The semantics of Hmong (in)definiteness

Quartz Colvin

May 13, 2025

1 Introduction

This project explores the definiteness, indefiniteness, and number semantics of nominal elements in White Hmong (Hmong-Mien). These elements include the indefinite article *ib*, classifiers *tus* and *cov*, three demonstratives *no*, *ko* and *ntawd,* and their possible DP configurations. Our research question is formulated in two parts. The first question is about the individual meanings of *ib, tus* and *cov*, and the second question is how these meanings compose to form the possible DP configurations in Hmong.

Exploring this in Hmong is interesting for two reasons. First, this paper provides a semantic account of Hmong's plural classifier. The other reason is that the indefinite marker in the language is a choice function. Analyzing it as such provides a more uniform explanation of the compositional semantics of plural indefinite phrases. Empirically, there is another reason to explore this in Hmong. Hmong is understudied and most linguistic sources and documentation of the language are from the 1900s, so this project will be a fresh view of how the language is used by speakers today.

The roadmap of this paper is as follows. First is some background information on Hmong §2 and then present the data §3 analyzed later in the paper. After presenting the data, the existing literature on nominal semantics will be discussed §4 and we will explain how the current literature cannot adequately address the data in Hmong §5. After this, the syntactic and semantic analyses will be presented §6. Finally, I discuss the implications of this project §7 and conclude.

2 Background on Hmong

This section serves as both an overview of the writing system this paper uses and Hmong's nominal domain. White Hmong's basic, pragmatically unmarked word order is SVO, but occasionally SOV is permitted, though it is unclear what exactly conditions SOV word order. Hmong is also known to have serial verb constructions (Riddle 1989) and a complex tonal system. It has seven tones (most of which are contour tones) in addition to distinguishing creaky and breathy voice. In the orthographic system this article uses (the Romanized Popular Alphabet), the tones are represented by the final written letter of the word (Figure (1)). Hmong is a classifier language, and its classifiers can encode shape

Figure 1: Tonal Inventory (Garellek and Esposito 2021)

	Description	RPA letter
(a.)	High-rising	-b
(b.)	Mid-level	∅
(c.)	Low-level	-s
(d.)	High-falling	-j
(e.)	Mid-rising	-v
(f.)	Low-falling *creaky*	-m
(g.)	High-falling *breathy*	-g

information (roundness, stacks or sheets of something, etc.), number information (plurals, groups or singular nominals), animacy information (humans and animals vs. inanimate objects), and definiteness. The main classifier in the data here is *tus* which is generally used for animate beings (humans and animals) and encodes definiteness, uniqueness, and singular number. To demonstrate the order (1) of nominal elements in Hmong, (2) shows the order of these in a Hmong sentence.

(1) [NUM + CLF + N + ADJ + DEM]

(2) Kuv pom [tzi tus aub nthaws ntawd]
1SG see [five CLF dog fat DEM]
"I see five fat dogs (over there)."

There is one existing source that mentions definite and anaphoric reference of DPs in Hmong. Bisang (1993) reports that Hmong has bare classifier phrases[1] and bare classifier phrases with demonstratives. The former has 'weaker referential power' than the demonstrative classifier phrases. Additionally, bare classifier NPs in Hmong typically have anaphoric reference (Bisang 1993). However, this is all that is said about the fine-grained semantics of these DP configurations in the language.

As for Hmong's pronoun system, the same form is used in all argument positions (3)-(4).

(3) kuv pom koj
1SG see 2SG
"I see you."

(4) koj pom kuv
2SG see 1SG
"You see me."

In (3) the subject 1SG pronoun is *kuv* and in (4) the object 1SG pronoun is also *kuv*. The pattern is the same for 2SG. For subject and object positions, this is the same for all singular, dual, and plural pronouns in Hmong.

3 Hmong nominal data

All data in this project comes from two White Hmong speakers from Wisconsin: one is 24 (KX), and the other is 34 (YX), and they are siblings. The Hmong examples are all written in the Romanized Popular Alphabet (RPA) as it is the most-utilized orthography by Hmong speakers in the US. Elicitation with YX was conducted on Zoom with cameras on, and elicitation with KX was conducted in person in New Jersey. In Zoom elicitations with YX, I prepared documents with images for the contexts or the English sentences for translation tasks, and during the Zoom meeting, I shared my screen so we could go through the material together. These Zoom meetings were video and audio recorded, but the video files were discarded (at the consultant's request).

Now let us turn to the nominal data this project investigates and note the particularly puzzling configurations. Figure (2) shows which DP configurations are possible in Hmong. Bare nouns and bare classifier phrases are acceptable DPs. Indefinite phrases are built from the bare classifier phrase, and demonstratives can combine with either an indefinite phrase or a bare classifier phrase.

Figure 2: Possible DP Configurations in Hmong

	Configuration	Example	Rough translation
1.	CLF + NP	*tus aub*	'the dog' (unique)
2.	CLF + NP + DEM	*tus aub ntawd*	'that dog' (exophoric)
3.	INDEF + CLF + NP	*ib tus aub*	'a dog' (non-specific)
4.	INDEF + CLF + NP + DEM	*ib tus aub ntawd*	'a dog' (exophoric)
5.	NP	*aub*	'dog'

It is important to note that whenever there is an indefinite article in Hmong, the classifier is still required. Demonstratives are noteworthy since they are grammatical with both definite and indefinite

[1] A "bare classifier phrase" is a phrase that only contains the noun and its classifier.

nominal bases. Additionally, bare nouns are possible in Hmong, but they are interpreted as kinds (5). For example, *tsov* is a bare noun, and if it had occurred with a classifier here, it would be *tus* (Heimbach 1980).

(5) Tsov ntxaus ntxaus ntshai
tiger intensifier REDUP fear
"Tigers are dangerous."

Figure (2) highlights a few challenges for a semantic account of Hmong DPs. First, it is unclear why an indefinite article *ib* with a classifier would result in a non-specific indefinite reading. If *ib* is a generalized quantifier, as indefinite articles often are, this does not explain why *ib* requires a classifier. Second, it is also unclear how combining an indefinite article with the maximal group classifier results in the translation 'some NP.' Lastly, it is interesting that an indefinite article like *ib* can co-occur with a demonstrative.

The two classifiers this paper focuses on are *tus* and *cov*. *Tus* has a singular definite reading and *cov* has a maximal definite reading in bare classifier phrases. In other words, *cov* selects a maximal group referent in the context (6a), while *tus* selects a single, unique referent in the context (6b).

(6) CONTEXT: Imagine you're in a room with three cats. How would you say 1) Keng sees the cats, referring to all of the cats? How would you say 2) Keng sees the cat, referring to a specific one of those three cats?

a. Keng pom cov miv
Keng see CLF cat
"Keng sees [all] the cats."

b. Keng pom tus miv
Keng see CLF cat
"Keng sees the [unique] cat."

Classifiers in Hmong cannot be used as anaphoric pronouns (7), at least for my consultants. This was unexpected because, as noted by Mottin (1978), some classifier languages allow their classifiers to be used as 3rd person anaphoric pronouns; even when there is no noun.[2] When asking about this construction, I said the Hmong example in (7) and asked the consultant if it made sense. The consultant said, "No, you can't just say *'I love the'*, which is basically what that sounded like to me."

(7) *Kuv pom tus aub$_i$. Kuv hlub tus$_i$.
1SG see CLF dog. 1SG love him.
"I see the dog. I love him"

As for Hmong indefinite configurations, *ib* always occurs with a classifier (8b) and it cannot be combined directly with a noun (8c). It always has a non-specific meaning.

(8) CONTEXT: Imagine you can see there is a dog on the other side of the barn. You don't know whose dog it is- you suspect it is probably a stray dog. How would you say "I see a dog" in this context? [This corresponds to the example in *b.* below]

a. FOLLOW-UP: Could you say this sentence without the "the" word?[3] [Then I repeated the Hmong sentence in *c.* back to them.]

b. kuv pom ib tus aub
1SG see INDEF CLF dog
"I see a dog."

c. *kuv pom ib aub
1SG see INDEF dog
"I see a dog."

[2]This may still be possible, but I was unable to elicit any grammatical examples.

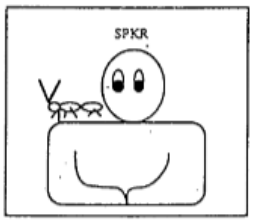
[3]Both of the consultants relate the singular Hmong classifiers to "the" in English, so they regularly call classifiers "the" or "'the' words."

I made it clear that neither of the people in the conversation knows who the dog is to make sure it is an unfamiliar dog. Also I asked if it would be said differently in this context if the addressee could see the dog too, and the consultants said it would still be *ib tus aub*. This indicates that *ib tus aub* truly does not require shared familiarity.

This data fits into Chierchia's typology as a 'Chinese-type' language (Chierchia 1998). Requiring a classifier is not novel in languages of this pattern. The denotation of this indefinite article *ib*, however, will have to be more than simply a generalized quantifier to explain why it has this requirement.

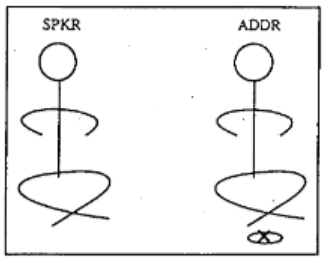
The final category discussed in the nominal domain is demonstratives. Hmong has three demonstratives: *no* (9), *ko* (10), and *ntawd* (11). *No* is the demonstrative used for referents near the speaker, while *ko* is used for referents near the addressee, and *ntawd* is the 'distal' demonstrative.

(9) CONTEXT: Imagine you are the person in this picture[4] and you just noticed there's a bug on your shoulder. How would you say that that bug is bothering you in Hmong?



a. tus kab no thab kuv
CLF bug DEM bother 1SG
"This bug (here) is bothering me."

(10) CONTEXT: Imagine the person on the left of this picture is you. You can see the book is right in front of the listener, but not within your reach. How would you say "I like that book" in Hmong in this context?



a. kuv nyiam phau-ntawv ko
1SG like CLF-writing DEM
"I like that book (by you)."

(11) CONTEXT: Imagine you're talking to someone and you notice a dog running around the yard, but it is far from both you and the person you're talking to. How would you say "I see a dog (over there)" in Hmong in this context?

a. kuv pom ib tus aub ntawd
1SG see INDEF CLF dog DEM
"I see a dog (over there)."

All of these Hmong demonstratives can occur with both definite (12) and indefinite bases (13). Since demonstratives are normally associated with definiteness, it is surprising that a phrase like (13) is grammatical.

(12) CONTEXT: Imagine you are talking to your sister (you both are very familiar with her dog Apollo) and you see Apollo alone in the yard. How would you say "I see the dog" in Hmong in this context?

a. kuv pom tus aub ntawd
1SG see CLF dog DEM
"I see the dog (over there)."

(13) [repeated from (11)]

a. kuv pom ib tus aub ntawd
1SG see INDEF CLF dog DEM
"I see a dog (over there)."

[4]These images are from Wilkins (1999). I created a worksheet with fill-in-the-blank tasks that the consultants and I went through in the demonstrative elicitation sessions.

Throughout the rest of this paper, these two demonstrative constructions will be called DDB and DIB (14).

(14) DDB vs. DIB
   a. DDB: demonstratives with definite bases (Hmong bare classifier phrases)
   b. DIB: demonstratives with indefinite bases (Hmong *ib*-phrases)

4 Literature on nominal semantics

Now that the Hmong data has been presented, we should zoom out and discuss definiteness and indefiniteness across language. After that, the paper turns to the typology of bare nouns, kinds, and plurality.

4.1 Definiteness

The general meaning of 'definiteness' is largely agreed upon, but the finer details differ between analyses. This definition is a good summary of definiteness as a general phenomenon.

> "[Definiteness] presupposes the existence and the unicity of the noun to which it refers. In addition, it implies that the referent has been previously introduced in the discourse and that it is familiar [in the sense of (Heim 1983)] to the speaker and to the hearer." (Pozzo 2022).

One of the oldest theories of definiteness treats a definite article like 'the' as a quantifier that asserts the existence and uniqueness of the nominal's descriptive content (Russell 1905). Others have said that existence and uniqueness are requirements that license the definite article as a referential element (Strawson 1950). These are all true in some way, but this paper follows a more recent notion of definiteness that describes the connection between uniqueness and existence (Roberts 2003). Researchers who categorize definite articles as referential also strongly correlate familiarity with the definite article (Heim 1983; Roberts 2003). For some authors, this familiarity property is all the definite article presupposes or requires (Szabó 2000).

Crosslinguistically, languages encode definiteness in different ways and many do not use a definite article. For example, some languages like Mandarin and Korean do not overtly mark definiteness (Cheng & Sybesma 1999; Lee 1992).

(15) gou yao guo malu (Mandarin; Cheng & Sybesma 1999)
dog want cross road
"The dog wants to cross the road." and **A dog wants to cross the road.*

(16) cha-rul takk-ass-ta (Korean; Lee 1992)
car-ACC wash-PAST-DECL
"I washed the car." or *"I washed cars."*

These Korean (16) and Mandarin (15) examples do not have classifiers in them and neither language has a definite article. Additionally, (16) is ambiguous between a singular definite reading and a plural reading of the noun 'car'.

4.1.1 Weak vs. strong definites

There are two types of definite articles cross-linguistically: strong and weak. Weak definite articles only have the uniqueness property and strong articles involve anaphoricity (Schwarz 2009). Anaphoricity entails familiarity, so we can say that strong articles have both the uniqueness and familiarity properties of definites.

Some languages mark these types of definites overtly have different forms for strong and weak definites and other languages do not mark the distinction (Royer 2022: 105). For example, the English

definite article never changes form based on which kind of definite it is. On the other hand, there are languages like Lakhota that mark strong (17c) and weak definites (17a) differently (O'Gorman 2011).

(17) a. eya' măpíya naŋ măkhá **kiŋ** lená thŏká-kágă-pi k'uŋ hé
well sky and earth **the**$_{weak}$ DEM.PL first-create-pass DEF DEM
"'Well, at the time when the sky and earth were created....'"

b. eháŋ ma-thŭŋ-pe ló
then 1-born-pass ASSRT
"...I was born.

c. akícita **k'uŋ** héna
soldiers **the**$_{strong}$ DEM.PL
"the soldiers mentioned before"

One of the ways to know if a definite marker is a weak definite or not is if it can refer to globally unique entities. In the Lakhota examples, *kiŋ* is the weak definite article (17a) since it can refer to a globally unique thing such as the earth. Strong definites involve reference to something already mentioned in Lakhota, so the strong definite *k'uŋ* is shown in (17c).

4.1.2 Definiteness in classifier languages

In some other languages, definiteness can be shown by classifier placement. For example, in Bangla, if the classifier comes after the noun (18a), then the noun has a definite reading. If the classifier is before the noun (18b), then the nominal is indefinite (Dayal 2012).

(18) a. boi ta
book CLF
"the book"

b. ɛk ta boi
one CLF book
"a book" or "one book"

Another property of classifier languages is that classifiers give rise to definite readings. One example is Vietnamese (19).

(19) Nguoi chong rat tot (Vietnamese; Daley 1998)
CLF husband very good
"The husband was very good."

Classifier languages may or may not have a definite article, but it is most common for a classifier language to not have one (Jiang 2017). Nuosu Yi and Indonesian are some of the few examples of languages that have both classifiers and at least one definite article.

(20) mu uo **ma su** nra jjy nra. (Nuosu Yi: Jiang 2017)
horse three **CLF DEF** fat very fat
"The three horses are very fat."

(21) a. Ada pun tatkala mêmbuat rumah **itu** (Indonesian: Chung 2000)
as for when make house **the**

b. tiga **orang** orang China kuli jatoh dari atas
three **CLF** person Chinese laborer fell from top
"In the course of its construction three of the Chinese workmen fell from the top."

In the Nuosu Yi example (20), the morpheme *su* appears at the end of the nominal domain and makes the whole DP definite, indicating that *su* is a definite article (Jiang 2017). The definite article in Indonesian is *itu* (21).

4.1.3 Demonstratives

Demonstratives are also a type of definite description (Ahn 2022). In the literature, there are two main approaches to analyzing demonstratives: deictic and anaphoric accounts. The deictic approach treats demonstratives as directly referential (Kaplan 1969) and the other approach says that demonstratives are another form of anaphora (Roberts 2002).

It has been noted that demonstratives presuppose uniqueness like definite articles do, but their uniqueness presuppositions are evaluated within different situations (Wolter 2006). The difference between demonstratives and definite articles under this idea is that demonstratives are evaluated in a non-default situation and definite articles are evaluated in a default situation (Stalnaker 1977). Default situations are topic situations associated with the main predicate and they determine the truth value of the whole proposition. Non-default situations are resource situations and are associated with nominal constituents. Both of these situations fix the reference of referential expressions like demonstratives.

Others have said that all uses of demonstratives are definite and they introduce a G(x) argument which is an index to an entity (Nowak 2019).

(22) Demonstratives (Nowak 2019)

a. that F = the x: [F(x) & G(x)]

b. Keng bought a book. That book was long.
$\rightarrow$ [[that book]] =ιx: book(x) $\wedge$ x=y

Some accounts of demonstratives also incorporate gestures into the semantics, since demonstratives are often used alongside gestures such as pointing. One such account claims that demonstratives are an indirect type of referential expression that takes an additional argument: either a gesture or an anaphoric index (Ahn 2022). Under Ahn's account, pointing gesture argument contributes direct reference, which is a way to explain the differences between direct and indirect uses of demonstratives.

4.2 Indefiniteness

In general, indefinites can be diagnosed by whether the referent is identifiable in the context or not. If the referent is not identifiable in the context, the referent is said to be indefinite (Heim 1983).

Cross-linguistically, indefinite expressions are often marked by indefinite determiners, usually derived from the word for 'one' in the language (Royer 2022). On the other hand, many languages show indefiniteness with bare nouns. (23) shows this for Hindi and (24) shows this for Russian.

(23) kamre meN cuuhaa hai (Hindi; Dayal 2004)
room in mouse is
"There's a mouse in the room."

(24) V komnate byli mal'chik i devocka (Russian; Dayal 2004)
in room were boy and girl
"A boy and a girl were in the room."

It has also been proposed that semantically, indefinites denote generalized quantifiers (Elbourne 2005). Support for this comes from English, where indefinites can have a wide scope reading over negation (25).

(25) Plurals and indefinite scope (Carlson 1977)

a. John didn't read a book. ($\neg\exists$ and $\exists\neg$)

b. John didn't read books. (only $\neg\exists$)

Another feature of indefinites is that indefinites do not need to denote a familiar entity. Dayal (2004) notes that singular bare nouns can only ever denote one unique individual in a situation, but these uses do not require that the referent is familiar.

The English word *some* is historically categorized as an existential plural determiner or a plural equivalent to *a*. More recently, authors have argued that determiners like *some* are choice functions, in agreement with literature from the 90s (Collins 2025). A choice function is a function that applies to a set and arbitrarily selects one of the members of the set. So, for a sentence like (26a), the choice function f selects x from the set of cats, yielding (26b) (Winter 1997).

(26) a. John loves some cat.
b. $\exists f[love(John, f(cat))]$

Some is similar to the English indefinite article *a* in that they both allow for wide-scope readings. But, they differ in whether they also allow generic readings or not (27).

(27) a. A dog barks. [generic interpretation possible]
b. Some dog barks. [generic interpretation impossible]

The contrast between the two sentences in (27) shows that *some* does not block a covert $\exists$ with plurals, while *a* does.

4.3 Noun typology

Generally nouns have two roles. Firstly they are predicates and restrictors of quantifiers. Second, they are arguments and are used for kind reference. Across the typology this section discusses, the way these options are actualized is different. This is instantiated as +/-arg and +/-pred features.

If a noun is [+predicate] (+pred), then members of that category can be predicates of type $\langle e,t\rangle$. If a noun is [+argument] (+arg), then members of that category can be type e or the type of generalized quantifiers ($\langle\langle s,\langle e,t\rangle\rangle,t\rangle$). Attested combinations of these features are shown in (28).

(28) Noun typology (Chierchia 1998)
a. Type A: [+arg, -pred]
b. Type B: [-arg, +pred]
c. Type C: [+arg, +pred]

Mandarin Chinese is the prototypical language of the first type in Chierchia's typology, which this paper refers to as Type A (29). This class of languages has the following features: bare nouns are interpreted as kinds, all bare nouns are mass nouns, there is no plural morphology on nouns, and they have general classifier systems.

(29) Type A languages [+arg,-pred]
a. generalized bare arguments
b. the extension of all nouns is mass
c. no PL
d. generalized system of classifiers

The difference between count and mass is neutralized in languages of this type since 'rice' (a semantically mass noun) and 'table' (a count noun) both need a classifier to count. (30a)-(30b) and (30c)-(30d) also show that there is no plural morphology anywhere in the phrase except the explicit numeral.

(30) a. yí lì mǐ
one CLF rice
"one (grain of) rice"

b. liǎng lì mǐ
two CLF rice
"two (grains of) rice"

c. yí zhāng zhuōzi
one CLF table
"one (piece of) table"

d. liǎng zhāng zhuōzi
two CLF table
"two (pieces of) tables"

In Type A languages, bare nouns are mass nouns by default, which explains why the bare nouns for 'rice' and 'table' both require a classifier to count. The other two types of languages in this typology are Romance as Type B (31) and the example for Type C is Germanic (33).

Romance languages (and others of Type B) have little to no tolerance for bare arguments. Plural morphology can be added directly to count nouns, including when counting, and mass nouns do not take plural morphology. Either classifier phrases or measure phrases are required for mass nouns.

(31) Type B languages [-arg, +pred]
- a. no or highly restricted occurrence of bare arguments
- b. count nouns with PL morphology and the possibility of directly combining with numerals
- c. mass nouns without PL morphology
- d. classifiers or measure phrases obligatory for mass nouns

(32) a. *cane ama giocare fuori
dog.SG loves to.play outside
"Dog loves to play outside."

b. *cane ama sta giocando è raro
dog.SG loves AUX play COP rare
"Dog loving to play outside is rare."

c. i cani amano giocare
DEF dog.PL loves to.play
"Dogs love to play."

d. *cani amano giocare fuori
dog.PL love to.play outside
"Dogs love to play outside."

e. *cani amano stanno giocando sono rari
dog.PL love AUX playing COP rare
"Dogs loving to play outside is rare."

f. l-acqua sta gocciolando
DEF-water AUX dripping
"Water is dripping."

Italian data from Chierchia (1998) shows that bare arguments are dispreferred (32). Neither singular nor plural count nouns can yield a kind reading. For kind readings, Italian uses a definite article with a plural noun (32c). Mass nouns do not have a plural form but still require a definite article (32f).

Germanic, as the Type C example, has the following nominal characteristics (33).

(33) Type C languages [+arg, +pred]
- a. mass/count contrast
- b. bare mass arguments
- c. bare count plural arguments
- d. no bare count singular arguments

(34) a. *Dog loves to play outside.

b. Dogs love to play outside.

c. Gible loves rice.

d. *Gible loves rices.

(34) shows that plural bare count nouns are kinds (34b) while singular bare nouns are not kinds (34a). Mass nouns in a language like English do not allow plural morphology (34d), but singular mass nouns retain a general mass reading (34c).[5] In sum, Type C languages are a mixture of the properties that Type A and Type B languages exhibit.

[5]Gible is the name of a tabby cat in this case, not a Pokémon.

4.3.1 Plurals

Across the literature on plurals, pluralities are modeled as sets and their structure is ordered with a part-of relation '$\leq$'. In (35)-(36), the lower-case variables correspond to individual dogs in the set, and 'AT(x)' means that x is atomic.

(35) **Pluralities** (Chierchia 1997)

a. "dogs" (pluralities) = [{f, b, s}, {f, b}, {f, s}, {b, s}]

b. "dog" = [f b s]

c. PL = $\lambda P \lambda x[\neg P(x) \wedge \forall y[y \leq ATx \rightarrow P(x)]]$

When the ι operator (or definite article, depending on the language) is added, it picks the largest plurality in the noun's extension. For example, when there is plural morphology, the operator picks out the largest plurality of the noun. A brief paradigm of how this works for definite pluralities is exhibited in (36). (36a) shows a blueprint for how the ι operator combines with singular or plural nouns. (36b) shows that ι plus a singular noun yields the unique singular dog in the context, and (36c) shows that combining ι with a plural noun yields the largest plurality of dogs in the context.

(36) **Definite plurals** (Chierchia 1997)

a. ιX = the largest member of X if there is one (else, undefined)

b. "the dog" = ι DOG = the only dog (if there is one)

c. "the dogs" = ι DOGS = the largest plurality of dogs

5 Taking stock

Given this short review of the literature on definiteness, indefiniteness, and number, there are a few questions that should stand out regarding the Hmong data presented in §3. First, it is odd that an indefinite article *ib* would result in a non-specific indefinite reading even though it still requires a classifier. If *ib* is a generalized quantifier, as indefinite articles are often described, that does not explain why *ib* has this classifier requirement. Second, it is unclear how combining an indefinite article with a 'maximal group' classifier results in the translation 'some NP.' Lastly, it is interesting that an indefinite article like *ib* can be used with demonstratives at all.

Now that we have broadly discussed definiteness and indefiniteness, we can zoom in on a case study of (in)definiteness in another classifier language: Chuj (Mayan). After the analysis of Chuj is explained, many of the tools can be extended to account for the Hmong data, but not everything works for Hmong. At that point, the present paper addresses the specific aspects of the analysis that need to be adjusted and transitions to the formal analysis of Hmong DPs.

5.1 Case study: Chuj

This case study explores several topics in the syntax and semantics of nominal expressions in Chuj, focusing on definite, indefinite, demonstrative, and pronominal elements (Royer 2022). Since there is minimal formal literature on definiteness and indefiniteness in classifier languages, this case study serves as a primary toolkit that needs to be modified to explain the Hmong data.

Before explaining Royer's analysis, we first need to explain the Chuj data. Figure (3) shows all of the possible configurations of nominal elements in the language. Most logically possible combinations of the indefinite article, classifiers, nouns, and deictic particles are available in Chuj, with each configuration yielding a different semantic representation. Bare classifiers and bare classifier phrases are grammatical; indefinites can combine with bare nouns or bare classifier phrases; and deictic particles can combine with either type of indefinite (rows 6-7) or a bare classifier phrase.

Figure 3: Possible DP configurations in Chuj (Royer 2022: 3)

	Configuration	Example	Rough translation
1.	CLF + NP	*nok' tz'i'*	'the dog'
2.	CLF + NP + DEIX	*nok' tz'i' chi'*	'the/that dog' (anaphoric)
3.	CLF	*nok'*	'it'
4.	INDEF + NP	*jun tz'i'*	'a dog'
5.	INDEF + CLF + NP	*jun nok' tz'i'*	'a (certain) dog'
6.	INDEF + NP + DEIX	*jun tz'i' chi'*	'that dog'
7.	INDEF + CLF + NP + DEIX	*jun nok' tz'i' chi'*	'that dog' (exophoric)

Like most languages with indefinite articles, in Chuj, it is acceptable to use an indefinite article with a bare noun (37a). The singular indefinite determiner *jun* is required for singular indefinites and the more standard plural form is *juntzanh* (37b). But, *jun* can also be reduplicated as *junjun* to yield a plural indefinite (37c).

(37) a. jun tz'i'
INDEF dog
"a dog"

b. juntzanh tz'i'
INDEF.PL dog
"some dogs"

c. junjun tz'i'
INDEF.PL dog
"some dogs"

A major goal of this dissertation is to explain classifier pronouns (row 2 in Figure (3)). These are classifiers that are used as 3RD person anaphoric pronouns. (38) contains two examples of classifier pronouns: *te'* and *winh*.

(38) a. ix-s-man jun te' onh winh winak.
PFV-A3-buy INDEF CLF avocado CLF man.
"The man bought an avocado."

b. ix-s-lo'-an **te' winh**
PFV-A3-eat.sweet-CON **CLF.PRON CLF.PRON**
"And then he ate it."

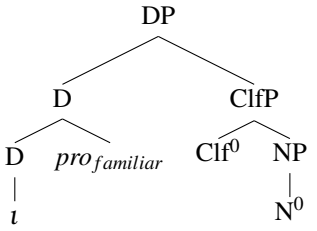
Another goal is to explain how deictic particles fit into this theory of definiteness and indefiniteness. In Chuj, deictics can combine with both definite (39) and indefinite bases (40).

(39) nok' tzi'i' tik
CLF dog DEIX.PROX
"the/this dog" (DDB)

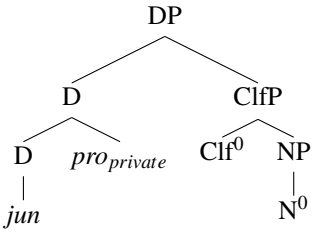
(40) jun (nok') tz'i' tik
INDEF (CLF) dog DEIX.PROX
"this dog" (DIB)

Royer's (2022) analysis of the Chuj data is as follows. First of all, classifiers contribute uniqueness and existence presuppositions. Combining a classifier with a noun results in a *weak definite*. Strong definites are formed compositionally from weak definites and require a deictic particle. When one combines a classifier with an indefinite article and a noun, it results in a *specific* indefinite. Classifier pronouns in Chuj can sometimes be *e*-type pronouns, and in this case, they are strong definite descriptions. Royer proposes that for definite descriptions (41), DP is headed by a null ι morpheme that must combine with a situation pronoun whose value is either known or familiar to the speech act participants. For indefinite descriptions (42), the situation pronoun's value does not need to be known by all discourse participants-the values can be 'private'.

(41) **Weak definites**



(42) **Indefinites**



As for demonstrative configurations, Royer argues that demonstratives are derived compositionally in Chuj and his theory of (in)definiteness can be extended to explain why deictic particles can combine with both definite and indefinite bases, though they have different uses according to what type of base the particle combines with. Deictic particles also need to combine with another functional nominal item to be licensed and they can appear in existential sentences with indefinite quantifiers.

To encode the differences between each definite configuration, recall that classifiers are weak definite determiners and they impose uniqueness and existence conditions non-assertively. Strong definites and classifier pronouns contribute a familiarity or anaphoric presupposition. Royer formalizes this by adding a null index ($\lambda x.x = g(i)$) for classifier pronoun uses.

Figure 4: Definite configurations with semantic output (Royer 2022: 60)

	Configuration	definite type
1.	CLF + NP	weak definite
2.	CLF + NP + DEIX	strong definite
3.	CLF + [λx.x = $g(i)$]	strong definite (= anaphoric pronoun)

Figure (4) is a summary of the three types of definites in Chuj. The weak definite is a bare classifier phrase, and there are two types of strong definites. Strong definites can either contain both a classifier and a deictic particle or be anaphoric pronouns which are composed of a classifier and a null index.

6 The proposed analysis

The two languages (Chuj and Hmong) have several differences. First of all, Hmong allows much fewer DP configurations than Chuj. Part of this is because the Chuj indefinite article can combine with an NP directly, whereas this configuration is ungrammatical in Hmong. Another reason is that weak and strong definites are overtly distinguished in Chuj, but Hmong uses the same surface phrase for both types of definites.

Concerning their semantic differences, the locus of number semantics is different and so is the general nature of the indefinite articles. In Chuj, plurality is encoded in a plural suffix on the indefinite article, but in Hmong, it is encoded in the classifier. This is not an affix added to the classifier, it is an entirely different classifier. As for their indefinites, Chuj indefinite phrases have a specific indefinite reading, but in Hmong, the indefinites are always non-specific.

With these key differences in mind, the following modifications are needed for an analysis of Hmong. First, Hmong demonstratives need to be specified differently to yield the correct meanings for DDBs and DIBs. Since their meanings are tied to the location of the referent, this article adopts a heavily deictic approach to demonstratives (Ahn et al. 2019). Second, for strong definites, I agree that there is an added index to get the anaphoric property, but instead of adding this as a separate node, it is incorporated into one of the null determiner options. This is called ι', following Jenks (2018)'s ι^x. Finally, due to the morphological differences in where number is encoded, this paper dedicates a chunk of the classifier analysis to explaining the lexical semantics of *cov* to show how it generates the appropriate plural readings.

6.1 Syntactic structures of nominal phrases

Bare nouns.

As mentioned in section §3, bare nouns in Hmong are grammatical but only as kinds. Following Chierchia (1998), this is because bare nouns are NPs. The following constituency tests show that bare classifier phrases and pronouns are the same category (43a), and pronouns and indefinite phrases are the same category (43b). These three constituents are DPs. (43c) shows that a pronoun, which is a DP, cannot coordinate with a bare noun, so it must be a different type of constituent. This follows if bare nouns are NPs

(43) a. Kuv pom **tus** **aub** thiab **koj**
1SG see **CLF** **dog** and **2SG**
"I see the dog and you."

b. *Kuv pom **aub** thiab **koj**
1SG see **dog** and **2SG**
"I see (a/the) dog and you."

c. Kuv pom **ib** **cov** **aub** thiab **koj**
1SG see **INDEF** **CLF** **dog** and **2SG**
"I see some dogs and you."

Classifiers.

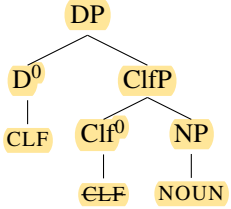
Under this account, classifiers are heads, not specifiers, because they select an NP and undergo head movement to D^0 for a definite reading (Simpson 2008). The following examples (44a) and (44b) show that bare classifier phrases with *tus* and *cov* can both be coordinated with pronouns, meaning that bare classifier phrases are DPs regardless of which classifier is used.

(44) a. Kuv pom **tus** **aub** thiab **koj**
1SG see **CLF** **dog** and **2SG**
"I see the dog and you."

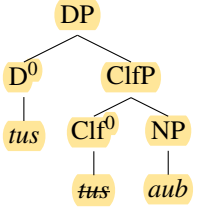
b. Kuv pom **cov** **aub** thiab **koj**
1SG see **CLF** **dog** and **2SG**
"I see all the dogs and you."

At this point, there is enough information established to begin putting the generalizations into structure. So far, we know that bare nouns are NPs and bare classifier phrases are DPs. A template for DP is shown in (45a) and a Hmong bare classifier phrase example is in (45b).

(45) a. **Template**



b. **Hmong example**



Following Simpson 2008, the classifier undergoes head movement[6] to D^0 when there is no indefinite 'one' and this is how the bare classifier structure yields a definite reading. One argument for this head movement is that although bare classifier phrases may naturally be interpreted as being definite. When a numeral precedes the classifier it yields an indefinite interpretation (46a).

(46) CONTEXT: How would you ask for someone to bring you "three chickens," where the three chickens could be any random chickens in the chicken coop?[7]

a. peb tus qaib no
THREE CLF chicken cold
"three cold chickens"

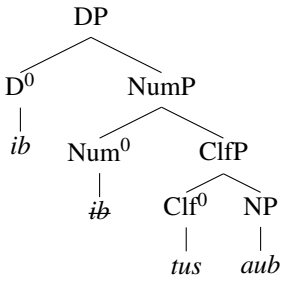
[6]There are no words or morphemes that can occur between D^0 and Clf^0, so I cannot do more standard distributional tests to further support that Clf^0 is a head.

[7]I omit the full question here, since we are only looking at this DP.

Indefinites.

As for indefinite phrases, we posit that the indefinite article also starts out as Num^0 and raises to D^0 to yield an indefinite interpretation. Adding a projection for NumP below D^0 yields the updated structure in (47).

(47) **Indefinite example**



Additional evidence for this is provided in (48a) where it is ungrammatical to say *three chickens* as *three* plus the usual indefinite structure (one CLF chicken). When asked to translate the English phrase "three chickens" the consultant provided the grammatical form in (48a), and when asked if the phrase in (48b) was correct, the consultant confirmed that it does not make sense. Since only one numeral is allowed at a time, *ib* must be generated on the same syntactic head as other numerals like *peb* (three).[8]

(48) a. peb tus qaib
three CLF chicken
"three chickens"

b. *peb ib tus qaib
three one.INDEF CLF chicken
"three chickens"

Demonstratives.

Lastly, demonstratives cannot be used without an accompanying noun (49a). Bare classifier phrases can be coordinated under one demonstrative (49c) and demonstratives can combine with both indefinite DPs and bare classifier phrases (=definite DPs). This data suggests that Dem syntactically selects a DP nominal.

(49) a. *Kuv pom no thiab ko
1SG see DEM.SPKR and DEM.ADDR
"I see this and that (by you)."

b. Kuv pom tzi tus aub nthaws ntawd
1SG see five CLF dog fat DEM
"I see five fat dogs over there."

c. kuv pom tus aub thiab tus qaib ntawd
1SG see CLF dog and CLF chicken DEM
"I see the dog and the chicken over there."

Hmong's underlying structure is all left-headed, but there is an EPP feature on D^0, forcing a smaller XP (which maximally contains Num^0, Clf^0 and the noun) to raise to Spec,DP (Simpson et. al 2015). This strands the demonstrative at the right edge of the phrase. The example repeated in (50) shows that we need both adjectives and demonstratives to be stranded at the right edge, which this analysis generates correctly.

(50) Kuv pom tzi tus aub nthaws ntawd
1SG see five CLF dog fat DEM
"I see five fat dogs over there."

[8] *peb* can either be a 1PL pronoun or the numeral three.

Motivating EPP within the DP.

This implementation of EPP follows Simpson (2008) and Simpson et al. (2011) in claiming that there is an EPP feature that drives movement of a smaller constituent to Spec,DP. The smaller constituent is the NP-Num-Clf unit, which we have established as a DP constituent. When this smaller DP (which I call *d*P) raises to Spec,DP, it yields the surface order seen in Hmong, with demonstratives and other syntactic modifiers being stranded at the end of the entire DP.

Support for this *d*P-movement[9] analysis can be found in several other languages. In Vietnamese, for example, the neutral DP order is [Num Clf NP], but in poetry and in "inventory forms" it can be [NP Num Clf]. Indonesian also has these two orders available (Simpson 2008). Additionally, in Greenberg (1975), and various other works, it is pointed out that linear sequences of noun or NP before numeral (and classifier) are found to occur particularly often in written list or "inventory" forms, as well as when people are involved in situations such as ordering food in a restaurant or buying commodities in a store.

Thus, placing *d*P at the beginning of the nominal phrase can be compared to presentational focus or topicalization in the sentential domain. In both CP and DP domains, nominal elements that are newly presented can be fronted so that they precede their modifiers. For languages such as Vietnamese and Indonesian, DP-internal movement is optional, but in some languages, such as Hmong, all nominal phrases need to have the smaller *d*P in the nominal-phrase initial position. Given these similarities with the sentential domain, it is necessary to state that in the Hmong DP, phrasal movement is driven by an EPP feature on D^0.

Admittedly, the label of this smaller *d*P is somewhat arbitrary. But, this category has motivation in other languages, so I take inspiration from that work (Cinque 2010). Calling it little *d*P enables us to uniformly posit that D^0 has an EPP feature and says nothing about an EPP on little *d*. An EPP feature on only a smaller DP would hardly explain why some DPs in the language have an obligatory EPP and some do not.

That being said, the adjectival arguments for little *d*P in Cinque (2010) do not apply to Hmong. Attributive adjectives in Hmong only surface at the very end of the entire DP, so they do not need different adjective positions available in this language. Again, I mainly employ this projection to make the nominal EPP generalization uniform across the language. But, when there is no DemP projection, I still assume that *d*P raises to Spec,DP, even though there is no surface evidence.[10]

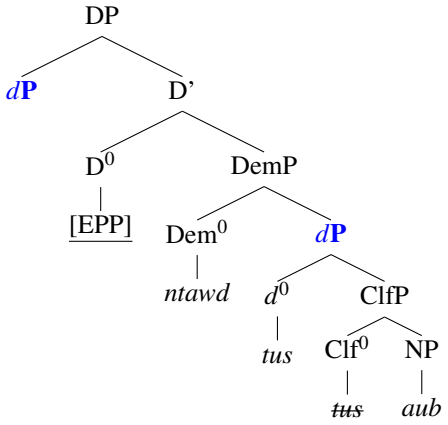
Demonstrative phrases.

What this means for demonstrative phrases is essentially that they are modifiers to the smaller, core *d*P, and when *d*P raises to Spec,DP the modifiers are stranded at the end of the nominal domain. For demonstratives that combine with definite bases (DDB), the syntactic structure is shown in (51a) and (51b) shows the structure for demonstratives with indefinite bases (DIB).

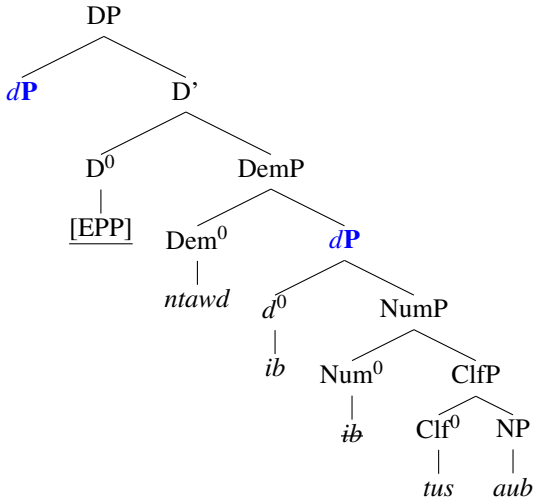
[9] It is DP-movement in our analysis, since we take an explicit stance on this unit's category. In Simpson et. al (2011), this is termed NP-movement.

[10] I assume this for continuity, but acknowledge there is no evidence in Hmong DP configurations without syntactic modifiers.

(51) a. **DDB**



b. **DIB**



Under this account, all three demonstratives are identical in the syntax and they only differ in their LF contributions, which is discussed in more detail in the semantic portion of the analysis presented here.

6.2 Semantics

This semantic discussion begins with lexical definitions for each of the individual words, then we move to bare classifier phrases and indefinite phrases. Our formal discussion ends with an account of the demonstratives and how their semantics are composed with definite and indefinite bases. The final subsection briefly addresses how our syntactic and semantic accounts are compatible with one another at their interface.

6.2.1 Lexical meanings

Classifiers. Both classifiers (*tus* and *cov*) turn a type e kind (bare noun) into a property (of type $\langle e,t\rangle$). This paper follows Chierchia (1998) in assuming that the bare noun is kind-denoting and is typeshifted with an 'up' operator ($\cup$) before it combines with the classifier. To simplify the denotations, this operator is left out of the classifier denotations. Both classifier phrases (those with *tus* or *cov*) need an ι operator to get a unique or maximal reading. In addition, *tus* (52a) must turn a bare noun into an atomic nominal, but *cov* does not.

(52) a. $tus^{g} = \lambda P.\lambda x.[P(x) \wedge AT(x)]$
b. $cov^{g} = \lambda P.\lambda x.P(x)$

The denotation in (52a) works by combining with a type-shifted nominal, and saying that that nominal is atomic. *Cov* is a plural classifier that can refer to atomic or non-atomic entities in the context (52b). This definition combines with a nominal property and says that x is a non-atomic or atomic entity. The reasons for *tus* having 'AT(x)' and *cov* not having it will become clearer when we derive the plural DP configurations in a later section.

Indefinite article. Recall that the indefinite article *ib* always results in a non-specific indefinite reading. In Hmong, I argue the indefinite article (53b) is a choice function (53), rather than an existential quantifier.

(53) **Choice Function** (Winter 1997)
a. A function that applies to sets and arbitrarily selects one of their members
b. $ib^{g} = \lambda P_{\langle e,t\rangle}.f_{cf}(\lambda y.P(y) = 1)$

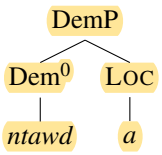
Indefinite *ib* requires a property of type $\langle e,t\rangle$ as its argument. Since bare nouns are type e kinds in Hmong, the classifier acts as a bridge between the noun and what type *ib* selects for, as bare nouns are not type-shifted to $\langle e,t\rangle$ unless they combine with a classifier.

Demonstratives.

Recall that Hmong has three demonstratives: *no, ko* and *ntawd*. Here we do not consider any anaphoric uses, so that we can focus on deictic uses. We also only look at the configurations with *ntawd* since the other two only differ in their presuppositional content (55b)-(55c).

The analysis of Hmong demonstratives is that the demonstrative takes two restrictions: the DP entity and location. It first takes a location argument a and then takes the whole DP (an entity of type e) x as an argument, and adds a presupposition that x is at a (Ahn 2022).

(54)



For the proximal demonstrative *no* (55b), this definition says that a is close to the speaker S and x is at this presupposed location (close to S). For the other proximal demonstrative *ko*, a is close to the addressee A, and x is located at a. The denotations below show the semantics at the Dem^0 level of the structures once they have combined with their location restriction. When we describe the DDB and DIB configurations in a later section, we will explain how the demonstrative combines with the entity restrictor.

(55) **Demonstratives**

a. $ntawd^g = \lambda a.\lambda x : \textsc{Loc}(x,a).x$

b. $no^g = \lambda a.\lambda x : \textsc{Close}(a,S) \wedge \textsc{Loc}(x,a).x$

c. $ko^g = \lambda a.\lambda x : \textsc{Close}(a,A) \wedge \textsc{Loc}(x,a).x$

We use the simple demonstrative definition (55a) for Hmong. Demonstratives like *ntawd* are often accompanied by pointing gestures but do not need to be, and they can be used anaphorically.[11] Additionally, (55a), there is no extra specification of where the location a is (55a). It more vaguely says that x is located at a location a.

Once again, *no* and *ko* have the same structure as *ntawd*, so nothing in (54) visually changes if a different demonstrative is used instead.

6.2.2 Bare classifier phrases

At this point, we address how definite readings are derived from bare classifier phrases. Weak and strong definite descriptions (56) are not distinguished overtly in Hmong, but something should account for both underlying uses of bare ClfPs.

(56) Strong vs. weak definites

a. Strong definites: strong (unique *and familiar*) definites' ι' operator combines with a familiar situation pronoun before combining with the ClfP.

b. Weak definites: weak (unique) definites' ι operator combines with a private situation pronoun before combining with the classifier-noun complex.

Adopting Jenks's approach to weak and strong definites, we distinguish them using two different ι operators (Jenks 2018). The regular ι operator (57a) is used for definites that are not anaphoric and

[11] I am not sure if *no* and *ko* can be used anaphorically or if their use is restricted to deictic uses. Either way, *ntawd* can be used both deictically and anaphorically.

(57b) is used for anaphoric definites. In practice, the only differences between these two operators are the added anaphoric property in (57b) and an explicit parthood relation between the variable in the situation (y) and the variable that refers to the broader set of things that would satisfy the nominal descriptive content (x) (57a).

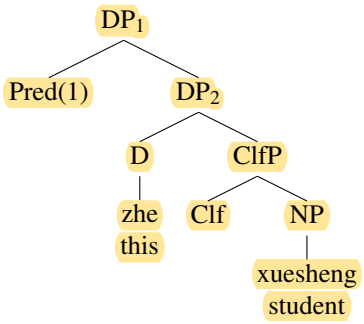
(57) a. $\iota = \lambda P : \exists x[P(x) \wedge \forall y[P(y) \rightarrow y \leq x]].\iota x[P(x) \wedge \forall y[P(y) \rightarrow y \leq x]]$

b. ι' $= \lambda P.\lambda Q : \exists! x[P(x) \wedge Q(x)].\iota x[P(x)]$

In the definition for ι' (which is ι^x in Jenks 2018), Q is the domain restriction, which can be satisfied by an index; in this case, a property. I follow Jenks in assuming $Q(x)$ is only part of the presuppositional component. This is what gives us the *anaphoric* property of strong definites (59d).

To show how the denotation for ι' works, the following is an example taken directly from Jenks (2018), where Jenks derives the following Mandarin structure (58) as an anaphoric definite.

(58)



The full steps for the structure in (58) are shown in (59) below. While the whole derivation of this phrase is shown here, the steps at DP_2 and DP_1 are the relevant ones for understanding how $Q(x)$ is saturated by Pred(1) in the tree structure.

(59) Example of λQ (Jenks 2018)

a. NP $= \lambda x.[student(x)]$

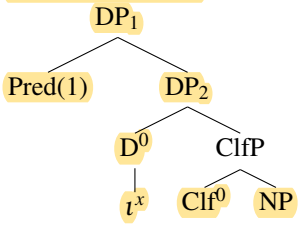
b. ClfP $= \lambda x.[student(x) \wedge AT(x)]$

c. $DP_2 = \lambda Q.\exists! x[student(x) \wedge AT(x) \wedge Q(x)].\iota x[student(x) \wedge AT(x)]$

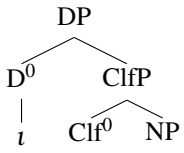
d. $DP_1 = \exists! x[student(x) \wedge AT(x) \wedge \mathbf{x = g(1)}].\iota x[student(x) \wedge AT(x)]$

Similarly in the Hmong analysis, the first thing either ι or ι' combines with is the ClfP. This results in the semantic structures for weak and strong definites, which are shown in (60a) and (60b).

(60) a. **Strong definites**



b. **Weak definites**



The corresponding semantic derivations for the Hmong strong and weak definite constructions with *tus* are shown below.

(61) Semantics of strong definite example

a. $\text{Clf}^0 = \lambda P.\lambda x[P(x) \wedge AT(x)]$

b. $\text{ClfP} = \lambda x[\text{DOG}(x) \wedge AT(x)]$

c. $\text{D}^0 = \lambda P.\lambda Q : \exists!x[P(x) \wedge Q(x)].\iota x[P(x)]$

d. $\text{DP}_2 = \lambda Q : \exists!x[\text{DOG}(x) \wedge AT(x) \wedge Q(x)].\iota x[\text{DOG}(x) \wedge AT(x)]$

e. $\textbf{DP}_1 = \iota x[\text{DOG}(x) \wedge AT(x)]$
defined iff: $\exists!x[\text{DOG}(x) \wedge AT(x) \wedge x = g(1)]$

(62) Semantics of weak definite example

a. $\text{Clf}^0 = \lambda P.\lambda x[P(x) \wedge AT(x)]$

b. $\text{ClfP} = \lambda x[\text{DOG}(x) \wedge AT(x)]$

c. $\text{D}^0 = \lambda P : \exists x[P(x) \wedge \forall y[P(y) \rightarrow y \leq x]].\iota x[P(x) \wedge \forall y[P(y) \rightarrow y \leq x]]$

d. $\textbf{DP} = \iota x[\text{DOG}(x) \wedge AT(x) \wedge \forall y[\text{DOG}(y) \wedge AT(y) \rightarrow y \leq x]]$
defined iff: $\exists x[\text{DOG}(x) \wedge AT(x) \wedge \forall y[\text{DOG}(y) \wedge AT(y) \rightarrow y \leq x]]$

So far, we have shown how the two types of singular definites are composed,[12] but what about plural definites? D^0 is formed similarly, with an ι operator. On the ClfP side of the structure, we encounter differences. As stated, the lexical contribution of the plural *cov* is to generate a group of atomic and non-atomic entities that fit the noun's descriptive content in the context.

First, *cov* combines with a type-shifted bare noun and generates a group of entities that fit the noun's description in the context (63a). Once D^0 combines with ClfP, the ι operator ensures that the largest non-atomic entity in the situation is the resulting meaning. This is shown in (63b) for a context where there are three dogs: Apollo, Mars and Copper.

(63) AUB_C: {Apollo, Mars, Copper}

a. $[_{ClfP}$ cov aub] = AMC, AM, AC, CM, A, M, C

b. $[_{DP}$ cov aub] = AMC

The semantic derivation for the plural definite configuration (with the same structure as in (60b)) is shown below.

(64) Semantics of plural definite example

a. $\text{Clf}^0 = \lambda P.\lambda x.P(x)$

b. $\text{ClfP} = \lambda x.\text{DOG}(x)$

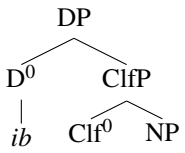
c. $\text{D}^0 = \lambda P : \exists x[P(x) \wedge \forall y[P(y) \rightarrow y \leq x]].\iota x[P(x) \wedge \forall y[P(y) \rightarrow y \leq x]]$

d. $\textbf{DP} = \iota x[\text{DOG}(x) \wedge \forall y[\text{DOG}(y) \rightarrow y \leq x]]$
defined iff: $\exists x[\text{DOG}(x) \wedge \forall y[\text{DOG}(y) \rightarrow y \leq x]]$

6.2.3 Indefinite phrases

Indefinite articles always select an $\langle e,t \rangle$ argument, which is a ClfP under this analysis and cannot be a bare noun of type e. The proposed structure for singular indefinites is shown in (65a).

(65) a. **Indefinite structure**



b. $ib^g = \lambda P_{\langle e,t \rangle}.f_{cf}(\lambda y.P(y) = 1)$

[12] Appendix A lists the DP-level denotations in one place for every configuration this paper derives, so they can be compared side by side if the reader desires to do so.

Recall that the indefinite article *ib* is a choice function that applies to a set and arbitrarily selects one of the members from that set. Its denotation as a choice function is restated in (65b). To show how this works, imagine the same context of three dogs: Apollo, Mars, and Copper. The choice functional *ib tus aub* arbitrarily selects any one of those three dogs in the context. In (66b) each of the options is shown.

(66) AUB_C: {Apollo, Mars, Copper}
  a. $[_{ClfP}$ tus aub] = A, M, C
  b. $[_{DP}$ ib tus aub] = A or M or C

The semantic composition for the singular indefinite structure in (65a) is shown step-by-step below.

(67) Semantics of singular indefinite example
  a. $\text{Clf}^0 = \lambda P.\lambda x.[P(x) \wedge AT(x)]$
  b. $\text{ClfP} = \lambda x.[\text{DOG}(x) \wedge AT(x)]$
  c. $\text{D}^0 = \lambda P_{\langle e,t\rangle}.f_{cf}(\lambda y.P(y) = 1)$
  d. **DP** $= f_{cf}(\lambda y.[\text{DOG}(y) \wedge AT(y)] = 1)$

The structure is the same for all indefinite phrases, but the number interpretation changes when *cov* is the classifier. To show how *cov aub* and *ib cov aub* differ, in the same context of three dogs, the following interpretation options hold. (68a) is the group of atomic and non-atomic dogs in the context that the classifier *cov* can refer to at the ClfP level of the structure.

(68) AUB_C: {Apollo, Mars, Copper}
  a. $[_{ClfP}$ cov aub] = AMC, AM, AC, CM, A, M, C
  b. $[_{DP}$ ib cov aub] = AMC or AM or AC or CM or A or M or C
  c. antipresupposition = {A, M} or {A, C} or {C, M}

Analyzing *ib* as a choice function allows us to see how a random entity is selected from the whole group of entities generated by *cov* classifier phrases. In the case of plural indefinites, the choice function does not select the largest entity, otherwise, ι would be the determiner chosen for the utterance instead of *ib*. The choice function also does not select an atomic entity, otherwise, the classifier *tus* would have been chosen instead of *cov*. This anti-presupposition yields the non-singular and non-maximal entities as the pragmatically sound options when *ib cov aub* is uttered (68c).

The semantic composition for plural indefinites is shown step-by-step below.

(69) Semantics of plural indefinite example
  a. $\text{Clf}^0 = \lambda P.\lambda x.P(x)$
  b. $\text{ClfP} = \lambda x.\text{DOG}(x)$
  c. $\text{D}^0 = \lambda P_{\langle e,t\rangle}.f_{cf}(\lambda y.P(y) = 1)$
  d. $\text{DP} = f_{cf}(\lambda y.\text{DOG}(y) = 1)$

6.2.4 Demonstrative phrases

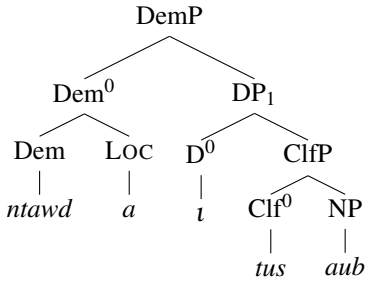
Finally, we revisit the semantic explanation of how demonstratives in Hmong combine with either definite bases (14a) or indefinite bases. As a reminder, the demonstrative in Hmong syntactically selects a DP. This is true for the semantic structure as well: the DP (the demonstrative's entity restriction) is composed exactly the same and then it combines with Dem^0. Dem^0 is assumed to be made up of the demonstrative and a location presupposition.

For DDB configurations, DP is formed from an ι operator and a ClfP consisting of a classifier and NP. It is derived through exactly the same process as described above for weak singular definites (without demonstratives). Dem^0 then combines with the type *e* DP, its entity restriction, and returns another type

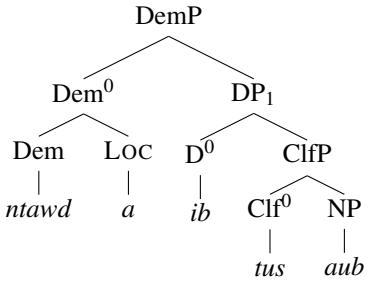
e entity, adding the location restriction to it. In the examples here, the location presupposition is that *x* is at some location *a*.

As for DIB configurations, DP is formed in the same way as previously shown for singular indefinites and then Dem^0 combines with the DP. Again, Dem^0 takes a type *e* DP and returns a type *e* entity with an additional presupposition that *x* is at a location *a*. The semantic structures for both DDBs (70a) and DIBs (70b) are shown here.

(70) a.



b.



The semantic compositions for both DDB and DIB configurations are shown step-by-step below.

(71) Semantics of DDB example

a. $\text{Clf}^0 = \lambda P.\lambda x[P(x) \wedge AT(x)]$

b. $\text{ClfP} = \lambda x[\text{DOG}(x) \wedge AT(x)]$

c. $\text{D}^0 = \lambda P : \exists x[P(x) \wedge \forall y[P(y) \rightarrow y \leq x]].\iota x[P(x) \wedge \forall y[P(y) \rightarrow y \leq x]]$

d. $\text{DP} = \exists x[\text{DOG}(x) \wedge AT(x) \wedge \forall y[\text{DOG}(y) \wedge AT(y) \rightarrow y \leq x]]$
$.\iota x[\text{DOG}(x) \wedge AT(x) \wedge \forall y[\text{DOG}(y) \wedge AT(y) \rightarrow y \leq x]]$

e. $\text{Dem}^0 = \lambda a.\lambda x : \text{Loc}(x,a).x$

f. $\text{DemP} = \exists x[\text{DOG}(x) \wedge AT(x) \wedge \text{Loc}(x,a) \wedge \forall y[\text{DOG}(y) \wedge AT(y) \rightarrow y \leq x]]$
$.\iota x[\text{DOG}(x) \wedge AT(x) \wedge \forall y[\text{DOG}(y) \wedge AT(y) \rightarrow y \leq x]]$

g. $\textbf{DemP} = \iota x[\text{DOG}(x) \wedge AT(x) \wedge \forall y[\text{DOG}(y) \wedge AT(y) \rightarrow y \leq x]]$
defined iff: $\exists x[\text{DOG}(x) \wedge AT(x) \wedge \textbf{Loc(x,a)} \wedge \forall y[\text{DOG}(y) \wedge AT(y) \rightarrow y \leq x]]$

(72) Semantics of DIB example

a. $\text{Clf}^0 = \lambda P.\lambda x[P(x) \wedge AT(x)]$

b. $\text{ClfP} = \lambda x[\text{DOG}(x) \wedge AT(x)]$

c. $\text{D}^0 = \lambda P_{\langle e,t \rangle}.f_{cf}(\lambda y.P(y) = 1)$

d. $\text{DP} = f_{cf}(\lambda y.[\text{DOG}(y) \wedge AT(y)] = 1)$

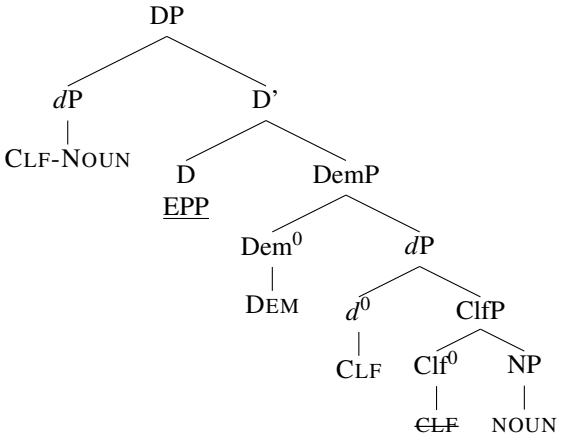
e. $\text{Dem}^0 = \lambda a.\lambda x : \text{Loc}(x,a).x$

f. $\textbf{DemP} = f_{cf}(\lambda y.[\text{DOG}(y) \wedge AT(y)] = 1)$
defined iff: $\text{Loc}(x,a)$

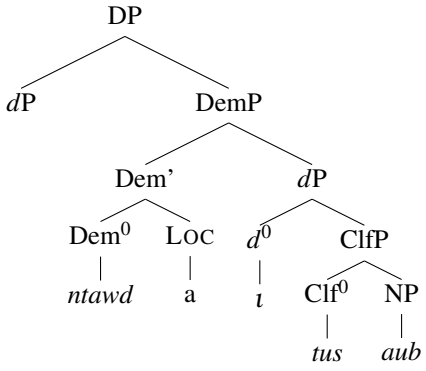
6.3 Syntax-semantics interface

Now that we have described the syntactic and semantic structures and composition, there are a few things left to say about the syntax-semantics interface. If we were to combine every assumption from both sections, we would have the structure in (73b) for a DDB example.

(73) a. **Syntax**



b. **Semantics (preliminary)**



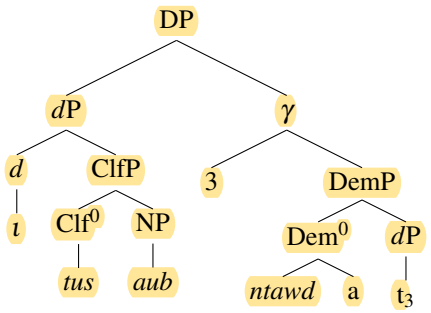
However, for the semantic structure to make sense with the assumed syntactic movement, we need to address three things. First, we need to explain *d*P movement to Spec,DP in the semantic representation. Second, we need to explain how the semantics of head movement to D^0 works. To explain phrasal movement, we need to use the Trace Rule (74a) and Predicate Abstraction (74b) (Heim & Kratzer 1998).

(74) a. **Trace Rule**
If α is a trace, g is a variable assignment and $i \in \text{DOM(g)}$, then $[\alpha_i]^{w,g} = g(i)$

b. **Predicate Abstraction (PA)**
For a branching node $[\gamma\ n\ \beta]$ where n is an index, $[\gamma]^{w,g} = \lambda x.[\beta]^{w,g[n \to x]}$

This is the same way one would account for subject movement in the clausal domain, which was our syntactic parallel when we invoked the EPP. In the structure (75), *d*P leaves behind a trace (t_3 in this case) when it raises to Spec,DP. We need Predicate Abstraction (PA) so that we can assign an index to DemP which contains a trace from the moved element once the structure encounters the full *d*P in Spec,DP.

(75)



Using PA like this unifies the syntactic and semantic movement of *d*P to DP. The semantics with PA at γ is (76a), which is simplified by the Trace Rule as (76b). Notice that the denotation in (76b) is essentially the same as at the Dem^0 level, but the index is added so that x is bound to and corefers with *d*P.

(76) a. $\lambda x_e.\text{LOC}(g[3 \to x](3), a).g[3 \to x](3)$

b. $\lambda x_e.\text{LOC}(x, a).x$

Previously we stated that in definite phrases, Clf^0 raises to D^0 to yield a definite interpretation and Num^0 raises to D^0 to yield an indefinite interpretation. Presumably, this is so the closest head X^0 can host definite or indefinite features. In a Y-model grammar (Chomsky 1973), movement happens before we get to the surface structure, which feeds into the semantics. So, while our phrasal movement analysis worked in the semantics, explaining how things move to D^0 for definiteness features in the semantics would be more complicated. This remains a topic to be expanded on in future work.

7 Discussion & Conclusion

To summarize, this paper addressed several puzzles within the semantics of definite and indefinite phrases in Hmong. The first puzzle was the 'plural' or 'maximal group' classifier, which has different interpretations in definite and indefinite contexts. This puzzle surrounding *cov* goes hand in hand with the second one; the indefinite word *ib*. To account for the data, *cov* must return both atomic and *non-atomic* entities, and *ib* must be analyzed as a choice function. The final puzzle is the lexical semantics of Hmong demonstratives and how they combine with both definite and indefinite bases. Some explicit implications of this project are as follows.

Classifiers. We established that the only defining characteristic of a classifier in Hmong is to convert a kind into a property. The difference between the singular and plural classifiers is that plural classifiers lack an 'AT(x)' in their denotation, while singular classifiers require it.

Indefinites. When *ib* combines with the ClfP, it picks a non-atomic and non-maximal entity from the full group of atomic and non-atomic entities that *cov aub* can refer to. Typically, *cov aub* is translated as 'some dogs' with a non-specific interpretation. Analyzing *ib* as a choice function allows us to see how this smaller entity is selected from the entire group of entities denoted by bare *cov* classifier phrases.

Demonstratives. Demonstratives in Hmong, when used deictically, are syntactic modifiers and primarily contribute spatial information about the nominal referent. Since they are adjuncts, along with other modifiers, they are always stranded at the end of the nominal domain when the lower *d*P raises to the specifier of DP. There is an EPP feature on D^0 that drives this movement.

Interface. In explaining how the syntactic representation feeds the semantic representation, we established that Predicate Abstraction and the Traces Rule (Heim & Kratzer 1998) are needed to derive phrasal movement within the nominal domain. This mirrors EPP-driven movement at the clausal level, which is also derived via these two rules. If other domains are argued to have an EPP feature, this account could be tested for the semantics of those domains as well. While this paper addressed *d*P movement within a DP, head movement to d^0 could be expanded on later in future work.

Empirical. Finally, there is minimal semantic work on Hmong and any previous discussion of classifiers is brief, descriptive, and focused on classifier choice (Bisang 1993). The present analysis is unique in that it explains (in)definiteness and number facts in all the possible DP configurations of the classifiers *tus* and *cov*, the indefinite article *ib*, and demonstratives *no, ko* and *ntawd*.

Appendix A Recap of formal definitions

All individual lexical items.

- $tus^{\,g} = \lambda P.\lambda x.[P(x) \wedge AT(x)]$
- $cov^{\,g} = \lambda P.\lambda x.P(x)$
- $ib^{g} = \lambda P_{\langle e,t\rangle}.f_{cf}(\lambda y.P(y) = 1)$
- $no^{g} = no^{g} = \lambda a.\lambda x : \textsc{Close}(a,S) \wedge \textsc{Loc}(x,a).x$
- $ko^{g} = \lambda a.\lambda x : \textsc{Close}(a,A) \wedge \textsc{Loc}(x,a).x$
- $ntawd^{g} = \lambda a.\lambda x : \textsc{Loc}(x,a).x$
- $\iota = \lambda P : \exists x[P(x) \wedge \forall y[P(y) \rightarrow y \leq x]].\iota x[P(x) \wedge \forall y[P(y) \rightarrow y \leq x]]$
- ι' $= \lambda P.\lambda Q : \exists!x[P(x) \wedge Q(x)].\iota x[P(x)]$

Definite descriptions.

- **strong singular definite**

  $= \iota x[\textsc{dog}(x) \wedge AT(x)]$

  <u>defined iff:</u> $\exists!x[\textsc{dog}(x) \wedge AT(x) \wedge x = g(1)]$

- **weak singular definite**

  $=\iota x[\textsc{dog}(x) \wedge AT(x) \wedge \forall y[\textsc{dog}(y) \wedge AT(y) \rightarrow y \leq x]]$

  <u>defined iff:</u> $\exists x[\textsc{dog}(x) \wedge AT(x) \wedge \forall y[\textsc{dog}(y) \wedge AT(y) \rightarrow y \leq x]]$

- **plural definite**

  $=\iota x[\textsc{dog}(x) \wedge \forall y[\textsc{dog}(y) \rightarrow y \leq x]]$

  <u>defined iff:</u> $\exists x[\textsc{dog}(x) \wedge \forall y[\textsc{dog}(y) \rightarrow y \leq x]]$

Indefinite descriptions.

- **singular indefinite**

  $= f_{cf}(\lambda y.\textsc{dog}(y) \wedge AT(y) = 1)$

- **plural indefinite**

  $= f_{cf}(\lambda y.\textsc{dog}(y) = 1)$

Demonstrative phrases.

- **DDB**

  $= \iota x[\textsc{dog}(x) \wedge AT(x) \wedge \forall y[\textsc{dog}(y) \wedge AT(y) \rightarrow y \leq x]]$

  <u>defined iff:</u> $\exists x[\textsc{dog}(x) \wedge AT(x) \wedge \mathbf{Loc(x,a)} \wedge \forall y[\textsc{dog}(y) \wedge AT(y) \rightarrow y \leq x]]$

- **DIB**

  $= f_{cf}(\lambda y.[\textsc{dog}(y) \wedge AT(y)] = 1)$

  <u>defined iff:</u> $\textsc{loc}(x,a)$

References

[1] Ahn, Dorothy. 2022. Indirectly direct: An account of demonstratives and pointing. *Linguistics and Philosophy* 45: 1345-1393.

[2] Ahn & van der Wal. 2019. What does that Lugwere demonstrative refer to? A semantic analysis of proximity and exteriority. *Studies in African Linguistics 48.*

[3] Bisang, W. 1993 . 'Classifiers, quantifiers and class nouns in Hmong', *Studies in Language* 17 : 1–51.

[4] Carlson, G. N. (1977). Reference to kinds in English. Doctoral dissertation, University of Massachusetts at Amherst.

[5] Cheng, L., & Sybesma, R.. 1999. Bare and not-so-bare nouns and the structure of np. *Linguistic inquiry* 30:509–542.

[6] Chierchia, Gennaro. 1997. Partitives, Reference to Kinds and Semantic Variation. *SALT VII,* 73-98. Cornell University. Ithaca, NY.

[7] Chierchia, Gennaro. 1998. Reference to kinds across language. *Natural Language Semantics* 6:339–405.

[8] Chierchia, Gennaro. 2010. Mass nouns, vagueness and semantic variation. *Language and Vagueness*, Vol. 174, No. 1, pp. 99-149.

«««< Updated upstream «««< Updated upstream

[9] Cinque, G. 2010. *The Syntax of Adjectives.* Cambridge MA: The MIT Press. =======

[10] Cinque, G. 2010. *The Syntax of Adjectives.* Cambridge MA: The MIT Press »»»> Stashed changes =======

[11] Cinque, G. 2010. *The Syntax of Adjectives.* Cambridge MA: The MIT Press »»»> Stashed changes

[12] Chomsky, N. 1973. "Conditions on transformations. In S. Anderson & P. Kiparsky (eds.), *A Festschrift for Morris Halle* (pp. 232-286). Holt, Rinehart and Winston.

[13] Collins, John. 2025. "Indefinites: Scope and Context," in Ernie Lepore, and Una Stojnić (eds), *The Oxford Handbook of Contemporary Philosophy of Language,* Oxford Handbooks (2025; online edn, Oxford Academic, 23 Jan. 2025), https://doi.org/10.1093/oxfordhb/9780192856852.013.32, accessed 10 Apr. 2025.

[14] Chung, S. 2000. On Reference to Kinds in Indonesian. *Natural Language Semantics* 8: 157-200.

[15] Daley, Karen Ann. 1998. Vietnamese classifiers in narrative texts. (Summer Institute of Linguistics and The University of Texas at Arlington Publications in Linguistics, 125.) Arlington: SIL. xii+209pp.

[16] Dayal, V. (2004), 'Number marking and (in)definiteness in kind terms'. *Linguistics and Philosophy* 27: 393–450.

[17] Dayal, V. 2012. Bangla classifiers: Mediating between kinds and objects. *Rivista di Linguistica* 24:195–226.

[18] Elbourne, Paul D. 2005. On the acquisition of Principle B. *Linguistic Inquiry* 36:333–366.

[19] Garellek, Marc and Esposito, Christina M.. Phonetics of White Hmong Vowel and Tonal Contrasts, 22 June 2021, pages.ucsd.edu/ mgarellek/files/SupplementaryFiles$_H$mong.html.

[20] Greenberg 1975. "Dynamic Aspects of Word Order in the Numeral Classifier," in C. Li (ed.) *Word Order and Word Order Change.* University of Texas Press, Austin, 27–46.

[21] Heim, Irene, and Angelika Kratzer. 1998. Semantics in generative grammar. *Oxford: Blackwell Publishers.*

[22] Heim, Irene. 1983. The semantics of definite and indefinite noun phrases. Doctoral dissertation, University of Massachusetts, Amherst, MA.

[23] Heimbach. 1979. White Hmong - English Dictionary. Southeast Asia Program, Cornell University, Ithaca NY.

[24] Jiang, Li Julie. 2017. Definiteness in Nuosu Yi and the theory of argument formation. *L&P*.

[25] Jenks, Peter. 2018. Articulated definiteness without articles. *Linguistic Inquiry* 49:501–536.

[26] Kaplan, David. 1969. Quantifying in. In Davidson, D. and Hintikka, J. (eds.) *Words and Objections*.

[27] Kratzer, Angelika. 2021. Situations in Natural Language Semantics. In *Stanford Encyclopedia of Philosophy*. Stanford, CA: CSLI Publications.

[28] Lee, C. 1992. Definite/specific and case marking in Korean. *Theoretical Issues in Korean Linguistics* 325–341.

[29] Mottin, Jean. 1978. Élé ments de grammaire Hmong Blanc. Bangkok: Don Bosco Press.

[30] Nowak, Ethan. 2019. Complex demonstratives, hidden arguments, and presupposition. *Synthese* 198:2865–2900.

[31] O'Gorman, Tim. 2011. Two kinds of definiteness in Lakhota. Ms., University of Colorado at Boulder.

[32] Pozzo, Lena. 2022. Definiteness and Indefiniteness. A Comparative Perspective on Finnish and Italian. *Lea* 11: pp. 355-372.

[33] Reinhart, T. 2006. *Interface Strategies: Reference-set Computation.* Cambridge, MA: MIT Press.

[34] Reinhart, T. 1997. Quantifier scope: how labor is divided between QR and choice functions. *Linguistics and Philosophy* 20:335-397.

[35] Riddle, E. 1989. Serial verbs and propositions in White Hmong. *Linguistics of the Tibeto-Burman Area,* 12(2), 1-13. https://dx.doi.org/10.32655/LTBA.12.2.01

[36] Roberts, Craige. 2002. Demonstratives as definites. In *Information sharing: Reference and presupposition in languages generation and interpretation,* eds. K. van Deemter and R. Kibble, 89–106. CSLI Publications.

[37] Roberts, Craige. 2003. Uniqueness in definite noun phrases. *Linguistics and Philosophy* 26:287–350.

[38] Royer, Justin. 2022. Elements of (in)definiteness and binding: A Mayan perspective. Dissertation submitted to McGill University.

[39] Russell, B. 1905. On denoting. *Mind* 479–493.

[40] Schwarz, Florian. 2009. Two types of definites in natural language. Ph.D. dissertation, University of Massachusetts.

[41] Schwarzschild. 1996. *Pluralities,* Springer.

[42] Simpson. 2008. "Classifiers and DP Structure in South Asia" in The oxford handbook of comparative syntax. Cinque, G., & Kayne, R. S. (Eds.). Oxford University Press, Incorporated. Created from rutgers-ebooks on 2025-02-17 20:33:03.

[43] Simpson et. al. 2011. Bare classifiers and definiteness. *Studies in Language* 35:1 (2011), 168–193. doi 10.1075/sl.35.1.10sim

[44] Stalnaker, Robert. 1977. Assertion. *Syntax and Semantics* 9: *Pragmatics*, 315–332.

[45] Strawson, P. 1950. On referring. *Mind* 59:320–344.

[46] Szabó, Zoltan Gendler. 2000. Descriptions and uniqueness. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 101:29–57.

«««< Updated upstream ««««< Updated upstream

[47] Wilkins, D. 1999. The 1999 demonstrative questionnaire: “This” and “that” in comparative perspective. In D. Wilkins (Ed.), *Manual for the 1999 Field Season,* pp. 1-24. Nijmegen: Max Planck Institute for Psycholinguistics. doi:10.17617/2.2573775. =======

[48] Wilkins, D. 1999. The 1999 demonstrative questionnaire: “This” and “that” in comparative perspective. In D. Wilkins (Ed.), *Manual for the 1999 Field Season,* pp. 1-24. Nijmegen: Max Planck Institute for Psycholinguistics. doi:10.17617/2.2573775. »»»> Stashed changes =======

[49] Wilkins, D. 1999. The 1999 demonstrative questionnaire: “This” and “that” in comparative perspective. In D. Wilkins (Ed.), *Manual for the 1999 Field Season,* pp. 1-24. Nijmegen: Max Planck Institute for Psycholinguistics. doi:10.17617/2.2573775. »»»> Stashed changes

[50] Winter, Y. 1997. Choice Functions and the Scopal Semantics of Indefinites. *Linguistics and Philosophy* 20, 399–467. https://doi.org/10.1023/A:1005354323136

[51] Wolter, Lynsey Kay. 2006. That’s that: The semantics and pragmatics of demonstrative noun phrases. Ph.D. dissertation, University of California Santa Cruz, Santa Cruz, CA.