduced into the narrative, even though

<sup>2</sup>While it is possible to use a demonstrative-noun phrase for either noun being used anaphorically, it is awkward to use the demonstrative with both at the same time. The effect of using the demonstrative here seems to be to emphasize the noun it combines with, so it may seem unnatural to emphasize both.

other objects, such as the tree and the storm, are discussed.<sup>3</sup> The anaphoric reference to the squirrel in (27) does not occur in a highly contrastive context, given that it is the only animate, third-person entity in the narrative. There also certainly is no contextually salient set of alternatives that are relevant to a QUD. This supports the idea that in Shan the bare noun anaphora is not licensed only by contrast in the sense of Jenks (2018).

(27) SHAN: NARRATIVE SEQUENCE (ANAPHORA)

**tsɔn**    **tõ**            **nuŋ** máa   mí   nǎ tonmâj ʔǎn   mí   hímtsǎm hǎn   háw  
squirrel CLF.ANML one   come exist on tree   COMP exist near   house 1  
nân sě   mǎnâj phǒn lóm   haan       hâaj nàa lɛ   hét haj   kɪŋ-mâj  
that and today   rain   wind appearance bad very and do   cause branch-tree  
jàat   tók njǎa   tẽ   phât njǎa **tsɔn**   (**tõ**        **nân**) pǎa   jâw  
break fall almost IRR hit   meet squirrel CLF.ANML that   with finish

‘A squirrel was on the tree near my house. Then one day, a bad storm caused a branch to break and fall almost hitting the squirrel.’

This data demonstrates that Shan bare nouns can express both unique and anaphoric definiteness, and this possibility is not constrained to subjects as has been noted for Mandarin. Shan definiteness follows the patterns predicted by the distribution of unique and anaphoric definiteness described by Schwarz (2009) for German and Jenks (2018) for Mandarin. Demonstrative-noun phrases are used in anaphoric definite contexts, but bare nouns can be used in both types of definite contexts.

### 3.3 Donkey anaphora

German, Thai, and Mandarin all use a strong/demonstrative form to refer back to a nominal in cases of donkey anaphora. However, Shan does not obligatorily use a demonstrative or strong definite article to express donkey anaphora. (28) gives an example of donkey anaphora in the second sentence. This example uses the conditional construction. The antecedent clause in (28) introduces *měw* ‘cat’ and *nǎ* ‘mouse’ using bare nouns and in the consequent the same bare nouns refer back to them.<sup>4</sup>

<sup>3</sup>However, it is likely relevant that there are other entities discussed in the example because otherwise, a pronoun or null pronoun might be more natural to use. Animals and non-animate entities are referred to using the same 3rd person pronoun, *mán*, so the pronoun might not uniquely identify the intended referent, the squirrel. This is related to Ahn’s (2019) notion of contrast. The relevance of this will be discussed further in section 5.4.

<sup>4</sup>The anaphoric expressions in donkey anaphora examples in different contexts can be a null pronoun, an overt pronoun, N, N Dem, N Cl Dem, or N Pl Dem. The interpretive differences between them are often subtle.

- (28) SHAN: CONDITIONAL  
 mếw lɛ nũ ʔəm mən kǎn. pố mếw hǎn nũ nǎj-tsũŋ, mếw lɔp lám  
 cat and rat NEG right together if cat see rat then cat follow chase  
 tĩ njóp nũ tàasè  
 grab snatch rat always

‘Cats and rats don’t get along together. If a cat sees a rat, the cat chases and catches the rat always.’

A demonstrative-classifier-noun phrase can be used more naturally as the anaphoric component of a donkey anaphora sentence in a structure like (29), but the demonstrative and classifier are not necessary. In (29), *tổ lǎj* (‘which one’) quantifies over individual cats, and the demonstrative-classifier-noun phrase can refer back to each cat.

- (29) SHAN: DONKEY ANAPHORA  
 mǎa nǎj hǎn mếw tổ lǎj kɔ tẽ lɔp mếw (tổ nân)  
 dog this see cat CLF.ANML which PRT will follow cat CLF.ANML that  
 tàasè.  
 always

‘Dogs, whichever cat they see they will always chase the/that cat’

In donkey anaphora constructions, my consultants did not typically use *mếw tổ nũŋ* ‘one cat’ in the antecedent clause. Shan does not have an indefinite article but uses the numeral ‘one’ in some cases where English uses an indefinite article. If ‘only’ is used, it is possible to use *mếw tổ nũŋ* in the antecedent, and in that case it is necessary to use a demonstrative to refer back to it, as in (30). When ‘one cat’ is used in a donkey anaphora construction, it sounds as if the number of cats is relevant to whether the dog chases the cat, in the same way it does in the English translation of (30).

- (30) SHAN: DONKEY ANAPHORA  
 mǎa ku tổ nǎj pố hǎn mếw tổ nũŋ kój kɔ tẽ  
 dog every CLF.ANML this if/when see cat CLF.ANML one only PRT will  
 lɔp lám mếw \*(tổ nân) tàasè.  
 follow chase cat CLF.ANML that always

‘Every dog, if it sees only one cat, will always chase that cat.’

Table 4 summarizes the range of definite expressions found in German, Thai, Mandarin, and Shan. This section has investigated the expression of definiteness in Shan in different kinds of definite contexts. In Shan, bare noun can be used in all of the contexts described by Schwarz (2009). Demonstrative-noun phrases are compatible with the contexts that require the strong definite form in German, Thai, and Mandarin, but they are not obligatory. The bridging anaphora data is not included here, but the data from Shan is consistent with the pattern discussed so far. Bare nouns can appear in both part-whole and producer-product bridging. Even contexts like cross-sentential and donkey anaphora allow for bare nouns.

**Table 4:** Expressions of definiteness in German, Thai, Mandarin, and Shan.

Type of Definite Use	German (Schwarz 2009)	Thai (Jenks 2015)	Mandarin (Jenks 2018)	Shan (Moroney 2019a)
Immediate situation	weak	bare	bare	bare
Larger situation	weak	bare	bare	bare
Anaphoric	strong	dem.	dem.	<b>bare/dem.</b>
Bridging: Producer-product	strong	dem.	dem.	<b>bare/dem.</b>
Bridging: Part-whole	weak	bare	bare	bare
Donkey anaphora	strong	dem.	dem.	<b>bare/dem.</b>

## 4 A revised typology of definiteness

This section argues for a revised typology of definiteness, adding a category where anaphoric definiteness is unmarked to the typology given above in Table 1. As shown in Section 3, Shan bare nouns can express both unique and anaphoric definiteness. In addition to Shan, there are several languages that could potentially fit into this ‘unmarked’ category, including Serbian (§4.1) and Kannada (§4.2). Section 4.3 revisits some previous work on the typology of definiteness marking to discuss where other languages might fit in the expanded typology. Section 4.4 gives the proposed revised typology.

### 4.1 The case of Serbian

Serbian presents a case where bare nouns can express anaphora even in object position, in contrast to Mandarin. Despić (2019) has previously identified that bare nouns can be used anaphorically in Serbian.

Example (31) demonstrates that a bare noun in object position can refer anaphorically in Serbian. (31) gives an anaphoric bare noun example using an inanimate noun. After the noun, *novčanik* ‘wallet’, is introduced, it is referred back to several times in the story using a bare noun. The first time it is referred to anaphorically, the bare noun is the object of the verb *uzela* ‘took’. This kind of anaphora using a bare noun is possible with animate nouns as well, such as *policaica* ‘policeman’.

## (31) SERBIAN: CROSS-SENTENTIAL ANAPHORA

Isidora Stojanović, ... pronašla je **novčanik** pun para gde se nalazilo  
 Isidora Stojanovic found is wallet full money where refl. located  
 4000 evra i 16.000 dinara a Isidora se nije dvoumila šta treba  
 4000 euros and 16.000 dinars and Isidora self did-not think\_twice what needs  
 da učini. Uzela je **novčanik** i otišla do MUP-a i predala  
 that do Took is wallet and left to police\_station and turned\_in  
 novac...  
 money

‘Isidora Stojanović, ... found a **wallet** full with money, which contained 4000 euros and 16.000 dinars. Isidora did not think twice what she needs to do. She took **the wallet** and went to the police station and turned in the money...’<sup>5</sup>

Bare nouns can also be used in donkey anaphora examples in Serbian, as shown in (32).<sup>6</sup> In these examples, it is better if there are two possible antecedents for the donkey anaphor, since a pronoun is often preferred to using a bare noun when expressing anaphora.

## (32) SERBIAN: DONKEY ANAPHORA

Moj jednogodišnji sin ponekad radi neočekivane stvari. Recimo, svaki  
 My one-year-old son sometimes does unexpected things. For instance, every  
 put kada vidi **koficu** i **loptu**, on pokušava da stavi **koficu** na **loptu**.  
 time when sees bucket and ball he tries to put bucket on ball.

‘My one-year old son sometimes does unexpected things. For instance, every time he sees a bucket and a ball, he tries to put the bucket on the ball.’

Serbian appears as though it might also fit into the ‘unmarked’ category within the typology of definiteness marking since bare nouns can be used anaphorically in many different contexts.

## 4.2 The case of Kannada

In a recent presentation, [Srinivas & Rawlins \(2020\)](#) demonstrated that Kannada bare nouns express both unique and anaphoric definiteness in much the same way that Shan bare nouns do. For example, (33) demonstrates that the anaphoric noun *pustaka-(d)alli* ‘in the book’ is not functioning as a subject. It is also not simply a unique definite, as there are many other books in the library.

<sup>5</sup>Source: <http://www.orsbap.com/test/rukometasica-sloge-isidora-stojanovic-vratila-novcanik-pun-para/>

<sup>6</sup>(32) is a sentence provided by one linguist native speaker of Serbian and judged as correct by another native speaker.



- (33) KANNADA: CROSS-SENTENTIAL ANAPHORA  
 Bengaloor-(i)na doDDa granthalaya-(d)alli halsinahaNN-(i)na bagge **ondu**  
 Bangalore-GEN big library-LOC jackfruit-GEN about one  
**pustaka** ide. Monne naanu alli-ge hogiddaaga **pustaka**-(d)alli  
 book COP Recently I-NOM there went book-LOC  
 halsinahaNN-anna he:ge kariyadu anta noDide  
 jackfruit-ACC how fry that saw

‘There is a book about jackfruit in the big library in Bangalore. Recently, when I was there, I looked in the book for instructions for frying jackfruit.’

(Srinivas & Rawlins 2020: (9))

Even donkey anaphora sentences, such as (34)-(35) use a bare noun to refer back to the donkey introduced in the first part of the sentence.

- (34) KANNADA: DONKEY ANAPHORA  
**katte-annu** uLLuva pratiyobba raitanuu **katte-ge** uuTa haakuttaane  
 donkey-ACC having every farmer-EMPH donkey-DAT food puts

‘Every farmer who has a donkey feeds the donkey.’

(Srinivas & Rawlins 2020: (7), added emphasis)

- (35) KANNADA: DONKEY ANAPHORA  
 Raita-na hattira **katte** iddare avanu **katte-ge** ooTa haakuttaane  
 Farmer-GEN near donkey have.if he donkey-DAT food puts

‘If a farmer has a donkey, he feeds the donkey.’ (Srinivas & Rawlins 2020: (8), added emphasis)

Two differences between Kannada and Serbian, on the one hand, and Shan, on the other hand, are that Kannada and Serbian both mark case and plurality where Shan does not. Shan instead uses word order to indicate grammatical relationships. Despite these differences, Kannada, Serbian, and Shan appear to have a similar distribution of anaphoric bare nouns.

### 4.3 Filling in the Typology

This section draws on previous work related to the typology of definiteness marking—including work by Simpson et al. (2011), Schwarz (2013), Jenks (2018), Ahn (2019), and Schwarz (2019), and identifies how these languages fit into the typology developed in this paper.

Ahn (2019) discusses several languages that could fall into the unmarked definiteness category. For example, Korean and American Sign Language (ASL) appear to allow bare nouns to express anaphoric definiteness fairly robustly. An example of Korean is shown in (36), here *namca-ka* ‘man’ is introduced as the subject in the first sentence and referred back to with a bare noun *namca-lul* as the object in the second sentence. Case marking accounts for their different forms. In ASL, simple inter-sentential anaphora might suggest that bare nouns cannot be anaphoric as shown in (37). However, (38) shows that a bare

noun is acceptable to use in some cases of anaphoric reference. Korean and ASL would then be languages that do not obligatorily overtly mark anaphoric definiteness.

- (36) **namca**-ka tulewa-ss-ta. na-nun **namca**-lul chyetapwa-ss-ta.  
 man-nom enter-past-decl I-top man-acc stare-past-decl  
 ‘A man entered. I stared at the man.’ (Ahn 2019: (24), added emphasis)
- (37) JOHN BUY **IX<sub>A</sub>** **BOOK**, **IX<sub>B</sub>** MAGAZINE. #(**IX<sub>A</sub>**) **BOOK** EXPENSIVE.  
 ‘John bought a book and a magazine. The book was expensive.’  
 (Ahn 2019: (333), citing Irani 2016; see Irani 2019: (27), added emphasis)
- (38) **BOY<sub>i</sub>** ENTER CLUB. MUSIC ON. **BOY<sub>i</sub>** DANCE.  
 ‘A boy<sub>i</sub> entered a club. Music was on. He<sub>i</sub> danced.’  
 (Ahn 2019: (330), Condition A for N, format and emphasis mine)

#### 4.3.1 One Marked or Unmarked?

The previous sections have discussed specific languages from different language families that have data available to demonstrate that bare nouns can be used to express anaphoric definiteness. There are several other languages, such as those mentioned by Despić (2019)—Japanese, Hindi, and Turkish—that have a bare noun/marked anaphoric definiteness contrast. All of these languages have a way of overtly marking anaphoric definiteness. What is not clear is whether this morpheme is obligatory in cases of anaphoric definiteness. This obligatoriness is what distinguishes languages with an anaphoric definite determiner from ones that do not have any overt definite determiners. This distinction has implications for the range of possible definite interpretations of bare nouns and overt lexical instantiations of unique and/or anaphoric definiteness. The semantic account of bare noun definiteness that will be proposed in section 5 is constrained by the lexical determiners found in the language.

Schwarz (2013) identifies two languages that use bare nouns to express unique definiteness and have a determiner that is only used to express anaphoric definiteness. These languages are Akan and Mauritian Creole. Schwarz (2013) notes that the anaphoric definite article is found largely in the same places as the strong definite article in Germanic languages. In addition to Thai (Jenks 2015) and Mandarin (Jenks 2018)—which will be discussed further in section 5.4, Jenks (2018) also classifies Akan as a language with an anaphoric definite article. Jenks (2018) further identifies Wu (Tibeto-Burman) as such a language.

The data available for Mauritian Creole from Wespel 2008 show that the determiner *la* is used in anaphoric and not unique definite contexts. There is no data that shows that bare nouns can express anaphoric definiteness. The data in (39) demonstrates that the determiner is used even in an apparent case of contrastive focus, which supports treating *la* as an anaphoric definite determiner in Mauritian Creole.

- (39) Enn garson ek enn tifi ti pe lager. **Garson la** ti paret an koler, **tifi**  
 one boy and one girl PST PROG argue boy DEF PST appear in rage girl  
**la** ti res kalm.  
 DEF PST stay calm  
 ‘A boy and a girl were arguing. The boy seemed furious, the girl stayed calm.’  
 (Wespel 2008: p.143)

Amfo (2007) identifies the Akan morpheme, *nó*, as a distal demonstrative determiner, which can also function as a definite determiner and a dependent clause marker. She seems to categorize this morpheme as expressing both unique and anaphoric definiteness. In contrast, Arkoh & Matthewson’s (2013) data on the same morpheme, which in their orthography is *nó*, demonstrates that *nó* is found in familiar/anaphoric contexts. They also report that bare nouns in Akan cannot be used in familiar contexts.

Wu has definite bare nouns, but also has a classifier-noun phrase that can either be definite or indefinite (Cheng & Sybesma 2005). Definiteness is associated with syntactic position, but it also can be marked with a specific tone (Cheng & Sybesma 2005; Sio 2006). Cheng & Sybesma (2005) and Sio (2006) do not discuss the distribution of bare nouns and classifier-noun phrases with respect to the uniqueness/familiarity contrast, but Simpson (2017) demonstrates that the variety of Wu spoken in Jinyun county in Zhejiang province does use the classifier-noun phrase to mark anaphoric definiteness. The data from Arkoh & Matthewson (2013) and Simpson (2017) on Akan and Jinyun Wu support categorizing these languages as marking anaphoric definiteness. If bare nouns do express anaphoric definiteness in these languages, they should instead be characterized as being unmarked for definiteness.

#### 4.3.2 Classifiers as definiteness markers

Several languages use classifiers to express definiteness, but the types of definiteness expressed can vary across languages. Simpson et al. (2011) provided information about expressions of definiteness in Vietnamese, Hmong, Bangla, Hong Kong Cantonese, and Malaysian Cantonese. The languages discussed in Simpson et al. 2011 express definiteness using either bare nouns or classifier-noun phrases (Clf + N).<sup>7</sup> These languages express overt marking of definiteness using classifier-noun phrases instead of demonstrative-noun phrases, which are the definite expressions used in contrast to bare nouns and pronouns in the languages that have already been discussed. While this distinction is important, it is still possible to compare marked versus bare noun definiteness in these languages. Simpson et al.’s (2011) investigation tested several categories of definiteness which overlap with those investigated by Schwarz (2009) and Jenks (2018).

Among the languages Simpson et al. (2011) describe, Hmong seems to pattern clearly as a language that marks both unique and anaphoric definiteness using a classifier-noun phrase. Li & Bisang (2012) note that the Hmong classifier seems to be a highly grammaticalized marker of definiteness, supporting the language’s inclusion in the typological category that marks both kinds of definiteness. The classifier-noun phrase, the overt

<sup>7</sup>Simpson et al. (2011) call these bare classifiers. Some languages allow the classifier to appear alone with a pronominal function (Chuj; Royer 2019), so I am calling them classifier-noun phrases to distinguish them.

marker of definiteness, was preferred in all contexts. The data from Bangla suggest that Bangla is a marked anaphoric language like Akan. The classifier-noun phrase was rated as more grammatical in anaphoric contexts and the bare noun was rated as more grammatical in unique contexts. A more detailed investigation of Bangla largely seems to confirm that (Simpson & Biswas 2016). As for Cantonese, which Jenks (2018) calls a ‘generally marked’ language, the Hong Kong Cantonese speakers are consistent with what one would expect from a language that uses the same definiteness marking for both kinds of definiteness, but the Malaysian Cantonese speakers patterned closer to the Bangla speakers.

Data for the Vietnamese speakers shows that while the classifier-noun phrase is well accepted in all definite environments, the bare noun data is also accepted in most contexts. Only in anaphoric contexts is the bare noun less preferred compared to the classifier-noun phrase. To definitively categorize these languages, it would be important to learn more about the contexts where bare nouns and classifier-noun phrases are preferred. However, Simpson et al. (2011) show clearly that definite expressions like classifier-noun phrases and bare nouns have different statuses across languages.

### 4.3.3 Both Marked

This paper mainly focuses on the contrast between languages that only mark anaphoric definiteness and languages that mark neither unique nor anaphoric definiteness. The other main typological category is the ‘both marked’ category.

English falls into the category where both unique and anaphoric definiteness are marked using the same determiner, *the*. As discussed in section 1, Schwarz (2009) identified that some Germanic languages mark both unique and anaphoric definiteness using separate definite articles. Other languages that express unique and anaphoric definiteness using two distinct overt morphemes include Lakhota and Hausa (Schwarz 2013).

### 4.3.4 Other patterns of definiteness

Some languages are more difficult to fit into the typology. For example, Schwarz (2013) notes that Haitian Creole shows a type of definite contrast different from the unique/anaphoric one. Haitian Creole typically uses the same definite article to express both unique and anaphoric definiteness, but there are certain definite contexts where a determiner is not used.

Optionality has been reported for some languages with overt definite articles, such as in Nuosu Yi (Jiang 2018) and Indonesian (Little & Winarto 2019). This means that there are contexts where either a bare noun or a determiner marked noun are acceptable. Within this typology, true optionality is unexpected of determiners. These languages would be important to consider further.

#### 4.4 Summary: A revised typology

A revised typology of definiteness marking is given in Table 5. This table adds the category ‘Unmarked’ definiteness to the typology and adds Shan, Serbian, and Kannada to that category.

**Table 5:** Revised Typology of Definiteness Marking.

	Both marked		One marked		Unmarked
	same	different	unique	anaphoric	
Unique	Def	Def <sub>weak</sub>	Def <sub>weak</sub>	∅	∅
Anaphoric	Def	Def <sub>strong</sub>	∅	Def <sub>strong</sub>	∅
Languages	Cantonese, English	German, Lakhota	( <i>unatt.</i> )	Akan, Wu, Mauritian Creole	Shan, Kannada, Serbian

Some languages clearly require overt marking of definiteness. These are languages like English. Of the languages that mark definiteness, some—like German—mark unique and anaphoric definiteness differently in some contexts.

Some languages use bare nouns to express unique definiteness and use another morpheme to mark anaphoric definiteness. Of these, there are languages that require anaphoric definiteness to be marked obligatorily. However, there are some languages that can express anaphoric definiteness using bare nouns, such as Shan, Serbian, and Kannada. The following section will build on the analysis of Jenks (2018) to include languages that can express anaphoric definiteness using bare nouns in the typology of definiteness marking.

## 5 Analysis

This section presents a type-shifting analysis of bare nouns, extending a previous account to capture the data in Shan and other languages that fall into the ‘unmarked’ typological category. This discussion includes motivating a semantic rather than pragmatic account of the unique/anaphoric distinction and discussing predictions of this account for definiteness marking connected to kinds. Additionally, this section discusses a pragmatic approach to accounting for the bare noun/overtly-marked definite alternations in definiteness marking.

Section 5.1 gives the background for the type shifting analysis used to derive the interpretations of bare nouns. Section 5.2 extends the analysis proposed by Jenks (2018) that distinguishes between unique and anaphoric definiteness. This allows us to explain why some languages obligatorily mark anaphoric definiteness and some do not. Section 5.3 describes predictions of definiteness marking for the typology based on the type shifting analysis. Section 5.4 discusses how an economy constraint connected to the definite expressions available in a language can account for variation of definiteness marking within languages. Section 5.5 summarizes this section.

## 5.1 Type shifting

A type shifting analysis of bare nouns (Chierchia 1998; Dayal 2004) makes the prediction that languages without overt articles should allow for bare nouns to express indefiniteness, definiteness, generics, and kinds as a result of being able to type shift. Some version of a type-shifting analysis has been used to account for the interpretations of bare nouns in many languages, including Mandarin (Yang 2001), Hindi (Dayal 2004), Nuosu Yi (Jiang 2018), and Teotitlán Del Valle Zapotec (Deal & Nee 2018).

Languages like Shan that lack overt definite determiners frequently use bare noun arguments. Languages like English that have overt determiners use bare noun arguments in a more restricted way. Only English plural and mass nouns can be bare arguments, as shown in (40)-(41).

(40) Dinosaurs are extinct.

(41) Gold is valuable.

Chierchia (1998) proposed a Neo-Carlsonian approach following Carlson (1977), claiming that bare plurals in English are type  $\langle e, t \rangle$ , and they can type shift to function as arguments. The type-shifting operators defined by Chierchia (1998) and updated by Dayal (2004), are given below:

(42) TYPE SHIFTING OPERATORS (Dayal 2004):

- a.  $\cap: \lambda P \lambda s \iota x[P_s(x)]$
- b.  $\cup: \lambda k_{\langle s, e \rangle} \lambda x[x \leq k_s]$
- c.  $\iota: \lambda P \iota x[P_s(x)]$
- d.  $\exists: \lambda P \lambda Q \exists x[P_s(x) \wedge Q_s(x)]$

Languages without determiners should be able to use all type shifting operations. Here, I will be focusing on the definite and kind interpretations of the type shifting analysis. There is some evidence that indefinite readings of bare nouns work somewhat differently. Dayal (2004) uses evidence from Hindi to argue that the type shifting operation that generates the wide-scope indefinite interpretation is not available unless the definite and kind generating operators are unavailable. In addition, there are other approaches to deriving the indefinite reading of bare nouns. These include the Mapping Hypothesis by Diesing (1992), which is connected to Existential Closure, and Restrict by Chung & Ladusaw (2003).

Using Dayal's (2004) update of Chierchia's (1998) analysis, Deal & Nee (2018) discuss the predicted available interpretations for bare nouns in Teotitlán Del Valle Zapotec. These interpretations discussed by Deal & Nee (2018) are all available for Shan bare nouns,<sup>8</sup> so this kind of analysis can account for the bare noun interpretations introduced in example (1), repeated below.

<sup>8</sup>Teotitlán Del Valle Zapotec distinguishes singular and plural nouns, but the same set of interpretations were predicted to be available when the noun is assumed to denote a kind or an  $\langle e, t \rangle$ -type predicate.



## (1) SHAN BARE NOUN INTERPRETATIONS

- a. **mǎa** hàw jù.  
dog bark IPFV  
'Dogs are barking.' *indefinite*
- b. **mǎa** hàw.  
dog bark  
'Dogs bark.' *generic*
- c. **mǎa** mət.wáaj.hǎaj kwàa jâw.  
dog disappear go finish  
'The dog(s) is/are barking.' *definite*  
'Dogs are extinct.' *kind*

This analysis relates to the typology of definiteness marking because it offers some constraints as to which type shifting operators are expected in which languages. The Blocking Principle (Chierchia 1998; Dayal 2004), defined in (43), constrains type-shifting by prohibiting use of covert type-shifting operators that are duplicated by overt determiners in a language. Essentially, if a language has an overt determiner form of a type-shifter, then covert type-shifting using that operator is unavailable. For example, in English *the* is said to correspond to  $\iota$ , which explains why bare nouns in English cannot type shift using  $\iota$ .

- (43) BLOCKING PRINCIPLE, Dayal (2004): For any type shifting operation  $\phi$  and any  $X$ :  
\* $\phi(X)$  if there is a determiner  $D$  such that for any set  $X$  in its domain,  $D(X) = \phi(X)$ .

A previous account using type shifting to capture the interpretations of bare nouns is from Despić (2019). This account specifically discusses anaphoric definiteness, claiming that whether a language uses number marking is relevant for what interpretations are available for bare nouns. Table 6 summarizes the distribution. In a language without definite articles,  $\iota$  can be used to generate a definite interpretation, as expected. However, only languages that mark number can use bare nouns to express anaphoric reference to a kind. This is because the singular count noun can be construed as ranging over the taxonomic domain as proposed by Dayal (2004).

**Table 6:** Languages without definite articles: Bare Nouns (Despić 2019: 28).

	+ Number						–Number			
	Kind-level			Object-level			Kind-level		Object-level	
	Mass	Count		Mass	Count		Mass	Count	Mass	Count
		SG	PL		SG	PL				
<i>Anaphoric</i>	*	✓	*	✓	✓	✓	*	*	✓	✓
<i>Type-Shift</i>	$\cap$	$\iota$	$\cap$	$\iota$	$\iota$	$\iota$	$\cap$	$\cap$	$\iota$	$\iota$

Shan is a language without definite articles or overt plural marking. Therefore, Shan should pattern with the –Number category in Table 6. This is consistent with what has been discussed in Section 3. However, in order to discuss the full typology of definiteness marking, I will extend the table to include a distinction between unique and anaphoric definiteness. This will be discussed in Section 5.3.

## 5.2 Two types of definiteness

Jenks (2018) follows Schwarz (2009) in supporting the existence of two types of definiteness. To account for the obligatory use of the demonstrative in anaphoric definite environments in Mandarin, Jenks (2018) provides the denotations of unique and anaphoric definiteness shown in (44), where (44a) is the type shifting operation  $\iota$  and (44b) is the denotation of the Mandarin demonstrative. Since the English determiner *the* is used in both unique and anaphoric definite contexts, Jenks (2015) proposes that *the* is ambiguous between the unique and anaphoric definite meaning.

- (44) a. UNIQUE DEFINITE ARTICLE: (Jenks 2018: (22))  

$$\llbracket \iota \rrbracket = \lambda s_r. \lambda P_{\langle e, \langle s, t \rangle \rangle}. : \exists ! x [P(x)(s_r)]. \iota x P(x)(s_r)$$
 b. ANAPHORIC DEFINITE ARTICLE:  $\iota^x$   

$$\llbracket \iota^x \rrbracket = \lambda s_r. \lambda P_{\langle e, \langle s, t \rangle \rangle}. \lambda Q_{\langle e, t \rangle}. : \exists ! x [P(x)(s_r) \wedge Q(x)]. \iota x P(x)(s_r)$$

According to Jenks (2018), a bare noun cannot express anaphoric definiteness due to a principle called *Index!*, defined in (45), which says that when an indexical expression is available, it must be used, meaning that a bare noun cannot be used to express anaphoric definiteness. The exception to this comes from Mandarin subject position, where both bare nouns and demonstratives can be used. Jenks claims that Mandarin subjects are topics and this topic marking negates the effect of *Index!*.

- (45) *Index!* (Jenks 2018 : (50))  
 Represent and bind all possible indices.

It is clear from the data in §3 that the Shan demonstrative is not an anaphoric definite determiner since it is not obligatory in all anaphoric contexts, as has been described for Thai. One option is to say that the pragmatic principle, *Index!*, introduced by Jenks (2018), can be over-ruled in Shan and other ‘unmarked’ languages in considerably more contexts than Jenks proposed. However, this seems unsatisfactory if we want the analysis to make strong predictions about what interpretations are available for bare nouns in a given language. Further, Dayal & Jiang (to appear) provide evidence demonstrating that Mandarin does allow bare nouns to express anaphoric definiteness. Instead, I propose that Shan has a null anaphoric type shifter  $\iota^x$  in addition to the  $\iota$  type shifter. These denotations would be the same as those in (44). This ambiguous type-shifting analysis would predict that the anaphoric interpretation of bare nouns should be available in contexts where anaphora is possible since  $\iota^x$  has more presuppositions. *Maximize presupposition* (Heim 1991)<sup>9</sup> would predict that when  $\iota$  and  $\iota^x$  are in competition,  $\iota^x$  should win out whenever there is an indexical property available.

We might at this point ask what the difference is between the Thai and Shan demonstratives such that the Shan demonstrative would not count as a definite determiner but the Thai one would. Using the Consistency test (Dayal 2004, based on Löbner 1985) to distinguish between demonstratives and true definites does not offer an explanation.<sup>10</sup>

<sup>9</sup>*Maximize Presupposition* says to ‘presuppose as much as possible’. Meaning, when choosing between competing expressions where one has more presuppositions than the other, choose the one that has the most true presuppositions.

<sup>10</sup>See more discussion of issues related to the Consistency test in Moroney 2019b.

For determiners like *the*, conjoining two clauses where a determiner-noun phrase is the argument of contradictory predicates, a contradiction results. Shan and Thai examples of the Consistency test with demonstratives are shown in (46) and (47), respectively.

(46) SHAN: CONSISTENCY TEST

(Context: I am holding a white cup and a black cup.)

kók hòj      nâj pěn sǐ      khǎaw. kók hòj      nâj pěn sǐ      lǎm.  
cup CLF.RND this be    color white    cup CLF.RND this be    color black  
'This cup is white. This cup is black.'

(47) THAI: CONSISTENCY TEST

dèk khon nán nɔɔn yùu tèe dèk khon nán mǎi.dâi nɔɔn yùu.  
child CLF    that sleep IPFV but child CLF    that NEG    sleep IPFV

'That child is sleeping but that child is not sleeping.' (cf. #the)

(Jenks 2015: (3), citing Piriawiboon 2010)

Here, both demonstratives pattern like demonstratives in causing no contradiction when using deixis. However, both sound contradictory when uttered out of the blue, when the demonstrative-noun expression in each clause is interpreted anaphorically as referring to the same individual. Anaphoric uses of the demonstrative in English sound similarly contradictory. The Consistency test does not seem to resolve the question of the difference between Thai and Shan. However, the discussion in 5.4 will actually support the idea that Thai and Shan definiteness is not so different.

In this section, I have proposed that Shan bare nouns can express both unique and anaphoric definiteness using type-shifting and that the Shan bare noun/demonstrative contrast parallels the English *the*/demonstrative contrast. Next, I will discuss the effects of this analysis on the predicted typology of definiteness marking.

### 5.3 The full typology

For a given language, it is possible to determine the available nominal expressions. From this, it is possible to infer what, if any, type shifting operations are available. Building on the table from Despić (2019) to include unique definiteness as one kind of definiteness marking, the following tables give the full predictions of this analysis, including information connected to a language's typological category of definiteness marking. Each table represents a different typological category. The tables are organized as follows:

- +/- Number represents whether a language in the typological category morphologically marks singular and plural.
- Object-level/Kind-level distinguishes between nominal expressions that refer to individual entities or to kinds/sub-kinds.
- Mass/Count distinguishes between mass and count nouns.
- For count nouns in the +Number category, singular (SG) and plural (PL) nouns are distinguished.
- Unique/Anaphoric (Anaph.) categorizes the type of definiteness.

- Type-Shift/Overt indicates whether that interpretation can come about through a type shifting operator or an overt determiner.

The assumptions for filling in the table, based on [Despić 2019](#) and [Dayal 2004](#), are the following:

- The available definite type shifting operators are  $\cap$  and  $\iota$ .
- $\cap$  cannot be used for singular nouns.
- $\cap$  cannot be used to express anaphoric definiteness.
- $\iota$  cannot be used if there is an overt determiner possible to express that interpretation.

### 5.3.1 Both Marked

Table 7 gives the predictions for languages that mark both unique and anaphoric definiteness, like English. In these languages, the determiner is required to express both unique and anaphoric definiteness. Bare nouns, both mass and plural, can get a kind level interpretation, but bare nouns cannot be used to refer to kinds anaphorically. This predicts, for example, that plurals can have a unique, kind interpretation either with a bare plural or with a determiner. The determiner relies on the taxonomic reading of the noun and the bare plural can type shift with  $\cap$ .

**Table 7:** Definite bare noun interpretations, both marked, adapted from [Despić 2019](#).

		+ Number						–Number			
		Kind-level			Object-level			Kind-level		Object-level	
		Mass	Count		Mass	Count		Mass	Count	Mass	Count
			SG	PL		SG	PL				
Unique	Type-Shift	$\cap$	* $\cap$ (SG)	$\cap$	-	-	-	$\cap$	$\cap$	-	-
	Overt	det.	det.	det.	det.	det.	det.	det.	det.	det.	det.
Anaph.	Type-Shift	* $\cap$	* $\cap$ (SG)	* $\cap$	-	-	-	* $\cap$	* $\cap$	-	-
	Overt	det.	det.	det.	det.	det.	det.	det.	det.	det.	det.

In a language where both unique and anaphoric definiteness are marked differently, like German, the predictions would be identical to those found in Table 7 except that the determiner form would vary depending on whether unique or anaphoric definiteness was being expressed. For example, with weak definite articles, both singular and plural nouns can have a kind reference. This the *Unique-Overt* row in the *Kind-Count* columns in the above table. [Schwarz \(2009\)](#) confirms that this is the case for singular nouns in German (p. 65, (66)) as well as in Fering, a Germanic language with a weak/strong definite determiner contrast that appears with both singular and plural nouns (p. 66, (67)). We would also expect that anaphoric reference to a kind would be possible with the strong definite article and not with bare nouns.

### 5.3.2 One Marked

Table 8 shows the predictions for a language where only anaphoric definiteness is marked. In these languages, the determiner is required to express anaphoric definiteness. Bare nouns can get a kind level interpretation, but they cannot be used to refer to kinds or objects anaphorically.

An example is Akan, which has the overt anaphoric definite marker *nó*. In such a language, it would be predicted that all anaphoric nouns require the overt use of the anaphoric definite determiner, and this anaphoric determiner could not be used for unique definiteness or to refer to a kind in the absence of anaphora. Unfortunately, Amfo (2007) and Arkoh & Matthewson (2013) do not discuss kinds, but Arkoh & Matthewson (2013) claim that *nó* is not compatible with unique definiteness (28).

As discussed in section 4.3.2, a classifier can be used as a marker of definiteness. Simpson & Biswas (2016) present data on Bangla showing that the classifier is obligatory with a noun in anaphoric reference, bridging contexts, and reference to salient visible entities.<sup>11,12</sup> Dayal (2014) notes that bare nouns can refer to kinds and that the plural classifier combined with a noun cannot be used to refer to a kind in Bangla except anaphorically. This is predicted with this typology if we treat the classifier as ‘det.’ in Table 8.

**Table 8:** Definite bare noun interpretations, one marked, adapted from Despić 2019.

		+ Number						–Number			
		Kind-level			Object-level			Kind-level		Object-level	
		Mass	Count		Mass	Count		Mass	Count	Mass	Count
			SG	PL		SG	PL				
Unique	Type-Shift	∩	<i>l</i>	∩	<i>l</i>	<i>l</i>	<i>l</i>	∩	∩	<i>l</i>	<i>l</i>
	Overt	-	-	-	-	-	-	-	-	-	-
Anaph.	Type-Shift	*∩	* <i>l</i>	*∩	* <i>l</i>	* <i>l</i>	* <i>l</i>	*∩	*∩	* <i>l</i>	* <i>l</i>
	Overt	det.	det.	det.	det.	det.	det.	det.	det.	det.	det.

### 5.3.3 Unmarked

Finally, in Table 9 are the predictions for languages like Shan and Serbian that do not obligatorily mark either unique or anaphoric definiteness. In these languages, the determiner is not required to express either type of definiteness. Bare nouns can get a kind

<sup>11</sup>There is an interfering sociolinguistic factor that prohibits a classifier when referring to a respected referent. See Simpson & Biswas 2016 for more details.

<sup>12</sup>According to Simpson & Biswas (2016), the classifier-noun form is used when the referent is ‘identifiable’.

level interpretation, and in number marking languages, singular bare nouns can be used to refer to kinds anaphorically.

Here, it is expected that bare nouns can be used to express anaphoric definiteness. For number neutral languages, bare noun anaphora is the primary difference between languages with anaphoric definiteness obligatorily marked and languages with unmarked definiteness. For languages like Serbian that are number-marking, singular bare nouns can express anaphoric reference to a kind. In number-marking languages, this is an important difference between languages with unmarked versus marked anaphoric definiteness. An account where anaphoric definiteness is incompatible with type-shifting cannot account for cases where a kind can be referred to anaphorically.

**Table 9:** Definite bare noun interpretations, unmarked adapted from [Despić 2019](#).

		+ Number						-Number			
		Kind-level			Object-level			Kind-level		Object-level	
		Mass	Count		Mass	Count		Mass	Count	Mass	Count
			SG	PL		SG	PL				
Unique	Type-Shift	$\cap$	$\iota$	$\cap$	$\iota$	$\iota$	$\iota$	$\cap$	$\cap$	$\iota$	$\iota$
	Overt	-	-	-	-	-	-	-	-	-	-
Anaph.	Type-Shift	$*\cap$	$\iota$	$*\cap$	$\iota$	$\iota$	$\iota$	$*\cap$	$*\cap$	$\iota$	$\iota$
	Overt	-	-	-	-	-	-	-	-	-	-

These tables generate the full predictions of bare noun and determiner marked definite available cross-linguistically. Whether a definite marker functions as the definite determiner has an impact on which kinds of definiteness bare nouns can express.

## 5.4 A pragmatic proposal

[Ahn \(2019\)](#) proposed a competition based analysis of anaphoric definiteness marking along with an economy principle *Don't Overdetermine!* which takes into consideration the available ways to mark anaphoric definiteness in a language. *Don't Overdetermine!* is an economy principle that ‘chooses the semantically weakest element in the scale that can uniquely identify the intended referent.’ ([Ahn 2019](#): 74). [Ahn \(2019\)](#) argued that conflict between the principles *Don't Overdetermine!* and *Index!* lead to mixed judgments in anaphoric definiteness marking among speakers.

[Ahn \(2019\)](#) proposes that Korean and Thai have the following options for marking anaphoric definiteness, shown in (48) and (49), respectively.<sup>13</sup> Each language has a ranked scale of anaphoric expressions, where the more semantically complex expression

<sup>13</sup>[Ahn \(2019\)](#) departs from [Schwarz \(2009\)](#) in separating the anaphoric index from the lexical expression. I am not adopting this assumption, but this choice is not essential to the analysis.



will only be used in case the less complex expression cannot uniquely identify the referent. Since Korean does not lexicalize a simplex pronoun, bare nouns are the simplest way to express anaphora. In Thai, there is a simplex lexical pronoun which is preferred to the bare noun in contexts where there is only one available antecedent.

(48) Korean:  $\langle N, ku\ N \rangle$  (Ahn 2019: (83))

- a.  $\llbracket N \rrbracket = \iota x : \text{entity}(x) \wedge \phi(x) \wedge P(x)$
- b.  $\llbracket ku\ N \rrbracket = \iota x : \text{entity}(x) \wedge \phi(x) \wedge P(x) \wedge R(x)$

(49) Thai:  $\langle \text{pronoun}, N, N\ nán \rangle$ <sup>14</sup> (Ahn 2019: (84))

- a.  $\llbracket \text{pronoun} \rrbracket = \iota x : \text{entity}(x) \wedge \phi(x)$
- b.  $\llbracket N \rrbracket = \iota x : \text{entity}(x) \wedge \phi(x) \wedge P(x)$
- c.  $\llbracket N\ nán \rrbracket = \iota x : \text{entity}(x) \wedge \phi(x) \wedge P(x) \wedge R(x)$

According to Ahn (2019), the effect of non-uniqueness inferred for demonstrative-noun phrases comes as a result of this ranking of anaphoric expressions.<sup>15</sup> If the bare noun is not used, that indicates that the bare noun phrase cannot uniquely identify the intended referent, so non-uniqueness can be inferred.

This competition-based account predicts that if the type-shifted (anaphoric) bare noun denotation is a subset of the denotation of a demonstrative-noun phrase, we would expect to find bare nouns preferred to the demonstrative-noun phrase in some contexts. When there are enough available referents in the discourse context that the pronoun cannot uniquely identify the intended referent, we would expect the bare noun to be preferred to the demonstrative-noun phrase. I am assuming that bare nouns are both less morphologically and semantically complex than more marked definite expressions.<sup>16</sup>

This type of analysis seems to work for Thai. Here are two examples with donkey anaphora which mirror other examples used in this paper. (50) is the analogue of (32), and (51) has a very similar meaning to (28).<sup>17</sup> The donkey anaphors in (50) and (51) are bare nouns ‘bucket’ and ‘ball’ in (50) and ‘rat’ in (51).

(50) THAI: DONKEY ANAPHORA

lúuk-chaai wai nùŋ khàap khǒŋ chǎn baaj kràŋ kô tham nai sǐŋ thī mǎi  
 child-male aged one year POSS 1 some time PRT do in thing that NEG  
 khâatkhit. nai thúk thúk khraŋ thī khǎw hǎn thǎŋ-nám lɛ lúukbɔn  
 expect. in every every time that 3 see bucket-water and ball  
 khǎw mák ca waaj thǎŋ-nám bɔn lúukbɔn.  
 3 often IRR put bucket-water on ball

‘My one-year old son sometimes does unexpected things. For instance, every time he sees a bucket and a ball, he usually puts the bucket on the ball.’

<sup>14</sup>In Thai, the word order for demonstrative-noun phrases is  $N\ nán$ , so I have used that word order here.

<sup>15</sup>An alternative approach to demonstratives by Dayal & Jiang (to appear) includes a presupposition of non-uniqueness within a widened domain.

<sup>16</sup>The relative importance of semantic and morphological complexity in this competition-based account will not be discussed here.

<sup>17</sup>These examples were translated into Thai from English by one native Thai speaker, and then judged as grammatical by a second native speaker.

## (51) THAI: DONKEY ANAPHORA

mew lɛ nǔ mǎi khǒi ca thùuk kan. mǔa-rài thī mew hěn nǔ, man  
 cat and rat NEG quite IRR correct together when that cat see rat, 3  
 mǎk ca lǎi-càp nǔ  
 often IRR chase-catch rat

‘Cats and rats don’t get along. Whenever a cat sees a rat, it usually chases and catches the rat.’

This account has more difficulty explaining the pattern of anaphora in Mandarin. This can be seen in the following two examples, repeated from (8a)-(8b), above. Here, we might expect to be unable to use a pronoun because *nansheng* ‘boy’ and *nǚsheng* ‘girl’ would be referred back to with the same pronoun.<sup>18</sup> In this case, we might expect the bare noun to refer back to the boy. However, it seems like the demonstrative-noun phrase must be used.

(8a) jiaoshi li zuo-zhe yī ge nansheng he yī ge nǚsheng,  
 classroom inside sit-PROG one CLF boy and one CLF girl

‘There is a boy and a girl sitting in the classroom...’

(8b) Wo zuotian yudao #(na ge) nansheng  
 I yesterday meet that CLF boy

‘I met the boy yesterday.’

(Mandarin, Jenks 2018: (15a,b))

If the demonstrative in a language has been grammaticalized to function as an anaphoric definite determiner, we might expect the Blocking Principle to disallow anaphoric definite type-shifting in the language. Therefore, a bare noun cannot be used in this sort of anaphoric context. Another possible explanation is that word order has an effect on the available bare noun interpretations with the result that the bare noun is not a potential anaphoric competitor in this syntactic context. It is well-known that there is an effect of sentence structure, where nominal expressions in the post-verbal position are typically indefinite (Li & Bisang 2012). Additionally, more recent work has suggested that the subject-object asymmetry in anaphoric bare nouns in Mandarin is not so robust. Ahn (2019) notes some variability in judgments from five Mandarin speakers looking at (8a)-(8b). Dayal & Jiang (to appear) give detailed evidence supporting the idea that non-subject bare nouns can be anaphoric in Mandarin.<sup>19</sup>

If an economy constraint approach can account for why bare nouns cannot appear in all contexts in Thai and Mandarin (perhaps with the effects of word order factored in), these languages seem to belong in the ‘unmarked’ typological definiteness category so as not to rule out the possibility of bare noun anaphora.

This section has shown that a competition based approach to definite expressions can explain some of the variation found both within and across languages. This analysis uses *Don’t Overdetermine!* from Ahn (2019) to choose which definite expression is appropriate in which context. In contexts where a pronoun fails to uniquely identify the referent,

<sup>18</sup>This is true in spoken language rather than written. See Ahn 2019 for more discussion.

<sup>19</sup>Bremmers et al. (2019) use a technique called Translation Mining to support Jenks’s (2018) claim that Mandarin bare nouns can denote unique definiteness. They suggest that bare nouns can also denote anaphoric definiteness, but do not go into detail about when a bare noun can be anaphoric.

a bare noun can express anaphoric definiteness, as the examples of donkey anaphora in Thai showed. This supports the idea that Thai is a language that does not obligatorily mark anaphoric definiteness with an overt morpheme, but instead has a bare noun as an option for anaphoric definiteness. In a similar way, this account can explain why certain anaphoric expressions might be preferred in specific contexts in Shan.

While I have argued that Thai and Mandarin belong in the ‘unmarked’ typological category, there are still languages that do appear to fall within the category of languages that obligatorily mark anaphoric definiteness, such as Wu, Akan, and Bangla. While I adopt a competition account of definiteness to explain the bare noun/demonstrative-noun phrase variation in Shan, I maintain that there is a distinction between unique and anaphoric definiteness that is represented overtly in some languages.

## 5.5 Summary

This section has proposed that the mechanism for deriving a definite interpretation for a bare noun is through type shifting as described by Chierchia (1998) and Dayal (2004). The choice between anaphoric definite expressions can, at least in part, be determined by an economy principle such as Ahn’s (2019) *Don’t Overdeterminate!*, which says that the minimal definite expression that uniquely identifies an intended referent is the one that should be used.

New donkey anaphora data from Thai combined with a competition based analysis of definite expressions suggests that Thai might better be categorized as a language that does not require overtly marked anaphoric definites. Since bare nouns can express anaphoric definiteness in some contexts in Mandarin, perhaps it should also be characterized as belonging in the ‘unmarked’ category. Akan as described by Arkoh & Matthewson (2013), the Jinyun dialect of Wu described by Simpson (2017), and Bangla as described by Simpson & Biswas (2016) could still be examples of languages that only overtly marks anaphoric definites. The final section discusses other factors that affect expressions of definiteness and concludes.

## 6 Discussion

There is a great deal of variation in how obligatorily languages mark definiteness both cross-linguistically and within one language. We are left with the question of what causes this variation beyond the set of available anaphoric definite expressions within a language. This section discusses some of these issues.

One important factor in definiteness marking is syntactic position. As Jenks (2018) noted, subject position in Mandarin allows for anaphoric bare nouns in that position. Others, such as Li & Bisang (2012), have noted that in some languages the post-verbal position is associated with indefiniteness for both bare nouns and classifier-noun phrases. In these languages, the syntactic position appears to license or prohibit certain semantic possibilities. This would affect which expressions would be competing in particular syntactic positions.

There are several other factors connected to pragmatics and information structure in determining the choice of definite expression. [Simpson et al. \(2011\)](#) identify several factors that could influence whether the bare noun or the overtly marked definite nominal expression is preferred:

- (52) i. The role of contrast
- ii. The role of relative sentential prominence
- iii. The role of disambiguation

This section offers a brief discussion of contrast (§6.1) and sentential prominence (§6.2) and some discussion of the role of ambiguity in definiteness marking (§6.3).

## 6.1 Contrast

Discussions of the role of contrast can be difficult since ‘contrast’ can refer to different things. As has been discussed in §3.2, for [Ahn \(2019\)](#) contrast is connected to the number of salient entities. [Ahn \(2019\)](#) claims that for ASL, for example, having multiple animate salient entities in a given contexts allows the bare noun to be used. For this reason, bare nouns are more common for languages like Shan, Serbian, and Thai—where a pronoun might be competing with a bare noun to express anaphora—in contexts where there is more than one individual in the discourse context.

This type of contrast is connected to, but importantly different from contrastive topics, discussed in relation to definiteness marking by [Jenks \(2018\)](#) and [Simpson \(2017\)](#). Contrastive topic is connected to information structure and generating alternative propositions. Even for contrastive topics, definiteness marking can behave differently than expected. According to [Jenks \(2018\)](#), contrastive topic ameliorates anaphoric definite use of the bare noun. However, [Simpson et al. \(2011\)](#) claimed that in a language that can use a classifier-noun phrase to mark definiteness, contrast leads to using a classifier-noun phrase instead of a bare noun.

There is more work to be done looking at the effects of contrast and contrastive topics. Some factors to consider might include overlap in: (i) pronoun morphology, (ii) semantic features such as animacy, (iii) grammatical  $\phi$ -features (number, gender), (iii) relative prominence in discourse, and (iv) probability of relevance within a particular semantic event (e.g., birds are more likely to be the subject of the verb ‘fly’ than frogs).

## 6.2 Sentential prominence

[Simpson et al. \(2011\)](#) noted that the classifier-noun phrase can indicate relative prominence of a referent within a sentence. Discourse old referents might be more likely to be bare nouns rather than classifier-noun phrases in languages like Vietnamese. [Jenks \(2015\)](#) noted a similar pattern for Thai, where in a long narrative, bare nouns can be used anaphorically, saying this happens ‘after an individual has been established and it is clear that they are the only individual of the relevant type’ (2015: 113, fn. 7).

Bare noun use in both of these cases seems compatible with what [Ahn \(2019\)](#) said about anaphoric bare nouns: they indicate that the property of being that noun is sufficient to uniquely identify the referent. However, sentential prominence is also connected

to information structural properties such as Givenness and Focus as discussed by [Krifka & Musan \(2012\)](#), among others. It would be worthwhile to investigate the connection between information structure and bare noun definiteness.

### 6.3 Ambiguity

The role of disambiguation in determining whether to use a bare noun or demonstrative expression might be connected to a pragmatic principle along the lines of *Don't Overdetermine!* from [Ahn 2019](#). In a context where there is more ambiguity due to a possible generic or plural interpretation, it might be expected that definiteness is more often marked (i.e., a demonstrative expression rather than a bare noun) in order to clearly identify the intended referent.

Perhaps what *Index!* from [Jenks \(2018\)](#) represents is the preference for a language to unambiguously specify an intended meaning. Since bare nouns can have a variety of interpretations, using a demonstrative-noun phrase can demonstrate clearly that a definite interpretation is intended. We would expect this to come into play whenever an expression is potentially ambiguous. This ambiguity might more often arise with bare nouns or with *pro*, both of which can have several interpretations.<sup>20</sup> Ambiguity could even arise with expressions like pronouns or other definite expressions.

Some predicates are only true of individuals or true of kinds, but some predicates can have both meanings. A generic or kind interpretation for a nominal expression can compete with the definite one. I propose that the bare noun can have competing meanings that could lead to using expressions with more complex denotations in ambiguous contexts. Therefore, for a given expression, both an economy principle like *Don't Overdetermine!* and a principle that avoids potential ambiguity must be taken into account. This type of analysis can begin to tease apart what factors come into play when we are deciding how informative we need to be.

As with lexicalization of definite expressions and syntactic effects on definiteness, the array of possible ambiguity will be language-specific. For example, a language like Serbian that overtly marks number will not run into ambiguity between singular and plural, whereas Shan might. The prediction is that expressions with more articulated semantics will be used in contexts where ambiguity could generate the wrong interpretation.

In a discussion of types of ambiguous expressions, [Sennet \(2016\)](#) included both flexible types (i.e., from the possibility of type shifting) and the generic versus episodic readings of sentences, which is connected to the contrast between kinds and individuals. Thus, bare noun expressions have already been identified as being potentially ambiguous. An example of this is shown in (53). Under [Chierchia's \(1998\)](#) analysis of the bare plural *dinosaurs*, the base denotation of the plural is a kind and the possible episodic and generic interpretations come from existential closure (as a result of [Chierchia's \(1998\)](#) Derived Kind Predication) and from a generic operator quantifying over instances of the kind, respectively.

(53) Dinosaurs ate kelp.

([Sennet 2016: \(31\)](#))

<sup>20</sup>[Ahn \(2019\)](#) analyzes *pro* in languages like Korean, Mandarin, and Thai as being different from the Romance style *pro* in whether they compete with other anaphoric expressions.

Returning to Shan, the competition between the generic and definite readings are apparent from the contrast between (54a) and (54b) as possible follow up sentences to (54).<sup>21</sup> (54) comes as part of a longer story where two friends are discussing possible pets that Jai Kham might have bought at the pet store. The addressee of this sentence knows that Jai Kham was interested in buying a particular black cat and white dog. Both (54a) and (54b) are judged as good follow ups to sentence (54). In (54a), the bare noun *méw* ‘cat’ is interpreted as the aforementioned black cat, but in (54b), the bare noun *măa* ‘dog’ is most easily interpreted as referring to dogs in general.<sup>22</sup> Both the definite and generic readings are possible for both sentences, but one reading is more salient than the other. Given that there is already one salient dog and cat in the context, the bare noun should be the most appropriate definite expression given the economy principle. However, some speakers do prefer a demonstrative-noun phrase in these sentences. I argue that this stems from the possible ambiguity between a definite and generic interpretation.

(54) SHAN: AMBIGUITY AND ANAPHORA

Tsáaj Khám lɿk sù **méw lăm tǔ** **nân**, kójkaa ʔəm laj lɿk  
 Jai Kham choose buy cat black CLF.ANML that but NEG get choose  
**măa khăaw tǔ nân.**  
 dog white CLF.ANML that

‘Jai Kham chose and bought that black cat, but did not choose that white dog.’

- a. jôn waa **méw** láklěm nàa.      b. jôn waa **măa** ʔəm láklěm.  
 because COMP cat clever very      because COMP dog NEG clever  
 ‘because the cat is very smart.’      ‘because dogs aren’t smart.’

Table 10 shows potential competing expressions (on the horizontal axis) and interpretations (on the vertical axis) for a subset of nominal expressions in Shan. For example, the [N dem.] expression indicates anaphoric definiteness, but it can be anaphoric to a kind, a singular entity, or a plural entity.

**Table 10:** Anaphoric and kind interpretations, potential ambiguity in Shan.

unique definite:	⟨... [N] ...⟩
kind:	⟨... [N] ...⟩
singular:	⟨... [N] ...⟩
plural:	⟨... [N] ...⟩
anaphoric definite:	⟨...[pronoun] [N] [N dem.] [N CLF/PL dem.] ...⟩
kind:	⟨...[pronoun] [N dem.] ...⟩
singular:	⟨...[pronoun] [N] [N dem.] [N CLF dem.] ...⟩
plural:	⟨...[pronoun] [N] [N dem.] [N PL dem.] ...⟩

<sup>21</sup>The complementizer *waa* here is distinct from the one that appears with relative clauses. See Jenks (2011) for a discussion of the cognate forms found in Thai.

<sup>22</sup>Bare nouns can be definite in sentences with negation in Shan.



We would expect that in contexts where a definite expression has more than one possible reading, we might instead use a more articulated expression that will not lead to the same ambiguity.

## 7 Conclusion

Shan can use bare nouns to express both unique and anaphoric definiteness. This pattern of bare noun anaphora has not previously been reported in languages like Thai and Mandarin. Other languages, such as Kannada and Serbian, also use bare nouns to express anaphoric definiteness. I have argued that languages that can express anaphoric definiteness using bare nouns should be placed in a new category of definiteness marking: the unmarked definiteness category.

This paper has shown that the Shan data matches well with the predictions of the type shifting analysis from Chierchia (1998) and Dayal (2004). Definite, generic, kind, and narrow scope indefinite readings are all available for bare nouns in Shan. Ahn's (2019) economy principle, *Don't Overdetermine!*, which provides a competition-based approach to anaphoric definite expressions can explain some of the variability in whether a bare noun or another definite expression is preferred in a particular anaphoric context. This approach can also explain why contexts with more than one possible antecedent expression, are suitable for bare nouns. If there is only one possible antecedent, a pronoun would be preferred based on the economy principle. This is consistent with previous work by Givón and Bisang on the preference of pronouns to other definite expressions. For example, Givón (2017) says that null pronouns or unstressed pronouns are the easiest things to use to continue discussion of a topical referent, and using a full NP signals discontinuous reference. Similarly, Bisang (1999) notes that the classifier-noun phrase in Vietnamese is used to refer to previously mentioned entities when a pronoun cannot be used.

In the way that the Blocking Principle of the type shifting analysis and the economy principle *Don't Overdetermine!* are constructed, Mandarin and Thai do not count as marked anaphoric languages in this revised typology. Instead they are unmarked, and all the factors that influence the choice of definiteness expression described here play a role in determining whether a bare noun or other expression is used. It remains an open question whether Wu and Akan obligatorily marks anaphoric definite expressions with an anaphoric definite determiner, but the data from Arkoh & Matthewson (2013) and Simpson (2017) suggest that they do. Banlga as described by Simpson & Biswas (2016) also seems to belong in this category. The proposed typology makes predictions for the available nominal expressions that make definite reference to both individuals and kinds within a particular typological category.

This paper also discussed other factors that influence the choice of definiteness expression and proposed that ambiguity is a significant factor. If a nominal expression can have different interpretations in different contexts, use of that nominal expression could lead to ambiguity and failure to uniquely identify an intended referent. I proposed, therefore, that in potentially ambiguous contexts, a semantically more complex definite expression

might be chosen. Taking the role of ambiguity into account could explain some of the patterns of data that the economy principle cannot account for.

## Abbreviations

1: first person, 3: third person, ACC: accusative, ANML: animal, CLF: classifier, COMP: complementizer, COP: copula, DAT: dative, DEF: definite, GEN: genitive, HUM: human, INDEF: indefinite, IRR: irrealis, IPFV: imperfective, LOC: locative, NEG: negation, NOM: nominative, PL: plural, POSS: possessive, PROG: progressive, PRT: particle, PST: past, RND: round, SG: singular

## Competing interests

The author has no competing interests to declare.

## References

- Ahn, Dorothy. 2019. *THAT thesis: A competition mechanism for anaphoric expressions*: Harvard University dissertation.
- Amfo, Nana Aba Appiah. 2007. Akan demonstratives. In *Selected proceedings of the 37th annual conference on African linguistics*. 134–148. Somerville: Cascadia Press.
- Arkoh, Ruby & Lisa Matthewson. 2013. A familiar definite article in Akan. *Lingua* 123. 1–30. <https://doi.org/10.1016/j.lingua.2012.09.012>.
- Bisang, Walter. 1999. Classifiers in East and Southeast Asian languages: Counting and beyond. *Nominal types and changes worldwide* 113–185.
- Bochnak, M. Ryan & Lisa Matthewson. 2015. *Methodologies in Semantic Fieldwork*. Oxford University Press.
- Bremmers, David, Jianan Liu, Martijn van der Klis & Bert Le. 2019. Definiteness across Languages: From German to Mandarin. In *The 22nd Amsterdam Colloquium*.
- Büring, Daniel. 2003. On D-trees, beans, and B-accent. *Linguistics and philosophy* 26(5). 511–545.
- Carlson, Greg N. 1977. A unified analysis of the English bare plural. *Linguistics and Philosophy* 1(3). 413–457. <https://doi.org/10.1007/BF00353456>.
- Cheng, Lisa L-S & Rint Sybesma. 2005. Classifiers in four varieties of Chinese. *Handbook of comparative syntax* 259–292.
- Cheng, Lisa Lai-Shen & Rint Sybesma. 1999. Bare and Not-So-Bare Nouns and the Structure of NP. *Linguistic Inquiry* 30(4). 509–542. <https://doi.org/10.1162/002438999554192>.
- Chierchia, Gennaro. 1998. Reference to Kinds across Language. *Natural Language Semantics* 6(4). 339–405. <https://doi.org/10.1023/A:1008324218506>.
- Chung, Sandra & William A. Ladusaw. 2003. *Restriction and Saturation*. MIT Press.

- Dayal, Veneeta. 2004. Number Marking and (in)Definiteness in Kind Terms. *Linguistics and Philosophy* 27(4). 393–450. <https://doi.org/10.1023/B:LING.0000024420.80324.67>.
- Dayal, Veneeta. 2014. Bangla plural classifiers. *Language and Linguistics* 15(1). 47–87.
- Dayal, Veneeta & Julie Jiang. to appear. The Puzzle of Anaphoric Bare Nouns in Mandarin: A Counterpoint to Index! *Linguistic Inquiry*.
- Deal, Amy Rose & Julia Nee. 2018. Bare nouns, number, and definiteness in Teotitlán del Valle Zapotec. In *Proceedings of Sinn und Bedeutung*, vol. 21. 317–334.
- Despić, Miloje. 2019. On kinds and anaphoricity in languages without definite articles. *Definiteness across languages* 25. 259.
- Diesing, Molly. 1992. Bare plural subjects and the derivation of logical representations. *Linguistic Inquiry* 23(3). 353–380.
- Eberhard, David M., Gary F. Simons & Charles D. Fennig. 2019. Shan. *Ethnologue: Languages of the World*.
- Frege, Gottlob. 1892. über Sinn und Bedeutung [On sense and reference]. *Zeitschrift für Philosophic und philosophische Kritik* 100. 25–50.
- Givón, Talmy. 2017. *The Story of Zero*. Amsterdam: John Benjamins Publishing Company.
- Hawkins, John A. 1978. *Definiteness and indefiniteness: A study in reference and grammaticality prediction*. Routledge.
- Heim, Irene. 1982. The semantics of definite and indefinite noun phrases.
- Heim, Irene. 1991. Artikel und definitheit. *Semantik: ein internationales Handbuch der zeitgenössischen Forschung* 487–535.
- Irani, Ava. 2016. Two types of definites in American Sign Language.
- Irani, Ava. 2019. On (in) definite expressions in American Sign Language. *Studies in Diversity Linguistics* (25).
- Jenks, Peter. 2011. *The hidden structure of Thai noun phrases*. Cambridge, MA: Harvard University dissertation.
- Jenks, Peter. 2015. Two kinds of definites in numeral classifier languages. *Semantics and Linguistic Theory* 25(0). 103–124. <https://doi.org/10.3765/salt.v25i0.3057>.
- Jenks, Peter. 2018. Articulated definiteness without articles. *Linguistic Inquiry* 49(3). 501–536.
- Jiang, Li. 2012. Nominal arguments and language variation. *Unpublished doctoral dissertation, Department of Linguistics, Harvard University, Cambridge, Massachusetts*.
- Jiang, Li Julie. 2018. Definiteness in Nuosu Yi and the theory of argument formation. *Linguistics and Philosophy* 41(1). 1–39. <https://doi.org/10.1007/s10988-017-9219-6>.
- Kadmon, Nirit. 1990. Uniqueness. *Linguistics and Philosophy* 13(3). 273–324.
- Krifka, Manfred & Renate Musan. 2012. Information structure: Overview and linguistic issues. *The expression of information structure* 5. 1–44.
- Li, XuPing & Walter Bisang. 2012. Classifiers in Sinitic languages: From individuation to definiteness-marking. *Lingua* 122(4). 335–355. <https://doi.org/10.1016/j.lingua.2011.12.002>.
- Little, Carol-Rose & Ekarina Winarto. 2019. Classifiers and the definite article in Indonesian. In Maggie Baird & Jonathan Pesetsky (eds.), *Proceedings of the 49th North East Linguistics Society*, vol. 2. 209–220. Amherst, MA: GLSA.

- Löbner, Sebastian. 1985. Definites. *Journal of Semantics* 4(4). 279–326. <https://doi.org/10.1093/jos/4.4.279>.
- Moroney, Mary. 2019a. Definiteness with Bare Nouns in Shan. In Jennifer Sikos & Eric Pacuit (eds.), *At the Intersection of Language, Logic, and Information* (Lecture Notes in Computer Science). 108–123. Berlin, Heidelberg: Springer. [https://doi.org/10.1007/978-3-662-59620-3\\_7](https://doi.org/10.1007/978-3-662-59620-3_7).
- Moroney, Mary. 2019b. Inconsistencies of the Consistency test. *Proceedings of the Linguistic Society of America* 4(1). 55–1–10. <https://doi.org/10.3765/plsa.v4i1.4560>.
- Piriyawiboon, Nattaya. 2010. *Classifiers and determiner-less languages: The case of Thai*. University of Toronto Doctoral dissertation.
- Roberts, Craige. 2003. Uniqueness in Definite Noun Phrases. *Linguistics and Philosophy* 26(3). 287–350. <https://doi.org/10.1023/A:1024157132393>.
- Royer, Justin. 2019. Domain restriction and noun classifiers in Chuj (Mayan). In *Proceedings of the Forty-Ninth Annual Meeting of the North East Linguistic Society*, vol. 3. 87–97.
- Russell, Bertrand. 1905. On Denoting. *Mind* 14(56). 479–493.
- Schwarz, Florian. 2009. *Two types of definites in natural language*. University of Massachusetts, Amherst dissertation.
- Schwarz, Florian. 2013. Different types of definites crosslinguistically. *Language and Linguistics Compass* 7(10). 534–559.
- Schwarz, Florian. 2019. Weak vs. strong definite articles: Meaning and form across languages. *Studies in Diversity Linguistics* (25).
- Sennet, Adam. 2016. Ambiguity. In Edward N. Zalta (ed.), *The stanford encyclopedia of philosophy*, Metaphysics Research Lab, Stanford University spring 2016 edn.
- Simpson, Andrew. 2017. Bare classifier/noun alternations in the Jinyun (Wu) variety of Chinese and the encoding of definiteness. *Linguistics* 55(2). 305–331.
- Simpson, Andrew & Priyanka Biswas. 2016. Bare nominals, classifiers, and the representation of definiteness in Bangla. *Linguistic Analysis* 40(3-4). 167–198.
- Simpson, Andrew, Hooi Ling Soh & Hiroki Nomoto. 2011. Bare classifiers and definiteness: A cross-linguistic investigation. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"* 35(1). 168–193. <https://doi.org/10.1075/sl.35.1.10sim>.
- Sio, Joanna. 2006. *Modification and reference in the Chinese nominal*. The Leiden University Centre for Linguistics (LUCL), Faculty of Arts dissertation.
- Srinivas, Sadhwi & Kyle Rawlins. 2020. Definiteness and the bare nominal in Kannada. *Talk presented at The 94th Annual Meeting of the Linguistic Society of America (LSA)*.
- Strawson, P. F. 1950. On Referring. *Mind* 59(235). 320–344.
- Wespel, Johannes. 2008. *Descriptions and their domains: The patterns of definiteness marking in French-related creoles*. Stuttgart: University of Stuttgart dissertation.
- Yang, Rong. 2001. *Common nouns, classifiers, and quantification in Chinese*. Rutgers, The State University of New Jersey dissertation.