

# 'Extra' *cái* Definites in Vietnamese

Aidan Sharma    aidan.sharma@rutgers.edu  
Ariela Ye        arielalu.ye@rutgers.edu

TEAL-14 @ USC, March 18-20, 2025

## 1 Introduction

Definiteness has been a long studied phenomenon in semantics, with long-standing efforts aiming to tease apart the presuppositions/assertions of definiteness encoding morphemes such as English *the* (Frege 1892; Russell 1905; Heim 1983; Roberts 2003 a.o.). In recent years, there have been a number of works on the differences in the syntactic and semantic representations of *familiar/strong* and *unique/weak* definites in natural languages (Schwarz 2009, 2013; Arkoh & Matthewson 2013; Jenks 2015, 2018; Šereikaitė 2019; Owusu 2022 a.o.).

While many interesting discoveries have been made by investigating the instantiations of different types of definites in various languages, there are also interesting patterns with regard to how definiteness is expressed in different types of languages. The focus of this paper is the syntax of definites in classifier languages, with emphasis on a particular pattern in Vietnamese – the so called 'extra' *cái* definites.

Generally, classifiers are elements that occur with numerals and nouns and appear to be a prerequisite for counting in classifier languages (1). Whether classifiers are needed for the sake of numerals (Bale & Coon 2014) or nouns (Chierchia 1998) is a topic of debate, and it is possible that the answer varies depending on the language (Simpson & Ngo 2018).

- (1) Vietnamese  
hai \*(con) chó  
two CL     dog  
"two dogs"

A great deal of work has been done on definites in classifier languages (Cheng & Sybesma 1999, 2005, 2012; Simpson 2005; Simpson et al. 2011; Simpson & Biswas 2016; Simpson 2017; Simpson & Ngo 2018; Jenks 2015; Syed 2015, 2016; Simpson & Syed 2016; Syed & Simpson 2017; Hsu & Syed 2020 a.o.), and one interesting structure that can license a definite interpretation in many of these

languages is the so-called "bare classifier" construction. In such constructions, classifiers occur with an NP but no numeral.

- (2) Vietnamese; Austroasiatic (Simpson et al. 2011: 170)

[*cái máy ảnh*]    rất    đắt tiền  
[CL camera]    very expensive  
'The camera was very expensive.'

- (3) Cantonese; Sinitic (Cheng 2012: 206)

[*zek<sup>3</sup> gau<sup>2</sup>*]    gam<sup>1</sup>jat<sup>6</sup>    dak<sup>6</sup>bit<sup>6</sup>    teng<sup>1</sup>waa<sup>6</sup>  
[CL dog]    today    special    obedient  
'The dog is specially obedient today.'

- (4) Wu; Sinitic (Cheng & Sybesma 2005: 5)

[*dyu<sup>8</sup> kau<sup>8</sup>*]    i<sup>5</sup>    tsau<sup>3</sup>-ku<sup>5</sup>    ka<sup>1</sup>-løy<sup>6</sup>  
[CL dog]    want walk-cross street  
'The dog wants to cross the street.'

- (5) Hmong; Hmong-Mien (Simpson 2005: 823) from (Jaisser 1987: 171)

[*tus tsov*]    tshaib    tshaib    plab  
[CL tiger]    hungry hungry stomach  
'The tiger was very hungry.'

- (6) Nùng; Tai (Simpson 2005: 823) from (Saul & Wilson 1980: 26)

lêo [tú mệ] đạ [tú pộ] vạ ...  
 then [CL wife] scold [CL husband] say ...  
 'Then the wife scolded the husband and said ...'

- (7) Bangla; Indo-Aryan (Simpson & Biswas 2016: 5)

[mEthor Ta] porisromi kintu [gard Ta] besh kuMRe  
 [janitor CL] hard-working but [guard CL] quite lazy  
 'The janitor is hard-working, but the guard is quite lazy.'

The goal of this paper is to see whether all these bare classifier phrases are the same cross-linguistically, and how Vietnamese can inform us about the differences or similarities in their structures.

## 2 Vietnamese

### 2.1 Classifiers

Vietnamese displays a rich classifier system typical of East and Southeast Asian languages. Nouns and numerals generally cannot directly combine without an intervening classifier (Simpson & Ngo 2018: 213-214; Thuong Bui p.c.).

- (8) **Obligatory-classifier nouns**

- a. hai \*(con) chó  
two CL dog  
'two dogs'
- b. bốn \*(cuốn) sách  
four CL book  
'four books'
- c. hai \*(chiếc) xe đạp  
two CL vehicle cycle  
'two bicycles'
- d. ba \*(quyển) vở  
three CL notebook  
'three notebooks'

However, for some Vietnamese nouns classifiers are optional (9) or not allowed at all (10) (Simpson & Ngo 2018: 214, 225).

- (9) **Optional-classifier nouns**

- a. hai (người) nhân viên  
two CL employee  
'two employees'
- b. tám (cái) làng  
eight CL village  
'eight villages'

- (10) **Non-classified nouns**

- a. hai màu  
two color  
'two colors'
- b. hai nước  
two country  
'two countries'

### 2.2 Indefinites

A range of syntactic structures can be interpreted indefinitely in Vietnamese, including bare nouns, bare classifier phrases, and numeral phrases.

- (11) a. Tôi muốn mua sách.  
1SG want buy book  
'I want to buy a book/books.' N
- b. Tôi muốn mua cuốn sách.  
1SG want buy CL book  
'I want to buy a book.' CL-N
- c. Hai cuốn sách rách rồi.  
two CL book torn already  
'Two books were torn.' Num-CL-N

### 2.3 Definites

Interestingly, these same structures can all have definite interpretations as well (Phan & Dong 2021: 7-10).

- (12) a. Sách rách rồi.  
book torn already  
'The book(s) was/were torn.' N
- b. Tôi làm rách sách rồi.  
1SG make torn book already  
'I made the book(s) torn.' N
- (13) a. con chó  
CL dog  
'the dog' CL-N
- b. cái bàn  
CL table  
'the table' CL-N
- (14) a. Hai cuốn sách là những người bạn tri-kỷ của tôi.  
two CL book COP PL CL friend soul of 1SG  
'The two books are my soul-mates.' Num-CL-N
- b. Tôi làm rách hai cuốn sách rồi.  
1SG make torn two CL book already  
'I made the two books torn.' Num-CL-N

This indicates that none of these structures are inherently definite or indefinite in Vietnamese.

## 2.4 'Extra' cái

Vietnamese also allows the use of an "extra" classifier. When the general classifier *cái* precedes the usual CL-NP sequence, it enforces a definite interpretation.

- (15) a. hai cái con chó  
two cai CL dog  
'the two dogs' (Simpson & Ngo 2018: 224)
- b. cái chiếc bàn này  
cai CL table DEM  
'this table' (Simpson 2005: 825 from Nguyen 1975)

Even though we saw previously that N, CL-N, and Num-CL-N can all be definite or indefinite in Vietnamese, 'extra' *cái* constructions requires a definite interpretation. Because of this, we follow Simpson & Ngo in assuming that

the primary contribution of 'extra' *cái* is definiteness. The most common view, however, is that 'extra' *cái* is a focus marker (Nguyen 1975, 2002, 2004; Đoàn et al. 2019; Phan & Dong 2021). We will return to this in Section 4.

This extra classifier phenomenon also appears to exist in Nùng.

- (16) câu tú pá dụ cháng áhn hơn nĩ, mưhn vạ ...  
CL CL fish at inside CL house FOC he say  
'The fish inside the house said ...' (Saul & Wilson 1980: 16)

## 3 Bare/Extra Classifier Definites

We return now to the question of bare classifier definites. Analyses vary regarding the amount of structure that is projected. In their examination of Mandarin and Cantonese (and later Wu and Min), Cheng & Sybesma (1999, 2005) argue that these definite interpretations of bare classifier phrases do not project more structure and are true "bare classifier phrases". Cheng & Sybesma state that CL in these constructions can perform the deictic function of picking out the individual that is described by the NP. The structure Cheng & Sybesma propose for (3) then looks as follows.

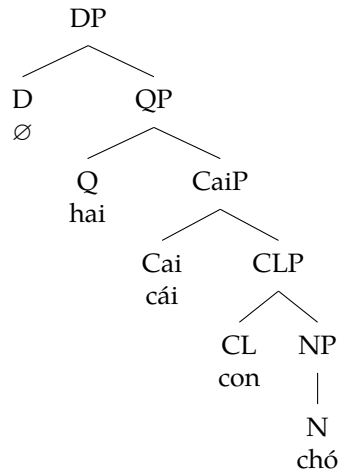
- (17) Cantonese
- ```

      CLP
     /  \
    CL   NP
   zek3 gau2
  
```

Meanwhile, on the basis of the 'extra' *cái* constructions, Simpson (2005); Simpson & Ngo (2018) propose that there is always structure projected above CLP. Simpson (2005) takes this to be DP but since numerals can occur above 'extra' *cái*, Simpson & Ngo: 224 revise this to be a dedicated projection for *cái* above CLP. In such constructions, the CL still performs the individuating function while *cái* performs the deictic function.

- (18) hai cái con chó  
two cai CL dog  
'the two dogs'

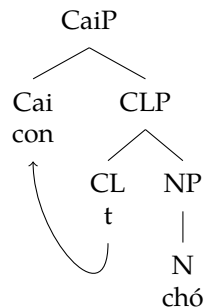
(19) Vietnamese



Presumably, the analysis for bare classifier definites assumed by Simpson & Ngo (2018) is CL to Cai movement (rather than CL to D movement as in Simpson 2005).

(20) [cái máy ảnh] rất đắt tiền  
 [CL camera] very expensive  
 'The camera was very expensive.'

(21) Vietnamese



The question of whether one of these structures should be adopted for both languages, or what information indicates which type of structure is present can best be answered by examining similar expressions in other languages.

### 3.1 CL Covariation

Nguyen (2004) notes that unlike the ordinary classifier *cái*, 'extra' *cái* appears with all nouns. However, this 'extra' *cái* occurs in addition to the usual classifier that occurs with any given noun. In bare classifier phrases, the classifier is consistent between Vietnamese indefinites and definites. This is also the case for Cantonese, Wu, Hmong, and Nùng.

(22) Vietnamese

- John đã gặp một người đàn ông và một người đàn bà trên đường  
 John PST meet one CL man and one CL woman on way  
 'John met a man and a woman on the way.'
- Người đàn ông đang la mắng, người đàn bà đang khóc.  
 CL man PROG scold CL woman PROG cry  
 'The man was scolding, the woman was crying.'
- tôi muốn mua cuốn sách  
 1SG want buy CLF book  
 'I want to buy a book.' (Phan & Dong 2021: 8)
- cuốn sách rách rồi  
 CL book torn already  
 'The book was torn.' (Phan & Dong 2021: 8)

(23) Cantonese

- saam<sup>1</sup> zek<sup>3</sup> gau<sup>2</sup>  
 three CL dog  
 'three dogs' (Cheng & Sybesma 2012: 209)
- zek<sup>3</sup> gau<sup>2</sup> gam<sup>1</sup>jat<sup>6</sup> dak<sup>6</sup>bit<sup>6</sup> teng<sup>1</sup>waa<sup>6</sup>  
 CL dog today special obedient  
 'The dog is specially obedient today.' (Cheng 2012: 206)
- ngo<sup>5</sup> soeng<sup>2</sup> maai<sup>5</sup> bun<sup>2</sup> syu<sup>1</sup> lei<sup>4</sup> tai<sup>2</sup>  
 I want buy CL<sup>volume</sup> book come read  
 'I want to buy a book to read.' (Cheng & Sybesma 2005: 8)
- bun<sup>2</sup> syu<sup>1</sup>  
 CL<sup>volume</sup> book  
 'the book' (Cheng & Sybesma 2005: 12)

(24) Wu<sup>1</sup>

- a. ɲ<sup>4</sup> ɕi<sup>3</sup> ma<sup>4</sup> dyu<sup>2</sup> si<sup>1</sup>  
I want buy CL dog  
'I would like to buy a dog.' (Cheng & Sybesma 2005: 5)
- b. dyu<sup>8</sup> kau<sup>8</sup> i<sup>5</sup> tsau<sup>3</sup>-ku<sup>5</sup> ka<sup>1</sup>-løy<sup>6</sup>  
CL dog want walk-cross street  
'The dog wants to cross the street.' (Cheng & Sybesma 2005: 5)
- c. ɲ<sup>4</sup> ɕi<sup>3</sup> ma<sup>4</sup> pan<sup>3</sup> si<sup>1</sup>  
I want buy CL<sup>volume</sup> book  
'I want to buy a book' (Cheng & Sybesma 2005: 5)
- d. ɲ<sup>4</sup> ɕi<sup>3</sup> ma<sup>4</sup> pan<sup>7</sup> si<sup>1</sup>  
I want buy CL<sup>volume</sup> book  
'I want to buy the book' (Cheng & Sybesma 2005: 5)

## (25) Hmong

- a. muaj ib tus tsov  
be one CL tiger  
'There is a tiger.' (Jaisser 1987: 171)
- b. tus tsov tshaib tshaib plab  
CL tiger hungry hungry stomach  
'The tiger was very hungry.' (Jaisser 1987: 171)
- c. ntau lub tsev  
many CL house  
'many houses' (Jaisser 1987: 171)
- d. lub tsev no  
CL house this  
'this house' (Jaisser 1987: 171)

## (26) Nùng

- a. lúhc mự ni mi tú má câu thàng  
moment once FOC be CL dog nine tail  
'Once there was a nine-tailed dog.' (Saul & Wilson 1980: 100)
- b. tú má khỏp dạ té slóng bặt  
CL dog bite woman that two time  
'The dog bit that woman two times.' (Saul & Wilson 1980: 80)

- c. áu áhn táhng nuhng ma  
take CL chair one come  
'Bring a chair.' (Saul & Wilson 1980: 27)
- d. nahng (áhn) táhng sláy này  
sit CL chair small this  
'Sit on this small chair.' (Saul & Wilson 1980: 30)

The classifier used in the definite has been the same as the one used in the indefinite for the languages shown so far. However, Bangla shows that in singular definites the classifier used is always the general classifier *Ta*, even if a different classifier is used in the indefinite.

## (27) Bangla

- a. laibreri-te Ek jon notun mEthor ar Ek jon  
library-LOC one CL<sup>human</sup> new janitor and one CL<sup>human</sup>  
gard rakha hoechhe  
new guard keep  
'The library hired a new janitor and a new guard.' (Simpson & Biswas 2016: 5)
- b. mEthor Ta porisromi kintu gard Ta besh kuMRe  
janitor CL hard-working but guard CL quite lazy  
'The janitor is hard-working, but the guard is quite lazy.' (Simpson & Biswas 2016: 5)
- c. Ram kalke Ek Ta goru bikri korechhe  
Ram yesterday one Ta cow sell did  
'Ram sold a cow yesterday.' (Simpson & Biswas 2016: 21)
- d. SObai bolchhe goru Ta naki OSuStho chhilo  
everyone is.saying cow CL PRT sick was  
'They say that the cow was not healthy.' (Simpson & Biswas 2016: 21)

This change in classifier for the singular definite is obligatory in Bangla, as shown below.

## (28) Bangla

- a. Ek jon Ofisar  
one CL<sup>human</sup> officer  
'one officer' (Simpson & Biswas 2016: 6)

<sup>1</sup>See Appendix A

- b. <sup>2</sup>Ofisar-Ta/\*jon  
officer-CL  
'the officer' (Surov Syed, p.c.)

Another Eastern Indo-Aryan language, Magahi, behaves similarly to Bangla. Magahi has two classifiers that are in free variation. Familiar definites, rather than using the classifier as in Bangla, are marked with a dedicated familiarity marker *-waa* (Alok 2022). In fact, *-waa* is usually referred to as the definite allomorph of the classifier (Kumar 2020; Lahiri 2021).

(29) Magahi

- a. kal ham e-go/ek Tho kutta dekhali  
yesterday I one-CL/one CL dog saw  
'Yesterday I saw a dog.'
- b. kutt-waa/\*go/\*Tho baRi sundar halai  
dog-FM very beautiful was  
'The dog was very beautiful.'

Given that *Ta* (Bangla) and *-waa* (Magahi) are uniformly used for all nouns, we take this to indicate that they are definite markers rather than classifiers, similar to 'extra' *cài*.

### 3.2 Mass Nouns

The proposal that these classifiers/classifier-like elements have become general definite markers is supported by their widened domain of application. They can attach not only to other count nouns that they usually do not combine with, but also mass nouns which usually require more semantically contentful measure phrases (*massifiers* in the terminology of Cheng & Sybesma 1998).

- (30) a. Bangla  
\*Ek Ta bhaat  
one CL rice  
Intended: 'one unit of rice'
- b. Magahi  
ek muTThi/\*Tho/\*go bhaat  
one handful rice  
'a handful of rice'

- (31) a. Bangla  
Ma olpo bhaat ranna koreche. Bhaat-Ta Tebil-er opor  
mom some rice cook did rice-CL table-LOC on  
rakha ache.  
keep COP  
'Mom cooked some rice. The rice is on the table.'
- b. Magahi  
Babuji thoRe sun bhaat banailkai. Bhat-waa tebal par hai.  
dad little CL rice made rice-FM table on is  
'Dad cooked a little rice. The rice is on the table.'

This observation has also been made for Vietnamese *cài* and the Nùng general classifier *cáh*.

- (32) Vietnamese  
cái thịt này  
cai meat this  
'this meat' (Nguyen 2004: 94)

- (33) Nùng  
a. muhn áu slám déht lợt ma thú  
he take three drop blood come test  
'He took three drops of blood to test.' (Saul & Wilson 1980: 19)
- b. muhn áu cáh lợt má páy  
he take CL blood dog go  
'He took the dog's blood.' (Saul & Wilson 1980: 19)

### 3.3 Analysis

The following table summarizes the data from the previous sections, categorizing the languages according to whether:

- The bare/extra classifier has a classifier that covaries with the noun in the same way as in indefinites.
- There is a single morpheme that marks definiteness for all bare/extra classifier constructions.
- The domain of the definiteness morpheme (if one exists) has extended to mass nouns.

<sup>2</sup>See Appendix B

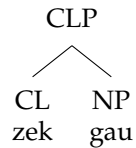
## Bare/Extra Classifier Definites

|            | covarying CL | Def morpheme | Def + Mass N |
|------------|--------------|--------------|--------------|
| Cantonese  | ✓            | ✗            | N/A          |
| Bangla     | ✗            | ✓            | ✓            |
| Magahi     | ✗            | ✓            | ✓            |
| Vietnamese | ✓            | ✓            | ✓            |
| Nùng       | ✓            | ✓            | ✓            |
| Wu         | ✓            | ✓            | ?            |
| Hmong      | ✓            | ?            | ?            |

The languages which have complete data split into three groups. We will use Cantonese, Magahi, and Vietnamese as the representative groups.

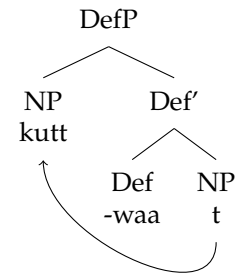
Cantonese bare classifier phrases show the usual classifier covariation with no additional definite morpheme. Thus, the bare CLP structure proposed by Cheng & Sybesma (1999) is perfectly suitable to explain this data.

## (34) Cantonese



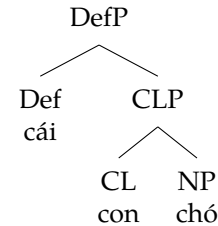
Magahi *-waa* definites do not have classifiers in them, and the use of *-waa* extends to mass nouns. So, rather than a bare CLP like in Cantonese, Magahi seems to project a bare DefP. While in Cantonese the classifier appears to have the unusual ability to perform the referential/deictic function, Magahi appears to have the reverse situation. The primarily referential/deictic head also appears to be able to perform the individuation function associated with classifiers.<sup>3</sup>

## (35) Magahi



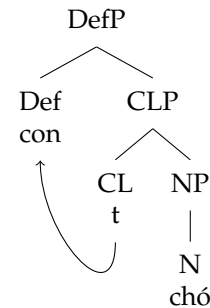
Finally, this brings us to Vietnamese, which has a single morpheme that indicates definiteness in all extra classifier constructions while still showing covarying classifiers. Vietnamese, then, appears to project DefP on top of CLP, as proposed by Simpson (2005); Simpson & Ngo (2018).

## (36) Vietnamese



For the case of the bare classifier construction in Vietnamese, since there is general evidence for a DefP projection above CLP, there is a reason to adopt the proposed CL to D(ef) movement in Simpson (2005).

## (37) Vietnamese

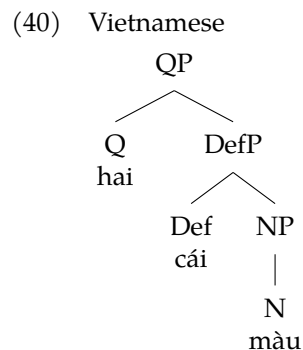
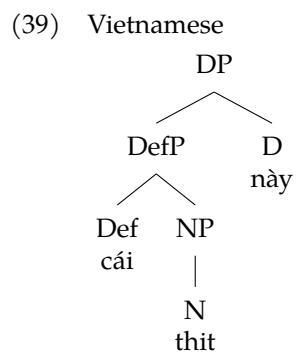


<sup>3</sup>See Appendix C

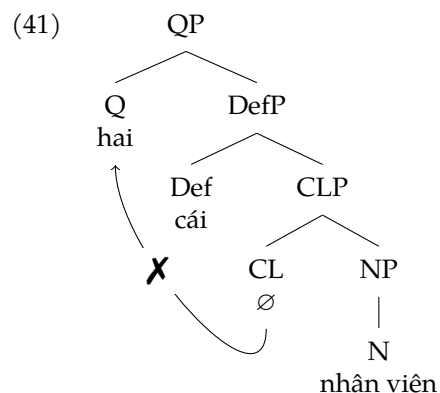
Interestingly, the optionality for optional-classifier nouns goes away with ‘extra’ *cái*, but non-classified nouns (like mass nouns) can still combine with ‘extra’ *cái* (Simpson & Ngo 2018; Nguyen 2004).

- (38) a. hai cái \*(cuốn) sách  
two cai CL book  
‘the two books’ [Obligatory-classifier noun]
- b. hai cái \*(người) nhân viên  
two cai CL employee  
‘the two employees’ [Optional-classifier noun]
- c. hai cái màu  
two cai color  
‘the two colors’ [Non-classified noun]
- d. cái thịt này  
cai meat this  
‘this meat’ [Mass noun]

We propose that (38c) and (38d) are evidence for Vietnamese shifting towards bare DefPs like in Magahi, with the numeral in a higher QP projection and demonstratives either in D or SpecDP (Simpson 2005; Simpson & Ngo 2018).



Finally, the loss of optionality for classifiers such as in (38b) needs to be explained. For these nouns, we assume that there is always a classifier present, but it is sometimes null. This null classifier, however, needs to undergo head movement to a higher projection (say QP in this instance). The occurrence of *cái* in DefP in these instances, however, blocks this head movement, forcing the classifier that can stay in-situ to be used (cf. the analysis of Swedish/Danish definites in Embick & Noyer 2001).



Thus, Vietnamese shows evidence for a split system, somewhere between Cantonese and Magahi, in which there is a marker that is always interpreted as definite and functions referentially/deictically but in certain cases still requires the use of other classifiers for individuation.

## 4 Relation to Focus and Honorificity

This section lays out some other similarities between ‘extra’ *cái* in Vietnamese and related phenomena in Magahi, Bangla, and Nùng as well as possible avenues for explanation, but no complete analysis.

Nguyen (1975, 2002, 2004); Đoàn et al. (2019); Phan & Dong (2021) all report that ‘extra’ *cái* is a focus marker. Part of the evidence for this, besides the intuition that ‘extra’ *cái* serves to focus on a particular individual, is that ‘extra’ *cái* is always stressed, while the classifier *cái* never is. Another fact seemingly related to focus is that some authors report ‘extra’ *cái* is only possible if the noun phrase contains a demonstrative or another kind of post-nominal modifier, such as a relative clause (Nguyen 2004; Đoàn et al. 2019; Phan & Dong 2021). This could indicate that ‘extra’ *cái* is sensitive to the presence of alternatives, a hallmark of focus Rooth (1992, 2014).

The connection between Bangla DPs and focus has also been explored by Syed (2015), who argues that DP-internal movements in Bangla are driven by focus/information structure. Finally, there might be another connection with the kinds of definites Magahi and Bangla encode with *-waa*/*Ta* respectively. Magahi *-waa* seems to not combine with large situation/global unique definites, e.g. ‘the sun’. However, in hypothetical situations where our solar system has two suns, speakers find it acceptable to use *-waa* (and a demonstrative if necessary) to talk about one of the suns. In a similar vein, Bangla *Ta* is compatible with ‘moon’ in certain scenarios but not others. Reference to the moon at a particular point of time allows for the use of *Ta* (42a), but stating a general property of the moon does not (42b).

(42) Bangla

- a. gOto rat-e chaMd Ta jhOlmOl korchhilo  
last night-LOC moon CL bright do  
‘The moon was very bright last night.’ (Simpson & Biswas 2016: 11)
- b. chaMd (\*Ta) prithibi-r upogroho  
moon CL earth-GEN satellite  
‘The moon is a satellite of the earth.’ (Simpson & Biswas 2016: 11)

These data points show that *-waa*, *Ta* might also be sensitive to alternatives. In the Magahi example, the hypothetical existence of a second sun creates an alternative. In the Bangla examples, statements about the moon at a particular time also allow for alternatives (e.g. the night before last night’s moon, last night’s moon, tonight’s moon, etc.). Whether a unified analysis of these facts

is possible or even warranted, is not clear at this point.

Finally, Nguyen (2004) reports that, while they do not agree with this assessment, *cái* is often claimed to have a perjorative connotation. Parallels are found in Magahi and Bangla once again. In Magahi, a speaker can use *-waa* with proper names, but only for those of lower social status, e.g. younger siblings but not older siblings or parents. Using *-waa* with the name of a social superior/elder is considered very disrespectful, and *-waa* can also be used to show contempt with other uses on common nouns (see Alok 2022 for details). Similarly, Bangla *Ta* can be used to express the speakers negative views towards another person. In the following example, the speaker is intending to show respect to the (former) peon and disrespect/contempt toward the officer.

- (43) 1990 Sal-e Ofis-e Ek jon pion ar Ek jon  
1990 year-LOC office-LOC 1 CL<sup>human</sup> peon and 1 CL<sup>human</sup>  
Ofisar rakha hoechhilo. Sei peon (\*Ta) Ekhon CEO hoe gEche,  
officer hire was that pion CL now CEO be went  
kintu Ofisar \*(Ta) khub kuMRe chhilo bole kono unnoti  
but Officer CL very lazy was COMP any improve  
korte pareni.  
do-PPL be-able-NEG  
‘The office hired a peon and an officer in the year 1990. The peon is now the CEO of the company; however, the officer was very lazy, so he could not get any promotion.’ (Simpson & Biswas 2016: 6)

Finally, (Saul & Wilson 1980: 26) reports that classifiers can occur with proper names in Nùng, but only for children. Thus, this use of classifiers might also be related to social hierarchy/honorificity.

## 5 Conclusion

Bare/extra classifier definites show a range of similar and related behaviors in various languages. We have tried to identify these similarities and build on the work of Cheng & Sybesma (1999, 2012); Simpson (2005); Simpson & Ngo (2018) by proposing a system that can capture various instantiations of the use of classifiers in definite expressions.

## References

- Alok, Deepak. 2022. The discourse particle-waa in magahi nominal domain. *Discourse* .
- Arkoh, Ruby & Lisa Matthewson. 2013. A familiar definite article in akan. *Lingua* 123. 1–30.
- Bale, Alan & Jessica Coon. 2014. Classifiers are for numerals, not for nouns: Consequences for the mass/count distinction. *Linguistic Inquiry* 45(4). 695–707.
- Cheng, Lisa Lai-Shen. 2012. Counting and classifiers. *Count and mass across languages* 199–219.
- Cheng, Lisa Lai-Shen & Rint Sybesma. 1998. Yi-wan tang, yi-ge tang: Classifiers and massifiers. *Tsing Hua journal of Chinese studies* 28(3). 385–412.
- Cheng, Lisa Lai-Shen & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of np. *Linguistic inquiry* 30(4). 509–542.
- Cheng, Lisa Lai-Shen & Rint Sybesma. 2005. Classifiers in four varieties of chinese. In Guglielmo Cinque & Richard S. Kayne (eds.), *The oxford handbook of comparative syntax*, 259–292. New York: Oxford University Press.
- Cheng, Lisa Lai-Shen & Rint Sybesma. 2012. Classifiers and dp. *Linguistic inquiry* 43(4). 634–650.
- Chierchia, Gennaro. 1998. Reference to kinds across language. *Natural language semantics* 6(4). 339–405.
- Embick, David & Rolf Noyer. 2001. Movement operations after syntax. *Linguistic inquiry* 32(4). 555–595.
- Frege, Gottlob. 1892. On sense and reference. In *P. geach and m. black (eds), translations from the philosophical writings of gottlob frege*, 56–78.
- Heim, Irene. 1983. File change semantics and the familiarity theory of definiteness. meaning, use and interpretation of language, ed. by reiner bauerle, c. schwarz and arnim von stechow, 164–89.
- Hsu, Brian & Saurov Syed. 2020. Cooccurrence of nonarticle determiners as evidence for split dps and bundled projections. *Language* 96(3). e114–e134.
- Jaisser, Annie. 1987. Hmong classifiers. *Linguistics of the Tibeto-Burman area* .
- Jenks, Peter. 2015. Two kinds of definites in numeral classifier languages. In *Semantics and linguistic theory*, 103–124.
- Jenks, Peter. 2018. Articulated definiteness without articles. *Linguistic inquiry* 49(3). 501–536.
- Kumar, Chandan. 2020. *Typology of nominal modifiers in major languages of bihar: Magahi, bhojpuri, maithili, angika & bajjika*: Jawaharlal Nehru University dissertation.
- Lahiri, Bornini. 2021. *The case system of eastern indo-aryan languages: A typological overview*. Routledge India.
- Nguyen, Phu Phong. 2002. *Những vấn đề ngữ pháp tiếng việt: loại từ và chỉ thị từ*. [aspects of vietnamese grammar: Classifiers and demonstratives]. Nhà xuất bản Đại học quốc gia Hà Nội.
- Nguyen, Tai Can. 1975. *Ngu pháp tiên việt: tiếng-từ ghép-doan ngu*. [Vietnamese grammar: morphemes, compounds and phrases]. Dai-hoc and Trung-hoc Chuyennghiep, Hanoi.
- Nguyen, Tuong Hung. 2004. *The structure of the vietnamese noun phrase*. Boston University.
- Đoàn, Quỳnh Ngọc Thị, Martin BH Everaert & Eric J Reuland. 2019. (in) definiteness of vietnamese noun phrases. In *Interdisciplinary perspectives on vietnamese linguistics*, 155–180. John Benjamins Publishing Company.
- Owusu, Augustina P. 2022. *Cross-categorial definiteness/familiarity*: Rutgers The State University of New Jersey, School of Graduate Studies dissertation.
- Phan, Trang & Lam Quang Dong. 2021. Decomposing definiteness in vietnamese. *Journal of the Southeast Asian Linguistics Society* 14(1). 1–18.
- Roberts, Craige. 2003. Uniqueness in definite noun phrases. *Linguistics and philosophy* 26. 287–350.
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural language semantics* 1(1). 75–116.
- Rooth, Mats. 2014. Alternative semantics .
- Russell, Bertrand. 1905. On denoting. *Mind* 14(56). 479–493.
- Saul, Janice E. & Nancy F. Wilson. 1980. Nung grammar .
- Schwarz, Florian. 2009. *Two types of definites in natural language*. University of Massachusetts Amherst.
- Schwarz, Florian. 2013. Two kinds of definites cross-linguistically. *Language and linguistics compass* 7(10). 534–559.
- Štereikaitė, Milena. 2019. Strong vs. weak definites: Evidence from lithuanian adjectives. *Studies in Diversity Linguistics* .
- Simpson, Andrew. 2005. Classifiers and dp structure in southeast asia. In Guglielmo Cinque & Richard S. Kayne (eds.), *The oxford handbook of comparative syntax*, 806–838. New York: Oxford University Press.
- 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 & Binh Ngo. 2018. Classifier syntax in vietnamese. *Journal of East Asian Linguistics* 27. 211–246.
- Simpson, Andrew, Hooi Ling Soh & Hiroki Nomoto. 2011. Bare classifiers and definiteness: A cross-linguistic investigation. *Studies in Language. Interna-*

*tional Journal sponsored by the Foundation "Foundations of Language" 35(1). 168–193.*

- Simpson, Andrew & Saurov Syed. 2016. Blocking effects of higher numerals in bangla: A phase-based analysis. *Linguistic Inquiry* 47(4). 754–763.
- Syed, Saurov. 2015. Focus-movement within the bangla dp. In *Proceedings of wccfl*, vol. 32, 332–341.
- Syed, Saurov. 2016. Decomposing definiteness: Arguments for a split d-domain in bangla. In *Proceedings of the 33rd west coast conference on formal linguistics*, 390–397.
- Syed, Saurov & Andrew Simpson. 2017. On the dp/np status of nominal projections in bangla: Consequences for the theory of phases. *Glossa: a journal of general linguistics* 2(1).

into the same group as Magahi in our classification, the bundling of the referential/deictic function and the individuation function is perhaps more clear, given that the form of the definite marker is the same as the general classifier. It should not necessarily be assumed that *Ta* still has an individuation function synchronically, however. But we do not touch on these issues further here.

## A Wu Tone in Definites

Note the tone change on the classifier in the definite examples in (24). Cheng & Sybesma (2005) show that while not all classifiers in definite noun phrases have the same tone, the tone of the classifier undergoes a predictable lowering depending on what the underlying tone of the classifier is. If Wu tone is characterized on a 3 point scale (HML), then the tone changes can be classified as  $H \rightarrow M$ ;  $M, L \rightarrow L$ . We take this to indicate that a floating L tone is the definite marker in these constructions in Wu. See Cheng & Sybesma (2005) for details.

## B Bangla Definites

Syed (p.c.) reports that while the use of the human classifier *jon* post-nominally cannot give the desired singular definite reading (cf. 28b), it can give a class reading. For example, *mEthor jon* can mean ‘the class of janitors’, but not ‘the janitor’ (cf. the use of *mEthor Ta* to mean ‘the janitor’ in 27b).

## C Magahi Classifiers

The assumption in Section 3.3 that the lack of classifier in Magahi familiar definites indicates that *-waa* in Magahi can perform the individuation function usually associated with classifiers in addition to a referential/deictic function is not the only possibility. It could be that Magahi is the kind of language described in Bale & Coon (2014), where the classifiers are only necessary for the numeral to be able to combine with an NP, rather than the classifiers being present to provide the necessary structure for the NP as described by Chierchia (1998). We do not rule this possibility out for Magahi. For Bangla, which fell