DEFINITENESS AND QUANTIFICATION: EVIDENCE FROM SHAN

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Mary Rose Moroney

August 2021

DEFINITENESS AND QUANTIFICATION: EVIDENCE FROM SHAN

Mary Rose Moroney, Ph.D.

Cornell University 2021

This dissertation argues for a syntax and semantics of a range of nominal expressions, expanding cross-linguistic theories of nominal reference. Looking at bare nouns, classifier constructions, and several types of relative clauses, I develop an account of definiteness and quantification that (i) connects the mass/count distinction to atomicity—providing a unified approach to count and measure quantification, (ii) argues that classifier morphology is required by cumulative nouns, (iii) proposes a typology of definiteness marking based on determiner inventory, and (iv) relates the distribution of some relative clauses and if-clauses to information structure. Data from Shan, a Southwestern Tai language of Myanmar, is compared with data from several other languages, including English.

In Chapter 2, I argue that bare nouns in Shan, Dënesųłiné, and English share several features in their distribution and interpretation, which arise from the absence of an overt determiner in these constructions. Building on observations by Chierchia (1998), Dayal (2004), and Deal (2017), I argue that in English only mass nouns (and fake mass nouns) have the semantic property of cumulativity, whereas in Shan all nouns are cumulative. Still, Shan count nouns are distinct from mass nouns in having identifiable atomic parts. I present a semantic analysis of bare noun interpretations, where nouns have an $\langle s, \langle e, t\rangle\rangle$-type interpretation. The definite, indefinite, kind, and generic interpretations are derived through type-shifting and generic/existential closure operations.

The property of cumulativity is connected to the use of classifiers that specify the unit of the noun they combine with. In Chapter 3, I propose that this property requires that a classifier be used when Shan bare nouns or English mass nouns combine with a numeral. With the proposed classifier semantics and syntax, I present a unified semantics of count and measure expressions, where I distinguish the two types of expressions based on whether the noun does (count) or does

not (mass) have identifiable atomic parts.

In Chapter 4, I present evidence that many languages, including Shan, Serbian, and Kannada, can express anaphoric definiteness with a bare noun due to the lack of an overt definite determiner in these languages. Based on this evidence, I give a revised typology of definiteness marking, assuming a split between unique and anaphoric definiteness (Schwarz 2009; Jenks 2018). I further argue that the choice between more/less overtly marked expressions of definiteness is motivated by the need to uniquely specify the intended referent, as suggested by Ahn (2019).

Finally in Chapter 5, I argue that left-located constructions in Shan serve to specify the situation argument by which a clause is interpreted. This is relevant for internally headed relative clauses, derived from free relatives, which only appear in a left-located position. Additionally, there is a particle *kɔ* which is involved in universal quantification over situation arguments. I demonstrate how nouns, classifiers, and expressions of definiteness apply to analyzing donkey anaphora in Shan, using the semantics developed for each type of expression.

My conclusions are presented in Chapter 6. The data presented in this dissertation comes mostly from my consultant work with speakers of Shan, a Southwestern Tai language of Myanmar. Shan is an understudied language, which is under-documented in academic contexts.

BIOGRAPHICAL SKETCH

Mary Rose Moroney grew up in Baton Rouge, Louisiana. She graduated from Princeton University with a minor in Linguistics and a major in Classics. She is graduating from Cornell University with a Ph.D. in Linguistics and a graduate minor in Cognitive Science, with a dissertation committee consisting of Sarah Murray and Miloje Despić (co-chairs), John Whitman, and Keir Moulton (University of Toronto).

This dissertation is dedicated to my parents, James and Patricia Moroney.

ACKNOWLEDGEMENTS

There are many who I would like to thank for their part in supporting me on the way towards completing this dissertation. I fear my words will not be adequate to the task of expressing my appreciation.

I could have written none of this dissertation without my Shan language consultants and other members of the Shan community who shared their language with me. The first Shan speaker I met was Aye Twei Soe, an Ithaca resident, who was the language consultant for a field methods class. Working with her sparked my interest in learning about Shan. Ying Zarm Hom, Sai Noom Hseng, Mwe Hkur, Zai Aorng Kham, Khu Sai Leong Kham, Khu Mai Hong, and Aacaan Ke taught me a great deal about Shan language and culture. They were all extremely generous with their time and resources. I am very glad to have gotten the opportunity to meet and work with so many wonderful people and hope to continue to do so going forward.

I have been lucky to have had great professors and advisors at Cornell. I have learned a great deal from my committee members Sarah Murray, Miloje Despić, John Whitman, and Keir Moulton. My co-chairs, Sarah Murray and Miloje Despić, have supported me at every point during my time at grad school. John Whitman is an expert on many topics and many languages, and I benefited from his knowledge and attention to detail. Keir Moulton provided excellent and much-needed comments on my research and writing. Without the help of all of my committee, this dissertation would not exist.

Funding from the Graduate School, the Southeast Asia Program, and an Engaged Grad Student Grant through the Office of Engagement Initiatives at Cornell University made this research project possible. Many thanks to Holly Boulia, Jenny Tindall and Gretchen Ryan for helping me navigate through administrative mazes associated with all of these.

I am very grateful to the members of the Semantics group and SynCirc at Cornell who gave me feedback on my research at various stages of development. Thanks also to the Cornell Southeast Asia program which was a great place for those researching aspects of Southeast Asia to meet and share ideas.

Much of the data in the dissertation was collected in Thailand, and I have several people to thank for their help in things related to language and logistics. Aacaan Ngampit Jagacinski taught me Thai, was an important resource for research, and took me on adventures around Chiang Mai. Oradi Inkhong connected me with the Shan community in Thailand. Sireemas Maspong has helped me with many translations and questions about Thai.

I could not have made it through grad school without the support of my colleagues and friends, particularly Shohini Bhattasali, Joy Das, Naomi Enzinna, Ryan Hearn, Andrea Hummel, Carol-Rose Little, and Brynhildur Stefánsdottír.

Finally, thanks to my family, especially my parents, for their love and support.

TABLE OF CONTENTS

Biographical Sketch . . . iii
Dedication . . . iv
Acknowledgements . . . v
Table of Contents . . . vii
List of Tables . . . x
List of Figures . . . xi
List of Abbreviations . . . xii

1 Introduction **1**
1.1 Goals and guiding questions . . . 1
1.2 Fieldwork . . . 1
1.3 Background and theoretical assumptions . . . 2
1.4 Overview . . . 4

2 Bare nouns cross-linguistically **9**
2.1 Introduction . . . 10
2.2 Bare noun interpretations in Shan . . . 11
2.2.1 Indefinite interpretation . . . 11
2.2.2 Definite interpretation . . . 15
2.2.3 Kinds and generics . . . 16
2.2.4 Shan mass nouns . . . 19
2.2.5 A comparison with English . . . 21
2.2.6 Summary: Bare noun interpretations . . . 26
2.3 Bare nouns: Predicates or kinds? . . . 27
2.3.1 The mass-plural-kind connection . . . 28
2.3.2 Nominal mapping parameter . . . 33
2.3.3 Type-shifting analysis basics . . . 35
2.3.4 Proposed semantics . . . 37
2.3.5 Sub-kinds . . . 43
2.3.6 Summary . . . 49
2.4 Demonstrative . . . 49
2.4.1 Shan: N DEM . . . 49
2.4.2 English: N DET . . . 56
2.4.3 The role of the demonstrative . . . 62
2.5 Chapter summary: Nouns . . . 63

3 Classifiers **65**
3.1 Introduction . . . 66
3.2 Shan classifiers . . . 67
3.2.1 Individual classifiers . . . 68
3.2.2 Measure terms . . . 74
3.2.3 Group terms . . . 75
3.2.4 Containers . . . 78

3.2.5 Summary of classifiers . . . 80
3.3 Structure of numeral-classifier expressions . . . 81
3.3.1 Repeater classifiers . . . 81
3.3.2 Quantification . . . 92
3.3.3 Structure . . . 96
3.3.4 Summary . . . 107
3.4 Classifier Semantics . . . 108
3.4.1 Classifiers and collectives . . . 111
3.4.2 A semantic analysis of Shan classifiers . . . 116
3.4.3 Clf Adj . . . 119
3.4.4 Properties of the plural morpheme . . . 123
3.4.5 Conclusion . . . 134
3.5 Count and measure: A unified analysis . . . 135
3.5.1 Unified analysis . . . 135
3.5.2 Count versus measure . . . 139
3.5.3 Mass nouns and container classifiers . . . 144
3.5.4 Nez Perce . . . 146
3.5.5 Mandarin classifiers . . . 148
3.5.6 Analysis . . . 154
3.6 Chapter summary: Classifiers . . . 157

4 Definiteness **162**
4.1 Introduction . . . 163
4.2 Two kinds of definiteness: Unique and anaphoric . . . 165
4.2.1 Background . . . 165
4.2.2 Uniqueness versus familiarity/anaphoricity . . . 167
4.2.3 Donkey anaphora . . . 170
4.3 Two kinds of definiteness with Shan bare nouns . . . 171
4.3.1 Background on Shan . . . 171
4.3.2 Uniqueness versus familiarity/anaphoricity in Shan . . . 173
4.3.3 Donkey anaphora . . . 176
4.4 A revised typology of definiteness . . . 179
4.4.1 The case of Serbian . . . 179
4.4.2 The case of Kannada . . . 181
4.4.3 Filling in the typology . . . 182
4.4.4 Summary: A revised typology . . . 187
4.5 Analysis . . . 188
4.5.1 Type-shifting . . . 188
4.5.2 Two types of definiteness . . . 191
4.5.3 The full typology . . . 194
4.5.4 A pragmatic proposal . . . 198
4.5.5 Summary . . . 202
4.6 Discussion . . . 203
4.6.1 Contrast . . . 204
4.6.2 Sentential prominence . . . 204

4.6.3 Ambiguity . . . 205
4.7 Chapter summary: Definiteness marking . . . 208

5 Left-located constructions **210**
5.1 Introduction . . . 211
5.1.1 Left-located (LL) constituents . . . 212
5.1.2 Shan IHRCs . . . 216
5.1.3 Thai . . . 221
5.2 Syntax of left-located constructions . . . 223
5.2.1 Position of LL constituents . . . 225
5.2.2 Islands . . . 229
5.2.3 Topic drop . . . 233
5.2.4 Binding . . . 237
5.2.5 Summary . . . 240
5.3 Semantic analysis . . . 244
5.3.1 IHRC construction . . . 246
5.4 The particle *kɔ* . . . 250
5.4.1 Particle kɔ: Semantics . . . 254
5.4.2 Donkey anaphora . . . 257
5.5 Chapter summary: Left-located constructions . . . 260

6 Conclusion **262**

LIST OF TABLES

1.1 Shan numerals . . . 3
1.2 Summary of ontology . . . 4

2.1 Mendia & Filip (inprep): (5) . . . 17
2.2 Comparison between Shan and Dënesųłiné . . . 20
2.3 Comparison between Shan and English bare nouns . . . 26
2.4 Properties of three types of English nouns Deal 2017: (15) . . . 29
2.5 Shan distribution . . . 32
2.6 English bare noun, N DET, and N DEM expressions . . . 61
2.7 Shan bare noun and N DEM (*nâj*) expressions . . . 61

3.1 Based on Chao's (1968) classification of nominal measures as cited by Li (2011) . . . 67
3.2 Basic numeral classifiers in Shan . . . 72
3.3 Basic measure terms in Shan . . . 74
3.4 Basic group terms in Shan . . . 76
3.5 Basic container terms in Shan . . . 78
3.6 Summary of classifier distribution . . . 81
3.7 N Clf Dem interpretation . . . 111
3.8 N Clf Dem interpretation . . . 119
3.9 Classifies, the plural morpheme, and group terms . . . 126
3.10 N Clf Dem interpretation . . . 132

4.1 Typology of definiteness marking (adapted from Jenks 2018) . . . 164
4.2 Types of definiteness described by Schwarz (2009), citing Hawkins (1978) . . . 167
4.3 Expressions of definiteness in German, Thai, and Mandarin . . . 171
4.4 Expressions of definiteness in German, Thai, Mandarin, and Shan . . . 178
4.5 Revised typology of definiteness marking . . . 187
4.6 Languages without definite articles: Bare Nouns (Despić 2019: 28) . . . 191
4.7 Definite bare noun interpretations, both marked, adapted from Despić 2019 . . . 195
4.8 Definite bare noun interpretations, one marked, adapted from Despić 2019 . . . 197
4.9 Definite bare noun interpretations, unmarked adapted from Despić 2019 . . . 198
4.10 Anaphoric and kind interpretations, potential ambiguity in Shan . . . 208

5.1 Summary of syntactic features of LL constructions in Shan . . . 241
5.2 Features related to topic prominence Paul & Whitman 2017 Table 1 . . . 241
5.3 Features related to topic prominence for Shan Paul & Whitman 2017 Table 1 . . . 243

LIST OF FIGURES

3.1 Complete atomic join semi-lattice 109
3.2 Group and parts of groups 117

LIST OF ABBREVIATIONS

1: first person
2: second person
3: third person
A: “Set A” person marking (ergative, possessive)
ACC: accusative
ACH: achievement
ANML: animal
AR: areal (agreement with an "areal" argument)
CAS: casual
CLF: classifier
COMP: complementizer
CONJ: conjunction
CONTR: contrastive
COP: copula
DAT: dative
DECL: declarative
DEM: demonstrative
DET: determiner
DETR: detrimental
DIM: diminutive
DISC: discourse particle
DISTR: distributive
EP: epenthetic insertion
ERG: ergative
EVID: evidential
EXT: existential particle
FEM: feminine
FUT: future
GEN: genitive
GN: generic classifier
HAB: habitual
HON: honorific
HUM: human
HUMB: humble
IMPF: imperfective
INDEF: indefinite
IPFV: imperfective aspect
IRR: irrealis
ITER: iterative
ITV: intransitive verb suffix
LOC: locative
MASC: masculine
MOD: modification marker

NEG: negation
NM: nominalizer
NOM: nominative
NUM.CL: numeral classifier
PASS: passive
PERF: perfective viewpoint aspect
PL: plural
PFV: perfective
POLAR.Q: polar question
POSS: possessive
PNS: possessed noun suffix
PROG: progressive
PRS: present
PRT: particle
PST: past
Q: question
REL: relative pronoun
REM.PAST: remote past tense
REP: reportative
RND: round
S: subject
SG: singular
SFP: sentence-final particle
STAT: stative
SUBJ: subject
TOP: topic

CHAPTER 1

INTRODUCTION

1.1 Goals and guiding questions

The main goals of this dissertation are to build up a syntax and semantics for nominals and expand the cross-linguistic theory related to different aspects of nominal reference. Data from Shan, a Southwestern Tai language of Myanmar, is compared with data from several other languages, including English. The assumptions that we make about reference affect many areas of syntax and semantics, since nominal interpretation and co-reference are used in a variety of tests for other theoretical issues. This body of work builds up the interpretation of Shan nominals starting from the basic noun interpretation. Even at such a basic foundation of theory—the structure and interpretation of a noun—data from Shan can give us insight into what is possible for nouns cross-linguistically.

1.2 Fieldwork

The data for this dissertation comes from fieldwork with Shan. Shan is a Southwestern Tai language in the Tai-Kadai family. The language is spoken in Myanmar and surrounding countries by approximately 4.68 million speakers (Eberhard et al. 2021). I collected data working with a total of five speakers of Shan, but all of the data from this dissertation is taken from examples by two speakers of the Southern Shan variety. Both speakers are from Southern Shan State, Myanmar. One comes from Keng Tawng in Shan State. She has lived in Thailand for over 10 years, and speaks Shan, Thai, and English. The other speaker is from Langkho who has lived in the U.S. in Jacksonville, Florida for approximately 9 years. He speaks Shan, Thai, Burmese, and English. Data was collected either in Thailand or by working remotely. The elicitation methods used in-

clude translation of stories based on storyboards and felicity judgments on grammatical sentences in specific contexts, following techniques described in Bochnak & Matthewson 2015. Elicitation sessions were conducted in English, with Thai used to clarify vocabulary.

1.3 Background and theoretical assumptions

This dissertation spans the literature of several topics, so it will be useful to define the terms I am using. Shan is an analytic, rigid-SVO language, as shown in (1). This means that the typical word order is SUBJECT-VERB-OBJECT. Given that it is an analytic language, there is little morphology on the nouns or verbs that marks grammatical gender, tense, or number. For example, in existential contexts, such as (2), the bare noun does not include morphology marking plurality, suggesting Shan is a number neutral language. BARE NOUNS are nominal expressions that include just the noun and no determiners, numerals, adjectives, or classifiers.

(1) háw hăn măa
1 see dog
'I see a dog/dogs.'

(2) tinaj mí măa
here have dog
'Here there is a dog/are dogs.'

DETERMINERS are morphemes that appear in a nominal expression. In English the determiners are *a* and *the*, for examples. In some contexts, quantifiers like *every* are also classified as determiners. Shan does not have an overt expression equivalent to *the* and *a* in English, but it does have demonstratives *nâj* 'this' and *nân* 'that'. The classification of these lexical items is a matter of some theoretical interest because it bears on the question of whether a languages has noun phrases (NPs) or determiner phrases (DPs) or some combination of the two.

NUMERALS are expressions like *three* that express amounts or numbers. The numerals in Shan from one to twenty can be seen in Table 1.1.

CLASSIFIERS are a varied range of 'noun categorization devices' found cross-linguistically

Table 1.1: Shan numerals

1	nɯŋ	11	síp ʔɛ́t
2	sɔ̌ŋ	12	síp sɔ̌ŋ
3	sǎam	13	síp sǎam
4	sì	14	síp sì
5	haa	15	síp haa
6	hók	16	síp hók
7	tsét	17	síp tsét
8	pɛ̀t	18	síp pɛ̀t
9	kaw	19	síp kaw
10	síp	20	sáaw

(Aikhenvald 2000: 1). This dissertation focuses on a sub-type of classifier called 'numeral classifiers', which appears in numerical and other quantifying expressions. A definition for classifiers is given in (3).

(3) N is a classifier if [(Num) N NP] denotes countable sums of NP-stuff/entities.

(Rothstein 2017: (3), p. 197)

Definiteness is reference to a specific entity. This is often split into *weak* and *strong* definiteness (Löbner 1985; Schwarz 2009). Weak definiteness is uniqueness in a context, and strong definiteness is familiarity or anaphoric reference—that is, referring to an individual that has previously been mentioned in a discourse. In English, definiteness is associated with the determiner *the* and the demonstratives *this* and *that*. Indefiniteness is reference to a non-specific entity or the introduction of an individual into the discourse context. This meaning is often associated with the English determiners *a* or *some*.

In building up the syntax and semantics of nominal expressions, I make several assumptions:

First, I assume that the nominal syntax and semantics can be built up compositionally. This means that it is useful to consider the most basic noun interpretation first and build the syntax and semantics morpheme-by-morpheme. To do so, it is also important to assume that the interpreta-

Table 1.2: Summary of ontology

Domain	Type	Variables	Metalanguage	Object language	Syntax
entities	e	x, y, z	a, b, c	*tsáaj lăawkhám* 'Jai Lao Kham'	DP
truth values	t	-	0, 1		
situations	s	w, w'	w_0, w_1		
predicates	$\langle e, t\rangle$	P, Q	$\lambda x.\textsc{dog}_w(x)$	*măa* 'dog'	NP
kinds	$\langle s, e\rangle$	k, k', k''	$\lambda w.\iota x[\textsc{dog}_w(x)]$	*măa* 'dog'	
propositions	$\langle s, t\rangle$	p, q	$\lambda w'.Gn\ w \leq w'$ $[\textsc{bark}_w(\iota x[\textsc{dog}_w(x)])]$	*măa hàw* 'dogs bark'	TP
intensional pred.	$\langle s, \langle e, t\rangle\rangle$	N, O	$\lambda w \lambda x.\textsc{bark}_w(x)$	*hàw* 'bark'	V

tion of a particular lexical item in a particular syntactic position will have the same syntactic and semantic properties.

I will also be assuming that nominal expressions with cross-linguistically similar semantic interpretations can be explained by a unified semantic analysis. I am assuming that we can use semantic tests in defined contexts to examine the meaning of expressions. Then, the difference between two linguistic constructions which vary with respect to one word in the same linguistic context comes down to the meaning of that word.

The basic semantics used in this dissertation follow Heim (1998) and the intensional implementation of Von Fintel & Heim (2011).

1.4 Overview

Chapter 2: Bare nouns

Chapter 2 provides the groundwork for the rest of the dissertation by examining the distribution and interpretation of bare nouns in Shan. This chapter engages with the literature on type-shifting and proposes an analysis of bare nouns that includes both existential/generic closure and type-shifting.

Chapter 2 demonstrates that while Shan shares some semantic features with Dënesųłiné (Northern Dene/Athabaskan) bare nouns, as described by Wilhelm (2014). Many of the semantic interpretations that are shared between Shan bare nouns and Dënesųłiné bare nouns are also found with bare nouns in English. There is a low scope existential reading with respect to negation that is consistently possible for bare nouns in these three languages. This can be explained if the operation that generates the existential interpretation for bare nouns is only available below negation.

English and Shan bare nouns can also have kind and generic interpretations. Kind interpretations arise when there is a predicate that applies to a kind argument, and generic interpretations arise through quantification over situations and individuals. Both types of interpretations are possible for bare nouns in Shan and English. I propose to use a base interpretation of $\langle s, \langle e, t \rangle \rangle$ for bare nouns in Shan (and cross-linguistically). The data here has not demonstrated a need to distinguish between bare nouns in Shan versus English. Instead, it seems that the significant differences between these two languages are (i) English has a definite determiner where Shan does not, and (ii) English has plural morphology where Shan does not.

The variety of interpretations available for Shan bare nouns comes about through a combination of existential closure and type-shifting. Type-shifting covers the definite and kind interpretations. The generic and indefinite interpretations can be analyzed through operations that bind the individual and situation variables associated with the nominal.

Chapter 3: Classifiers

Chapter 3 addresses classifiers and nominal structure above the noun. This chapter describes the variety of classifiers found in Shan and proposes that the basic semantic contribution of the classifier is to specify the unit by which the nominal expression is measured. The classifier distribution in Shan provides evidence to support a syntactic structure where the noun and classifier first combine together before combining with a numeral. This accounts for the repeater classifier constructions

and the evidence that numeral-classifier expressions do not modify nouns in the way that adjectives do. There is also evidence that classifiers in Shan overtly represent the atom/unit-identifying component of stubbornly distributive predicates, which are predicates that apply to the atomic individuals that make up the nouns they modify.

This chapter uses the proposed classifier semantics and the apparent syntax of numeral-classifier expressions to propose a unified semantics of count and measure. The head noun and classifier semantics determines whether counting or measuring is taking place. In the case of count nouns, which have inherent identifiable atomic parts, only counting can occur. When there is a mass noun, which does not have inherent identifiable atomic parts, the default is for measuring to take place. However, in some cases, such as in Chinese, there is evidence that the interpretation includes both counting and measuring. This is due to the classifier–which is derived from a noun–maintaining some nominal semantics with the result that the classifier measures both the volume of the mass noun that describes the contents and the number of containers that are described.

In Mandarin Chinese, there is some evidence that the count and measure expressions are syntactically distinguished. When a morpheme *de* intervenes between a classifier and its noun, a measure interpretation is required. This can come about if the classifier semantics is somewhat different than the one proposed for Shan. Instead of being something of type $\langle\langle e,t\rangle,\langle e,t\rangle\rangle$, the classifier is of type $\langle e,t\rangle$ making it semantically compatible with the numeral semantics proposed. As a result, the numeral-classifier expression can form a constituent before it combines with a noun. This unified semantics for count and measure is desirable when one considers languages where there is even less evidence of a distinction between counting and measuring syntax and morphology. This can be seen in Nez Perce.

Chapter 4: Typology of definiteness

Chapter 4 addresses the typology of definiteness marking based on evidence from definiteness marking in Shan. This chapter shows that Shan can express both unique and anaphoric definiteness using bare noun expressions. This is contrary to the typology that had been proposed by Jenks (2018) based on evidence from Thai and Mandarin.

A variety of languages, including Shan, Serbian, and Kannada, can express anaphoric definiteness using a bare noun. Based on this evidence, I give a revised typology of definiteness marking, including the possibility of kind interpretations in a language. There is also some discussion of the pragmatic factors involved in choosing between definite expressions. I assume the goal of more morphological marking connected to definiteness is motivated by a need to uniquely specify the intended referent, as suggested by Ahn (2019). This chapter ends by discussing issues related to different kinds of ambiguity in connection with definiteness marking.

Chapter 5: Left-located constructions

Chapter 5 argues for a semantics of left-located constructions in Shan that specifies the situation argument by which a clause is interpreted. This chapter examines the constructions in Shan that can appear in a left-located position. These include nominals and prepositional phrases. One type of construction that can appear in this position is what looks like an internally headed relative clause. These internally headed relatives are nearly identical to both externally headed and headless relatives, except for the position of the head. This chapter also introduces the particle *kɔ* which is involved in universal quantification over situation arguments.

Chapter 5 takes the semantics that was developed for the components of nominal constructions over the course of dissertation and demonstrates how these apply to analyzing donkey anaphora in Shan. These constructions involve universal quantification over situations. The donkey anaphors

are evaluated with respect to the quantified situation. This can be used to predict what morphology is required for donkey anaphora.

This is the first compositional semantic analysis of topic positions. This can be used to explain various phenomena associated with certain left-located construction, such as why in the ordering of topics it is necessary to go from a larger situation to a smaller one.

Takeaways

This dissertation demonstrates the value in a complete compositional analysis of nominal expressions in a language. Having an understanding of the basic nominal constructions in a language are essential to understanding more complicated constructions, like donkey anaphora. Additionally, it proposes several cross-linguistically applicable analyses. Examining data from an understudied language, like Shan, can give us insight into constructions that are found in English and cross-linguistically.

CHAPTER 2

BARE NOUNS CROSS-LINGUISTICALLY

DEFINITENESS AND QUANTIFICATION: EVIDENCE FROM SHAN

Mary Rose Moroney, Ph.D.

Cornell University 2021

This chapter introduces issues related to the semantics of bare nouns cross-linguistically. It does so by focusing on the distribution and interpretation of bare nouns in Shan, a Southwestern Tai language of Myanmar. Fieldwork data demonstrates that Shan bare nouns have many semantic parallels to bare nouns cross-linguistically, including English bare plurals and mass nouns. This leads to the proposal that bare count nouns in Shan are semantically similar to nouns like *furniture* in English and that Shan mass nouns are similar to English mass nouns (Chierchia 1998; Dayal 2004). I propose that bare nouns cross-linguistically are type $\langle s, \langle e, t \rangle \rangle$ which can be existentially/generically closed over or type-shifted to other interpretations, and these type-shifting possibilities depend on the particular syntax and nominal morphology in a language.

2.1 Introduction

The main bare noun interpretations that are found in Shan, summarized in (4), are the indefinite, definite, generic, and kind interpretations. These are consistent with interpretations reported for bare nouns in Mandarin (Yang 2001), Hindi (Dayal 2004), Teotitlán del Valle Zapotec (Deal & Nee 2018), among others.

In (4a), the bare noun is interpreted as either definite or indefinite. There is an imperfective aspect marker *jù*, which makes the episodic reading more salient in a similar way to how 'is barking' does in English. This is important because episodic interpretations with this verb are not compatible with a generic interpretation. In (4b), the generic interpretation is more salient because there is no imperfective aspect marker. This expression is about dogs generally rather than about specific dogs or instances of dogs barking. In (4c), the nominal has a kind interpretation. This is due to the kind verb *wɔt.wáaj.hăaj* 'go extinct'.

(4) Shan Bare Noun Interpretations

a. măa hàw jù.
dog bark IPFV
'Dogs are barking.' *indefinite*

'The dog(s) is/are barking.' *definite*

b. măa hàw.
dog bark
'Dogs bark.' *generic*

c. măa wɔt.wáaj.hăaj kwàa jâw.
dog disappear go finish
'Dogs are extinct.' *kind*

This chapter investigates how to analyze and account for the bare noun interpretations that are available in Shan. In section 2.2, I discuss the distribution of bare noun interpretations in Shan and how they compare to bare nouns in other languages. Bare nouns in Dënesųłiné, Shan, and English

can be interpreted as low scope existential expressions, generics, or kinds freely, but Dënesųłiné and Shan bare nouns can be interpreted as definite, where they cannot in English. Section 2.3 compares a predicate versus kind interpretation of bare nouns in Shan and discuss my proposed analysis of Shan bare nouns. I argue that a combination of type-shifting and existential/generic closure can account for the distribution of interpretations that are available for bare nouns in Shan and other languages. Section 2.4 adds data on NOUN DEMONSTRATIVE (N DEM) and NOUN DETERMINER (N DET) expressions and integrates this information into the proposed analysis. Section 2.5 concludes this chapter.

2.2 Bare noun interpretations in Shan

2.2.1 Indefinite interpretation

This section focuses on the indefinite interpretation of Shan bare nouns, comparing it to Dënesųłiné (Northern Dene/Athabaskan, Wilhelm 2014), another language with bare noun arguments.

As previously discussed, Shan bare nouns are number neutral, meaning they can be used to refer to one or more individuals. A bare noun, such as *măa* in (5), could describe one or more individuals. In (6), the noun *măa* does not change based on the number it combines with. Thai, a related language, has similarly been described as having number neutral nouns (Jenks 2011).

(5) măa hàw jù.
dog bark IPFV
'(The) dogs are barking.'
'A/The dog is barking.'

(6) háw mí măa nɯŋ/sɔ̌ŋ tǒ
1 have dog one/two CLF.ANML
'I have one/two dog(s).'

There are other bare argument languages that are number-neutral, such as Dënesųłiné, shown in (7), and Mandarin Chinese (MC), shown in (8).

(7) John be-bes-é
J. 3-knife-PNS
'John's knife/knives'

(Dënesųłiné, Wilhelm 2014: (2))

(8) Hufei mai shu qu le.
Hufei buy book go SFP
'Hufei went to buy a book/books.'

(MC, Cheng & Sybesma 1999: (1a))

Shan is like other languages in requiring that an individual be overtly introduced before an anaphoric pronominal can be used to refer to the individual. This is the Formal Link Condition described by Postal (1969), Kadmon (1987), and Heim (1990). In (9), *mé* 'wife' is an overt argument of the first clause, so *mán-náaŋ* 'she' can be used to refer back to her. In (10), the first clause says that Jai Kham married, but does not overtly mention any wife. This makes using a pronoun in the next sentence to refer to the wife infelicitous. Overt reference to the wife in the second clause with a possessive pronoun (*mé mán* 'his wife'), as in (11), is required in order to refer to this individual.

(9) tsáaj khám mí mé jâw kójkaa mán-náaŋ ʔàm lajtsăj mán-tsáaj / tsáaj khám
Jai Kham has wife PFV but 3-FEM NEG like 3-MASC / Jai Kham
'Jai Kham has a wife already, but she doesn't like Jai Kham.'

(10) tsáaj khám tɛ̀ naahɤ́n jâw kójkaa #mán-náaŋ ʔàm lajtsăj tsáaj khám
Jai Kham start household PFV but #3-FEM NEG like Jai Kham
'Jai Kham is married already, but *she doesn't like Jai Kham.'

(Note: *tɛ̀ naahɤ́n* is used to mean 'to marry')

(11) tsáaj khám tɛ̀ naahɤ́n jâw kójkaa mé mán ʔàm lajtsăj mán-tsáaj
Jai Kham start household PFV but wife 3 NEG like 3-MASC
'Jai Kham is married already, but his wife doesn't like him.'

In (12), we see evidence that bare nouns in Shan can introduce discourse references that can be referred back to using a pronoun. Two significant characters in this story are introduced with bare nouns. First, *kón-waan* 'villager' is introduced with a bare noun and then referred back to using a pronoun *mán-tsáaj* 'he' in the following sentence in (13). Then, *mǐ* 'bear' is introduced with a bare noun in (12) and referred back to using a N (CLF) DEM expression in (13).

(12) ti waan nɯŋ mí kón-waan khaw kwàa hăa khɔ̆ŋ-pàa ti
at village one have person-village enter go search_for things-forest at

náj thɤ̀n-pàa sĕ laj kwàa ɲáa mĭ̌.
in jungle-forest and ACH go encounter bear

'One day a villager went looking for forest things in the forest and came upon a bear.'

(13) mán-tsáaj kɔ júɪ́ sàj mĭ̌ nân kójkaa mĭ̌ tŏ nân ʔàm tăaj lɛ
3-MASC PRT shoot into bear that but bear CLF.ANML that NEG die and
'He then shot the bear, but that bear did not die and...'

The Dënesųłiné example in (14) demonstrates that bare nouns can introduce discourse referents that can be referred back to. Here, *Dëneyu Tsádhekúę hots'į* 'a man from Edmonton' introduces a discourse referent, and the following sentences gives the name of that specific individual.

(14) Scenario: In a store.

A: Ts'éré nets'į hú?
blanket 2SG.from Q

'Do you have blankets?'

B: Dódí sį. Dëneyu Tsádhekúę hots'į horelyú ts'éré nádághéłnígh.
3S.nothing DISC man Edmonton AR.from all blanket DISTR.PERF.3S.buy

'No, I'm all out. A man from Edmonton bought them all.' (∃ > ∀)

K'ąbí (dé) nįda xa. Paul húlye sį.
morning (when) ITER.3S.one_goes_IMPF FUT Paul 3S.named DISC

'He will come back tomorrow. His name is Paul.' (Dënesųłiné, Wilhelm 2014: (19))

As (15) shows, bare noun objects in Shan with an existential interpretation can only take narrow scope with respect to negation. We would expect this example might be felicitous with a definite interpretation. This would require that there was some unique dog in the situation known to both the speaker and hearer. Since that is not the case in this context the definite interpretation is ruled out. This example also demonstrates that specific indefinites in the sense of Fodor & Sag (1982) are not possible in Shan, since there is a specific dog that the speaker does not see in the context of (15).

(15) kǎw ʔàm laj hǎn mǎa
1.CAS NEG ACH see dog
'I didn't see a dog.' (¬ > ∃, *∃ > ¬)

a. (Context: I saw three of four dogs.) - # infelicitous

b. (Context: I saw no dogs.) - ✓felicitous

The examples in (16) demonstrate how bare noun subjects work with respect to negation. These are non-anaphoric uses of the bare noun, since no dogs have been mentioned previously. (16) has the bare noun subject *mǎa* which linearly precedes negation. However, as shown by the compatibility of possible follow ups in (16a)–(16b), the only interpretation available is where negation takes scope over the existential. If (16) had the meaning 'there are dogs that did not see me', (16a) would be compatible with that interpretation. The compatible follow up is (16b) which has the interpretation where negation takes scope over the existential.

(16) mɤnâj háw laj lâklɔ̌m khaw náj hóŋjǎa-sát sě mǎa ʔàm laj hǎn háw
today 1 ACH sneak enter in hospital-animal and dog NEG ACH see 1
'Today, I sneaked into the vet, and dogs did not see me.'

a. #mǎaŋ tǒ hǎn háw sě hàw kójkaa mǎaŋ tǒ nɔ́n jù
some CLF.ANML see 1 and bark but some CLF.ANML sleep IPFV
intended: 'Some saw me and barked, but some were sleeping.' (*∃ > ¬)

b. kɔpnǎn tsɤ́-khǎw ʔàm hàw háw
therefore PL-3PL NEG bark 1
'Therefore, they did not bark.' (¬ > ∃)

Bare nouns in Dënesųłiné, similarly, can have only low scope existential interpretations in the presence of negation, as in (17).

(17) Cue: There are 3 bear cubs outside, running into the bush. Andrea looks outside, but only sees two of the cubs (the first one is already gone).

#Andrea sasaze ghe?ı́ híle.
Andrea bear.DIM PERF.3S.see NEG
'Andrea didn't see a bear cub.' (¬ > ∃; *∃ > ¬) (Dënesųłiné, Wilhelm 2014: (20))

Thus far, Dënesųłiné and Shan bare nominals seem to have similar properties with respect to number-neutrality, scope possibilities for indefinite readings, and introducing discourse referents.

2.2.2 Definite interpretation

Looking back at the Dënesųłiné example (14), Wilhelm (2014) notes that *dëneyu* 'man' can be used to refer to the person named *Paul* subsequently. This is an anaphoric definite bare noun use. Similarly, Wilhelm (2008) says that Dënesųłiné bare nouns are translated as definite or indefinite, depending on the context.

Shan also allows definite interpretations of bare nouns. In part of another story about bears (18)–(20), *mǐ* 'bear' in (20) refers back to the bears that ate the fish in (18). The bare noun *pǎa* 'fish' in *ʔàm hǎn pǎa* '(they) didn't see (the) fish' in (19) is compatible either an interpretation where the existential quantifier takes low scope with respect to negation or where *pǎa* refers anaphorically to the fish that the people had and which the bears ate.

(18) phɤ̀ thŭŋ kǎaŋ-kham máa, mǐ khaw máa náj hɣ́n sě kǐn pǎa pɛt
then until middle-night come bear enter come in house and eat fish cast_away
'Then at night, bears came into the house and ate the fish.'

(19) mɤ khǎw tùn khɯn máa nân tsɣ́-khǎw ʔàm hǎn pǎa hǎn mɛn hɔ́j tǐn mǐ
when 3.PL wake up come that PL-3.PL NEG see fish see right mark foot bear
pɔ́ těm kwàa seŋ
extent full go end
'When they woke up, they did not see the fish but saw bear footprints everywhere.' (ι)
'When they woke up, they did not see any fish but saw bear footprints everywhere.' (¬ > ∃)

(20) khǎw kɔ ʔùp kǎn sě waa tě hét tsɣ́ŋhŭ tĵɲɔ́p mǐ (tsɣ́ nân)
3.PL PRT talk together and say IRR do how catch bear (PL that)
'They talked about how to catch the bears.'

In the same way, (21) has two examples of anaphoric bare nouns. First, *waan* 'the village' refers back to a village that was introduced in an earlier part of the story. Then we see the bear that

is one of the main characters of this story be referred to anaphorically twice. One uses a N CLF DEM construction *mǐ tǒ nân* 'that bear' and the other uses the bear noun *mǐ* 'the bear'.

(21) mí wán nɯŋ mǐ tǒ nân khaw máa náj waan sě pɔ́ hăn kón kɔ lám khóp
have day one bear CLF.ANML that enter go in village and if see person PRT chase bite

lám jáaj thŭŋ ti mí kón-tsáaj kɔ̂ nɯŋ ɲáa mǐ khóp thŭŋ ti pɔ́
chase scratch until at have person-MASC CLF.HUM one encounter bear bite until at extent

laj lûtăaj kwàa jù jâw
ACH die go stay PFV

'One day that bear came into the village and if it saw a person would chase, bite, and scratch them until one man was bitten by the bear and died.'

This section has demonstrated that bare nouns in both Dënesųłiné and Shan can have a definite interpretation. I will go over the interpretations of N DEM constructions in Shan in section 2.4, and a more detailed discussion of definiteness marking in Shan will be presented in Chapter 4.

2.2.3 Kinds and generics

I am following Diesing (1992) and Carlson & Pelletier (1995) among others in assuming that generics involve quantification over individuals and situations. This is distinct from kinds which are the intensional groups consisting of individuals with the property of being a member of that kind. Table 2.1 demonstrates different kinds of generic characterizing sentences in contrast to episodic sentences. One issue is that the nature of characterizing sentences makes it difficult to distinguish kind referring from object referring nominal expressions. Episodic sentences with kind reference, such as *Marconi discovered the radio*, more clearly demonstrate that the nominal can refer to a kind.

Table 2.1: Mendia & Filip (inprep): (5)

	Generic Characterizing sentence	Episodic sentence
Kind Reference	*Dogs* bark/ *The potato* contains vitamin C.	Marconi discovered the radio.
Individual Reference	*John* smokes.	*John* smoked last night.

Kinds

This section discusses the kind interpretations available with bare nouns and noun-demonstrative expressions in Shan. Wilhelm (2014) notes that Dënesųłiné bare nouns can be interpreted as kinds, as (22) shows. Shan bare nouns can also be interpreted as kinds.

(22) ts'ékui yú-é kánesta sį
woman clothing-PNS IMPF.1SGS.search DISC
'I'm looking for women's clothing' (e.g., in a store) (Dënesųłiné, Wilhelm 2014: (38b))

I will assume that the subject of the verb *wɔt.wáaj.hăaj kwàa jâw* 'go extinct' and the object of the verb *khŏŋsaaŋ* 'invent' must be kinds. In (23), the subject of the verb meaning 'go extinct' is the bare noun *nôk* 'bird', so the bare noun can refer to a kind.

(23) nôk wɔt.wáaj.hăaj kwàa jâw.
bird disappear go finish
'Birds are extinct.'

(24) and (25) demonstrate that Shan bare nouns can refer to kinds even in object position. In (24), the verb *khŏŋsaaŋ* 'invent' takes the bare noun object *káa*, which refers to cars as a kind. In (25), the bare noun *măa* 'dog' refers to dogs as a kind.

(24) Karl Benz pĕn kô khŏŋsaaŋ káa ʔɔ̆ntáaŋ pɤn
Karl Benz COP CLF.HUM invent car before others
'Karl Benz is the first person who invented cars.'

(25) mɤ nɯŋ mù̀n pĭ pôn máa nân kón ʔɔ̆nkăn lêŋ măa ʔɔ̆ntáaŋ pɤn
when one ten.thousand year past come that people together raise dog first others
'10,000 years ago, humans first took care of dogs.'

Sub-kinds

A sub-kind is an expression that is compatible with a kind interpretation that refers to part of a larger kind. For example, *owl* is a sub-kind of *bird*. (26) sets up a response where a sub-kind of bird (*kâw* 'owl') and a sub-kind of big cat (*sў-láajkàajkɔ̆n* 'royal tiger') can be referred to anaphorically. As (26a) shows, a bare noun denoting the complete kind (*nôk* 'bird') or (*sў* 'big cat') cannot be used to refer to its sub-kind anaphorically. (26b) shows a felicitous follow up to (26) which includes the bare nouns that describe the specific sub-kinds.

(26) laat nɛ lɔŋ kâw lɛ sў-láajkàajkɔ̆n tsɔ̀ŋ laj khaa.
say show about owl and big.cat-bengal POLAR.Q can HUMB
'Could you tell me about owls and Bengal tigers?'

a. #sў lɛp tě wɔtwáajhăajkwàa jâw kójkaa nôk tê jáŋ mí năm nàa náj lumfâa
big.cat likely IRR extinct PFV but bird CONTR still have many very in world
intended: 'The big cat will likely go extinct soon but the bird is still common in the world.'

b. sў-láajkàajkɔ̆n lɛp tě wɔtwáajhăajkwàa jâw kójkaa kâw tê jáŋ mí năm nàa náj lumfâa
big.cat-bengal likely IRR extinct PFV but owl CONTR still have many very in world
'Bengal tigers will likely go extinct soon but owls are still common in the world.'

Generics

It is also possible for bare nouns in Shan to have a generic interpretation. In (27), the bare noun subject *măa* 'dog' is interpreted generically, and in (28) the object *môt* 'ant' is interpreted generically.

(27) măa kâj lɯp-lám hăaŋ tŏ-tsawkàw mán khɯ́n
dog often chase-bite tail body-self 3 back
'Dogs often chase and bite their own tails.'

(28) háw ʔàm lajtsǎj môt
1 NEG like ant
'I do not like ants.'

Dënesųłiné bare nouns can also be generic, as in (29).

(29) sas xaye k'étłá thetez
bear winter to_end_of IMPF.3S.several_sleep
'Bears sleep all winter.' (Dënesųłiné, Wilhelm 2014: (6), emphasis mine)

It is also possible to have generic (characterizing) interpretations when the nominal refers to a specific individual, as shown in (30). Here, the pronoun *mán* 'he/she/it' refers to a salient dog. The sentence describes a characteristic of the dog.

(30) mán tɯ́ɯn hàw tsɤ̂ŋ-nǎn tàasè
3 always bark how-like.that always
'He always barks like that.'

Summary

Table 2.2 demonstrates the similar features of Shan and Dënesųłiné bare noun interpretations. For all the possibilities tested, they have the same available interpretations. These include existential interpretations that can introduce discourse referents but only take low scope with respect to negation. Bare nouns can also have definite, kind, and generic interpretations in both languages.

2.2.4 Shan mass nouns

The previous discussion of bare nouns in Shan has included only nouns that tend to be COUNT NOUNS, which refer to individuals that are inherently countable in some sense. This section demonstrates

Table 2.2: Comparison between Shan and Dënesųłiné

	Dënesųłiné	**Shan**
introduce dref	✓	✓
Ex.	(14)	(12)–(13)
¬ > ∃, *∃ > ¬	✓	✓
Ex.	SUBJ: (17)	SUBJ: (16) OBJ: (15)
Def	✓	✓
Ex.	-	(18)–(20)
Kind	✓	✓
Ex.	(22)	(24)
Generic	✓	✓
Ex.	(29)	(27)

that MASS NOUNS, which refer to things that are not inherently countable, pattern in the same way as count nouns with respect to the possible interpretations.

Indefinite

As (31)–(32) demonstrates, bare mass nouns in Shan have the low-scope existential reading, just like count nouns. There is more to say about the scope of negation and serial verb constructions in this example, but I will not address that here.

(31) kăw hăa nâm ʔàm hăn
1.CAS look_for water NEG see
'I didn't find water.' (¬ > ∃)

(32) khaa ʔàm pàj laj kǐn nâm
1.HUMB NEG yet ACH drink water
'I have not drunk water.' (¬ > ∃, *∃ > ¬)

Definite

As shown in (33), mass nouns can be used anaphorically.

(33) mɤnâj háw sûː khaw lɛ thò.
today 1 buy rice and bean

thò nâj kaakhăn ʔàm pɔ́ jàj. khaw tɛ̂ mí kaakhăn jàj nàa.
bean this price NEG quite big rice CONTR have price big very

'Today I bought rice and beans. The beans were not expensive, but the rice was very expensive.'

Kind and Generic

Bare mass nouns can have a kind interpretation, as (34) shows for *khaw* 'rice' and (35) shows for *khám* 'gold'. They can also be interpreted generically, as with *khám* 'gold' in (36).

(34) mɤlěw nâj hăa khaw jàap nàa
now this find rice difficult very
'Now it is difficult to find rice.'

(35) haa hěŋ pǐ pôn máa nân kón tsâj khám ʔɔ̌ntáaŋ sút
five thousand year past come that people use gold first absolutely
'Humans first used gold about 5000 years ago.'

(36) khám mí kaakhăn jàj tɛ̂tɛ̂
gold have value big really
'Gold is really valuable.'

Both Shan count nouns and mass nouns can have low-scope indefinite, definite, kind and generic interpretations.

2.2.5 A comparison with English

The previous sections have identified some semantic parallels between bare nouns in Dënesųłiné and bare nouns in Shan. However, it seems important to point out that bare nouns in English have some of the same possible semantic interpretations.[1]

[1]There has even been some debate about whether English bare plurals have a number neutral interpretation (Sauerland et al. 2005).

Indefinite

Bare nouns in English also have the property of taking low scope with respect to other operators, specifically bare plurals as in (37) and mass nouns as in (38). English bare plurals are well known to only allow narrow scope readings (Heusinger 2019).

(37) Today, I did not buy books. ($\neg > \exists$, $*\exists > \neg$)

(38) Today, I did not drink water. ($\neg > \exists$, $*\exists > \neg$)

These, of course, can introduce discourse referents, as (39)–(40). The *books* introduced in the first sentence of (39) is referred back to using *they* in the second sentence. Similarly, *water* in the first sentence of (40) is referred back to using the pronoun *it* in the second sentence.

(39) Today, I bought books at the store. They were expensive.

(40) Today, I drank water after going jogging. It was delicious and refreshing.

Looking more carefully at how existential scope possibilities of bare nouns interact with other operators, there is clearly a subject-object asymmetry. In comparing the bare noun interpretations (42) to nouns with indefinite expressions *a*/*one* (43) and *some* (44), there are clear differences in scope possibilities. For each of these, the (a.) follow up is compatible with a high-scope existential interpretation and the (b.) follow up is compatible with a low-scope existential interpretation. (42)–(44) all involve indefinite subjects with negation. We see that the bare noun subject is only compatible with the low scope existential interpretation, and the nouns with *a*/*one*/*some* are only compatible with a high-scope existential interpretation.[2]

[2]For (43) and (44) there is a cancelable implicature that the low scope interpretation is false. So it is possible to follow up (43) and (44) with *In fact, none of the students are wearing shorts.*

(41) CONTEXT: When I was in school, we had to wear the uniform shorts.

(42) Visiting the school now, I see that students are not wearing shorts.

a. #Many are wearing shorts, but some wear trousers or skirts. (*$\exists > \neg$)

b. They all are wearing trousers or skirts. ($\neg > \exists$)

(43) Visiting the school now, I see that a/one student is not wearing shorts.

a. Many are wearing shorts, but she is wearing trousers. ($\exists > \neg$)

b. #They all are wearing trousers or skirts. (*$\neg > \exists$)

(44) Visiting the school now, I see that some students are not wearing shorts.

a. Many are wearing shorts, but some are wearing trousers. ($\exists > \neg$)

b. #They all are wearing trousers or skirts. (*$\neg > \exists$)

The same pattern seems to hold when the relevant indefinite expression is the object under negation. The object bare plural in (46) is only compatible with a low scope interpretation. With the expressions *a*/*one*/*some* in (47)–(48), the high scope interpretation is more prominent.

(45) CONTEXT: When I was in school, the students erased all the blackboards after class.

(46) When I visited the school today, the students did not erase blackboards.

a. #They erased many blackboards, but some were not erased. (*$\exists > \neg$)

b. None of the blackboards were erased. ($\neg > \exists$)

(47) When I visited the school today, the students did not erase a/one blackboard.

a. The students erased many blackboards, but one was not erased. ($\exists > \neg$)

b. None of the blackboards were erased. ($^{?}\neg > \exists$)

(48) When I visited the school today, the students did not erase some blackboards.

a. The students erased many blackboards, but some were not erased. ($\exists > \neg$)

b. #None of the blackboards were erased. (*$\neg > \exists$)

We do not necessarily expect that Dënesųłiné bare nouns, Shan bare nouns, and English mass nouns and bare plurals are all licensed by the same syntactic constructions. However, what might be more useful to take away from this data is that bare nouns that are syntactically licensed (by whatever means) can have low scope existential interpretations.

Definite

This section goes over definite interpretations of bare nouns in English. (49)–(50) give examples with the mass noun, *rice*. With a bare mass noun in (49), the definite *rice* in the second sentence is ungrammatical.[3] In contrast, when there is an overt definite determiner as in (50), it is grammatical. (51)–(53) give examples with count nouns. Again, the example with a bare count noun cannot have a definite interpretation, as in (51), but when it does have an overt determiner, as in (53), it is grammatical.

(49) I bought some rice and beans. #Rice was very expensive.

(50) I bought some rice and beans. The rice was very expensive.

(51) I adopted a cat and a dog. *Cat is very fluffy.

(52) I adopted cats and dogs. #Cats are very fluffy. (KIND: ✓, DEFINITE: ✗)

(53) I adopted a cat and a dog. The cat is very fluffy.

(54) I adopted cats and dogs. The cats are very fluffy.

English, unlike Shan, has an overt definite determiner *the*. Shan uses a demonstrative *nâj*/*nân* 'this'/'that' in some of the places that English would use a definite determiner. However, they have a different semantic effect. The determiner is obligatory to express anaphoric definiteness in English, where it is optional in Shan (Moroney 2021). The relationship between having a definite

[3] In other contexts, *rice was very expensive* is grammatical in English. However, it cannot be used to refer back to the rice describe in the first sentence of (49). Only the generic reading, not the definite, is available for *rice*.

determiner in English and the lack of definite bare noun interpretations is discussed in more detail in chapter 4.

Kind

The following are English examples that involve reference to a kind. (55) gives an example with a mass noun. (56) gives an example with a count noun. Finally, (57) is an example with a *furniture/footwear* noun, which will be discussed in section 2.3.1.

(55) Rice was first discovered in 12,000BCE.

(56) Dinosaurs are extinct.

(57) Furniture was first invented in 12,000BCE.

Generic

English bare nouns can also have a generic interpretation, as in (58). In this example, there are two reported interpretations. One (58a) is a generalization about a definite bare noun *this part of the Pacific* in (58). The other (58b) is a generalization about the kind described by *typhoons*.

(58) Typhoons arise in this part of the Pacific. (Milsark 1974; Carlson 1989; Mendia 2020)

a. Object-referring interpretation:
   This part of the Pacific has the property that some typhoons arise in it.

b. Kind-referring interpretation:
   It is a property of typhoons that they arise in this part of the Pacific.

Example (59) has a bare mass noun with a generic interpretation, and (60) has a bare plural count noun with a generic interpretation.

(59) Rice grows in paddy fields.

(60) Tigers eat meat.

2.2.6 Summary: Bare noun interpretations

Table 2.3 compares the semantic interpretations that are possible with English and Shan nouns. It compares count and mass nouns in the two languages and also bare nouns and nouns combined with a determiner in English or a demonstrative in Shan. The Dënesųłiné data is left out of the table because I do not have the distribution across the count/mass categories. With the bare noun interpretations that I have discussed for Dënesųłiné thus far, it does pattern with Shan.

Table 2.3: Comparison between Shan and English bare nouns

	English		**Shan**	
	mass	count	mass	count
Low ∃	✓	✓	✓	✓
Ex.	(40)	(39)	(32)	(12)
High ∃	✗	✗	✗	✗
Ex.	(38)	(37)	(32)	(15)
Def	✗	✗	✓	✓
Ex.	(49)	(51)	(33)	(21)
Kind	✓	✓	✓	✓
Ex.	(55)	(56)	(34)	(23)
Generic	✓	✓	✓	✓
Ex.	(59)	(60)	(36)	(27)

What this table shows is that bare nouns in Shan and English largely pattern together with respect to kind, generic, and low scope existential interpretations. The significant way in which the languages differ is that the definite determiner is obligatory in English and optional in Shan to get a definite interpretation. To account for this, I will propose in section 2.3.3 that bare nouns in Shan can type-shift to a definite interpretation because there are no definite articles in the language.

2.3 Bare nouns: Predicates or kinds?

It is relatively broadly accepted that nominal expressions can denote objects or kinds (Carlson & Pelletier 1995), and that certain verbs are kind-selecting and certain verbs are object-selecting. There is some discussion about the base interpretation of bare nouns cross-linguistically.

Krifka (1995) first proposed that bare nouns in Chinese are kinds, citing two main reasons for assuming that bare nouns represent kinds: (i) every language that allow bare nouns uses them to refer to kinds, and (ii) kinds seem 'ontologically prior to specimens' (399).

Chierchia (1998) argues that a cluster of properties in languages like Chinese can be explained if NPs denote kinds. These properties are: (i) bare nouns can be arguments, (ii) all nouns are mass nouns, (iii) no morphologically expressed plural, and (iv) a generalized classifier system.

Feature (i) is type based: bare nouns in Shan and Mandarin and bare plurals in English can serve as direct arguments of verbs whereas singular nouns in English cannot. This is taken as evidence that bare nouns in classifier languages are type e (even though kinds are typically represented as $\langle s, e\rangle$). Feature (ii) notes that bare nouns in languages like Chinese pattern similarly to mass nouns in English. This is related to feature (iii) which notes that nouns in languages like Chinese are not marked for singular/plural. Feature (iv) is that there must be a classifier whenever a bare noun combines with a numeral. The kind analysis explains why bare nouns cannot combine with numerals without having a classifier as well (Krifka 1995; Chierchia 1998; Jenks 2011; Dayal 2012).

Li (2011) gives evidence to show that bare nouns in Mandarin are similar to bare plurals in English in that they take narrow scope with regard to other quantifiers and they can be used with kind-denoting verbs. For example, (61) and (62) demonstrate that bare nouns can only have low scope interpretations with respect to universal quantifiers, in contrast to indefinite NUM CLF N expressions, which can have both scope interpretations. While Carlson (1977) and Chierchia (1998)

propose that bare plurals have a kind semantics, they do not rule out the possibility that bare plurals can simply shift into a kind interpretation.

(61) mei-ge ren dou zai kan yi feng guanyu jiaxin de shu.
every-CLF man all PROG read one CLF about add-salary MOD book
'Everybody is reading a letter about raising salaries.' [$\forall > \exists$ OR $\exists > \forall$]

(62) mei-ge ren dou zai kan guanyu jiaxin de xin.
every-CLF man all PROG read about add-wage MOD letter
'Everybody is reading letters about raising salaries.' [$\forall > \exists$]

OR 'Everybody is reading the letter about raising salaries.' [definite]

(Mandarin Chinese, Li 2011: 63, (19))

As demonstrated in the previous section, Shan bare nouns, Dënesųłiné bare nouns, and English bare plurals pattern like Mandarin in that bare nouns take narrow scope with respect to negation and bare nouns can denote kinds. Mandarin Chinese nouns and Shan nouns share the feature that number is not morphologically marked on the noun in either language. However, I do not take these facts as evidence in itself that the base denotations of Shan nouns are necessarily kinds. Instead, I will argue that the base denotation is type $\langle s, \langle e, t \rangle \rangle$, and type-shifting can account for the kind and definite interpretations.

2.3.1 The mass-plural-kind connection

Deal (2017) summarized the distinction between different types of nouns in English as in Table 2.4. This includes many of the distinctions that Chierchia (1998) uses to support analyzing bare nouns in Chinese and mass nouns in English as kinds.

Here we see that there are three types of nouns in English: COUNT NOUNS like *cat*, FAKE MASS NOUNS like *footwear* or *furniture*, and MASS NOUNS like *water*.[4] For properties (a)–(d), *furniture*

[4]The noun category for *furniture* nouns has been called many things, including 'count mass nouns' (Doetjes 1997),

and *water* nouns pattern together: neither can be morphologically pluralized (a), combine directly with numerals (b), or combine with distributive quantifiers like *each* (c), but both can combine with quantificational terms like *much* and *less*. For properties (e)–(g), count nouns like *cat* and *furniture* nouns pattern together: both can combine with stubbornly distributive adjectives (count adjectives) (e), both involve comparison of amounts based on number (f), and neither involve comparison of amounts based on mass or volume (g). Based on these properties, count nouns and mass nouns are distinct in English, and *furniture* nouns have overlapping properties with both types of nouns.

Table 2.4: Properties of three types of English nouns Deal 2017: (15)

	cat	footwear	water
(a) pluralization	✓	*	*
(b) direct combination with numerals	✓	*	*
(c) combination with *each*, *many*, *fewer*	✓	*	*
(d) combination with *much*, *less*	*	✓	✓
(e) combination with 'count adjectives'	✓	✓	*
(f) comparison based on number	✓	✓	*
(g) comparison based on mass/volume	*	*	✓

These properties can be tested in Shan to gain an understanding of how nouns should be categorized. Property (a) is pluralization.[5] As previously mentioned, Shan nouns are number neutral, meaning that plurality tends not to be morphologically marked on the noun. (63) shows that the noun *măa* can be interpreted as plural even when there is no overt plural morphology. The way to mark plurality overtly is with the morpheme *tsɣ́*, shown in (64). This morpheme has the same distribution as classifiers in Shan. I will discuss this morpheme and its relationship with semantic plurality in Shan more in section 3.4.4 in the next chapter.

(63) măa nân jàj nàa
dog that big very
'That/Those dog(s) is/are very big.'

(64) măa tsɣ́ nân jàj nàa
dog PL that big very
'Those dogs are very big.'

'fake mass nouns' (Chierchia 2010), 'neat mass nouns' (Landman 2016a), a.o. I am simply calling them '*furniture* nouns'.

[5]Under Sauerland et al.'s (2005) account of plurality, even 'count' nouns are number neutral and number marking is introduced higher up based on whether a presupposition of singularity is met (leading to singular morphology) or not (leading to plural morphology). Under this account, the distinction between nouns like *cat* and nouns like *footwear* would be whether this distinction was morphologically represented.

Shan is an obligatory numeral-classifier language, meaning that whenever a numeral and noun combine, there must also be a classifier morpheme, as shown in (65) and (66). This means that Shan nouns cannot combine directly with numerals, so all Shan nouns pattern like English *furniture* and mass nouns with respect to property (b). (67) and (68) give examples with nominal expressions consisting of a noun, numeral, and classifier. Shan classifiers will be discussed more extensively in chapter 3.

(65) mă̆a să̆am *(tŏ)
dog three *(CLF.ANML)
'three dogs'

(66) nâm să̆am *(kɔ́k/taw)
water three *(cup/bottle)
'three cups/bottles of water'

(67) ti hɣ́n háw nâj mí mă̆a lɛ mɛ́w să̆am tŏ
at house 1 this have dog and cat three CLF.ANML
'At my house, there are three dogs and cats.'

(68) thɛŋ păn nâm să̆am kɔ́k sàj náj mɔ
add give water three cup put in pot
'Add three cups of water to the soup.'

Property (c) says that English count nouns can combine directly with distributive quantifiers like *each* and *many*, but *furniture* and mass nouns cannot. Shan has a set of distributive quantifiers that require a classifier. These include *ku* 'every' in (69) and *lă̆ajlă̆aj* 'many' in (70). Thus, Shan nouns pattern together with English *furniture* and mass nouns.

(69) tsáaj khám lup mă̆a lɛ mɛ́w ku *(tŏ)
Jai Kham pet cat and dog every *(CLF.ANML)
'Jai Kham pet every dog and cat.'

(70) háw lò kĭ̌n nâm lă̆ajlă̆aj *(taw) jù mɣnâj
1 must drink water many *(bottle) IPFV today
'I need to drink many bottles of water.'

According to property (d), there are quantificational terms like *much* and *less* that combine with *furniture* and mass nouns but not count nouns in English. The Shan equivalent of *much* is *nă̆m* 'much/many' which can combine with both count nouns as in (71) and mass nouns as in (72).

(71) mí mɛ́w nǎm nàa
have cat many very
'There are many cats.'

(72) mɤnâj hàw kǐn wáaj nǎm nàa.
today 1 drink wine many very
'Today I drank a lot of wine.'

A significant way in which count nouns and mass nouns in Shan pattern differently from one another is with respect to property (e), the ability to combine with 'count adjectives'. Count nouns and furniture nouns can combine directly with what Deal (2017) calls 'count adjectives' and Schwarzschild (2007) calls 'stubbornly distributive predicates'. These are predicates like *big* that distribute to the atomic parts of the nominal they modify. For example, *these dogs are big* means that the individual dogs are big, rather than the collection of dogs. These predicates cannot directly combine with English mass nouns felicitously (e.g., *#The water is big.*). What we find in Shan is that typical count nouns like *mǎa* 'dog' in (73) can combine directly with these count/distributive adjectives like *jàj* 'big' and *lêk* 'small', but mass nouns like *nâm* 'water' in (74) cannot. The role of the optional classifier in these examples will be discussed more in section 3.4.1 in the next chapter.

(73) ʔǎn-nâj pěn mǎa (tǒ) jàj. ʔǎn-nân pěn mǎa (tǒ) lêk.
CLF.GN-this COP dog (CLF.ANML) big CLF.GN-that COP dog (CLF.ANML) small
'This is a big dog. This is a small dog.'

(74) #nâm jàj
water big
intended: 'big water'

The last two properties pertain to whether comparison using a particular kind of noun tends to compare the number of individuals described by the noun or the volume/mass described by the noun. Count nouns and *furniture* nouns in English involve comparison by number and mass nouns involve comparison by mass or volume. The context in (75) sets up a situation where two people have mangoes; one has more by number and one has more by volume. When asked about

which person has more mangoes as in (75a), the answer indicates that the count noun *máakmâj* 'fruit' preferences comparison based on number. It is possible to compare based on volume using a slightly different question (75b) and incorporating a mass noun *nɤ̂* 'meat'.

(75) (Context: One person has six big mangoes, one person has ten small mangoes)

a. phǎj mí máakmâj nǎm lɤ̌ phǎj
who.Q have fruit many exceed who.Q
'Who has more fruit than who?'

Answer: Person with 10 small mangoes.

b. phǎj laj kǐn nɤ̂ màakmâj nǎm lɤ̌ phǎj
who.Q ACH eat meat fruit many exceed who.Q
'Who gets to eat more fruit than who?'

Answer: Person with 6 big mangoes.

Table 2.5 summarizes the distribution of nouns in Shan with respect to these features. Distributionally, Shan count nouns pattern like English *furniture* type nouns (also called 'fake mass nouns', among other things), and mass nouns pattern like English mass nouns. Chierchia (2010) similarly said that Mandarin Chinese count nouns behave much like 'fake mass nouns'. Cheng & Sybesma (1999) also identified a mass/count distinction in Mandarin Chinese.

Table 2.5: Shan distribution

	cat	water
(a) pluralization	- (63)	-
(b) direct combination with numerals	*	*
(b.i) combination with NUM+CLF	✓(67)	✓(68)
(c) direct comb. with *ku* 'each', *lǎajlǎaj* 'many'	*	*
(c.i) comb. with *ku* 'each', *lǎajlǎaj* 'many' +CLF	✓(69)	✓(70)
(d) combination with *nǎm* 'much'	✓(71)	✓(72)
(e) combination with 'count adjectives'	✓(73)	* (74)
(f) comparison based on number	✓(75a)	*
(g) comparison based on mass/volume	*	✓(75b)

Deal (2017) argues that languages like Chinese make a distinction between nouns like *furniture* and nouns like *water*, where the difference lies in whether the noun has stable/identifiable atomic

parts or not rather than whether the noun denotes objects or kinds. This distinction between mass nouns and count nouns has been similarly described by Chierchia (2010), Landman (2011), and Grimm (2012), among others, though the precise difference is still debated. Deal (2017) defines GENERALIZED HOMOGENEITY as in (77) to capture these possible analyses.

(76) CUMULATIVITY (adapted from Deal 2017: (6))

A predicate P is cumulative iff any sum of parts that are P is also P.

(77) GENERALIZED HOMOGENEITY (G-HOMOGENEITY) (Deal 2017: (9))

A noun is g-homogeneous iff it denotes a g-homogeneous predicate. A predicate is g-homogeneous iff it is both cumulative and one or more of the following:

a. lacking in minimal parts (divisive)

b. lacking in stable minimal parts

c. lacking in non-overlapping minimal parts

d. lacking in non-strongly-connected minimal parts

Since Shan count and mass nouns seem to pattern with English *furniture* and mass nouns, respectively, I argue that Shan count nouns and mass nouns are CUMULATIVE, defined in (76), but only mass nouns are G-HOMOGENEOUS.

2.3.2 Nominal mapping parameter

Chierchia (1998) claimed that bare plurals in English denote $\langle e, t \rangle$ type predicates which can type-shift to kinds, and mass nouns denote kinds, following Carlson's (1977) analysis of bare plurals. Part of Chierchia's (1998) Nominal Mapping Parameter analysis takes for granted that in languages like Chinese the mass/count distinction does not affect whether the nouns denote properties or kinds. As already has already been mentioned, Cheng & Sybesma (1999) argue that there

is a count/mass distinction for Chinese. Deal (2017) argued that not only does Chinese have a mass/count distinction, but even languages with a less clear distinction can be shown to differentiate mass and count nouns. As has been shown in section 2.3.1, Shan also has a mass/count distinction.

- Chierchia's 1998 argument of why bare nouns are kinds:
  - bare and mass nouns can function as arguments
  - bare and mass nouns cannot be pluralized because already plural
  - bare and mass nouns need a classifier to individuate before combining with a numeral

A significant part of the arguments about noun denotations has to do with the fact that bare nouns in some languages and bare plurals in other languages pattern together in certain features. This I do not dispute. However, these properties (such as low scope existential interpretations with respect to negation) do not directly rely on a kind interpretation of the bare noun. Most arguments in support of analyzing bare nouns as kinds rely on demonstrating that they have features similar to other bare noun languages, features that do not directly connect to a kind-denotation of the noun.

I am going to raise the following objections to Chierchia's (1998) arguments for a kind analysis of bare nouns:

- The fact that bare nouns function as arguments does not mean they must denote kinds if we allow for type-shifting, which is already part of Chierchia's (1998) analysis
- Assuming that Shan nouns are equivalent to *furniture* and mass nouns in English can account for the properties that distinguish count nouns from Shan nouns
  1. We can explain why Shan nouns are not morphologically pluralized: English *furniture* and mass nouns similarly cannot be pluralized.

2. We can explain why Shan nouns need classifiers to combine with numerals: the same is true of *furniture* and mass nouns in English.

2.3.3 Type-shifting analysis basics

This section introduces the standard type-shifting analysis of bare nouns. I will propose to use a slightly modified version of this. Chierchia (1998) proposed a Neo-Carlsonian approach following Carlson (1977), claiming that bare plurals in English are type $\langle e, t\rangle$, and they can type-shift to function as arguments. The type-shifting operators defined by Chierchia (1998) and updated by Dayal (2004), are given below:

(78) TYPE-SHIFTING OPERATORS (Dayal 2004):

a. $^{\cap}$: $\lambda P \lambda s.\iota x[P_s(x)]$

b. $^{\cup}$: $\lambda k_{\langle s,e\rangle}\ \lambda x.x \leq k_s$

c. ι: $\lambda P.\iota x[P_s(x)]$

d. $\exists$: $\lambda P \lambda Q.\exists x[P_s(x) \wedge Q_s(x)]$

In a language where there are are no determiners, you would expect all type-shifting operations to be available, but according to Dayal (2004), $\exists$-type-shifting does not occur in these languages. Dayal (2004), revising Chierchia (1998), proposes that the type-shifting operators follow a hierarchy, where kind-forming $^{\cap}$ and entity forming ι must be ruled out before $\exists$ becomes available. This rule is defined in (79). The justification is that using $^{\cap}$ or ι is a less drastic change because it does not introduce quantificational force. Dayal (2004) claims that bare nouns are equally allowed to form kinds or entities, so they must be ranked equally.

(79) MEANING PRESERVATION: $\{^{\cap}, \iota\} > \exists$

$\exists$ is only available as a type-shifting operator when the others are ruled out. There is a narrow scope existential reading available for bare plurals in English that can be seen in negated sentences, as in (80). Chierchia (1998) analyzes this reading as coming from Derived Kind Predication, defined in (81), which fixes type mismatches between the verb and the argument.

(80) John didn't see machines. (Chierchia (1998): (49d′))

(81) *Derived Kind Predication (DKP)* (Chierchia (1998): 364)

If P applies to objects and k denotes a kind, then $P(k) = \exists x[^{\cup}k(x) \wedge P(x)]$

As mentioned in the previous paragraph, Dayal (2004) proposes the ranking of type-shifting operators in (79). This ranking is based on the data from Hindi in (82)–(83). Under the previous ranking, either the $\exists$ (existential) or ι (definite) type-shifters would be available if the $\cap$ (kind) shifter was blocked. (82) shows that *is mashin ke TukRe* 'parts of this machine' is not compatible with a verb that requires a kind-denoting argument, meaning type-shifting to a kind is blocked for this expression. Then, it would be predicted that a wide scope existential reading would be available. However, (83) shows that this is not the case. Dayal (2004) argues that the possibility of type-shifting with ι blocks $\exists$.[6]

(82) is mashin ke TukRe aam haiN
this machine of parts common are
'Parts of this machine are common.'

(83) anu-ne is mashin ke TukRe nahiiN dekhe
Anu-ERG this machine of parts not see-P
'Anu didn't see any/the parts of this machine.' (Dayal (2004): (45a,b))

Since the only linguistics expressions that are known to block $\cap$ type-shifting are expressions that include a definite component—*is mashin* 'this machine' in (82)–(83))—there will be no type-shifting using $\exists$ predicted in a language without an overt definite article. Given that Shan does

[6]In English, this blocking does not occur since ι type-shifting is unavailable due to the presence of an overt definite determiner in the language.

not have an overt definite article, we would expect that wide scope indefinite readings would be impossible under this analysis.

2.3.4 Proposed semantics

I propose that the basic interpretation of bare nouns cross-linguistically is type $\langle s, \langle e, t\rangle\rangle$.[7] Existential/generic closure can account for the broadly available low scope existential and generic interpretations. Kind and definite interpretations come about through type-shifting in Shan, since it lacks an overt definite determiner.

Let's say that a situation is a part of a world.[8] They are the same type s. I propose that the declarative sentences are formed by saying that the situation described by the clause is a part of a given world. If a proposition is a set of worlds, then a clause describes a situation that makes up a part of each individual world in that set of worlds. For example, (84) would be the semantics for the proposition *a dog barked*. The extension of a property is given at a particular situation, which may be a world or just a part of a world.

(84) $\lambda w.\exists w' \leq w \exists x[^{*}\textsc{dog}_{w'}(x) \wedge \textsc{bark}_{w'}(x)]$

Indefinite

Rather than using Derived Kind Predication from Chierchia (1998), I will follow Diesing (1992) in assuming that bare nouns below a certain syntactic position, at the verb phrase level, are in-

[7] Von Fintel & Heim (2011) propose two distinct syntactic types of world pronouns which leads to the idea that "predicates inside nominals can be freely indexed but that the ones inside predicates are captured by the closest λ operator" (108). They connect this difference to the presence or absence of a determiner on the nominal, but I will assume that both types of world pronouns are possible with bare nouns in Shan. I assume this because we see both bound variable world interpretations (generic/existential) and pronominal (anaphoric). See Percus 2000 for discussion of how this should be constrained.

[8] I will refer to them that way where a world is the maximal situation and a situation is a part of a world.

terpreted existentially if they are not a definite or kind. Bare nouns above this syntactic level are interpreted generically or as a definite. This can straightforwardly account for the combinations of generically and existentially bound variables that make up the distribution of characterizing and generic sentences. I am going to assume that all verbal arguments are introduced below negation, and that existential closure or something like it takes place below negation.

I will assume that a situation pronoun or variable first combines with the characteristic property, which is type $\langle s, \langle e, t\rangle\rangle$, and that this variable is bound either by the situation that is depicted by the clause or by the world variable associated with the proposition. In example (84), the situation variable w' binds the situation variable associated with the noun DOG, and the world variable w is related to w' by virtue of w' being part of w.

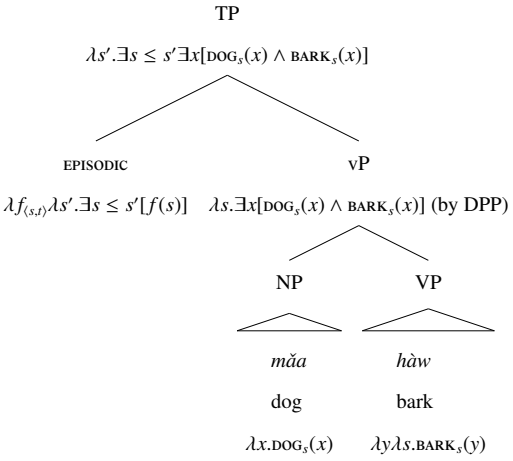
To be more precise, I am going to assume that something like existential closure takes places at the point where the bare nominal combines with a verb. This would take place before the nominal had combined with a situation variable, so the situation variable associated with the noun and verb must be the same. This could be carried out in your favorite way for predicative-type nouns to combine with verbs compositionally. I am going to use an altered version of Derived Kind Predication, which involves an intensional predicate rather than a kind, which I will call Derived Predicate Predication, shown in (85). This is preferable because it is no longer necessary to worry about whether a nominal is a typical kind or not. The derivation for an episodic sentence with an existential interpretation for *mă̆a* 'dog' can be seen in (86).

(85) *Derived Predicate Predication (DPP)*

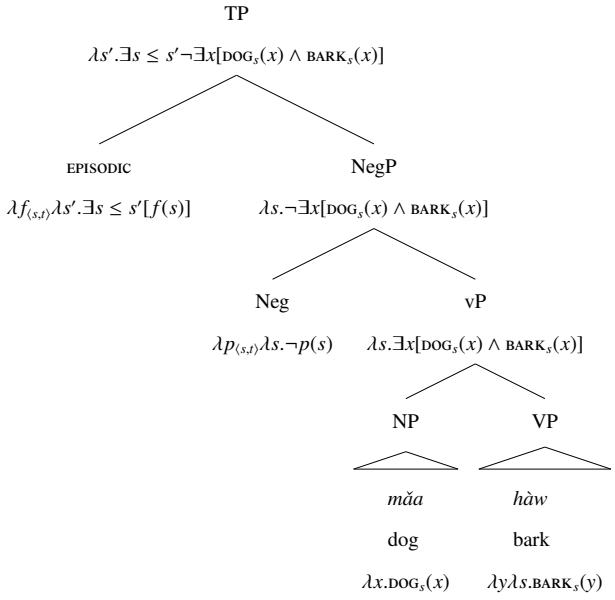
If P applies to objects of and Q denotes an intensional predicate of type $\langle s, \langle e, t\rangle\rangle$, then $P(Q) = \lambda w.\exists x[Q_w(x) \wedge P_w(x)]$

If there is negation in a sentence, the negation would be higher than the highest position where DPP could take place. This would lead to obligatory low scope with respect to negation for bare nouns. This derivation is shown in (87).

(86) EXISTENTIAL, EPISODIC

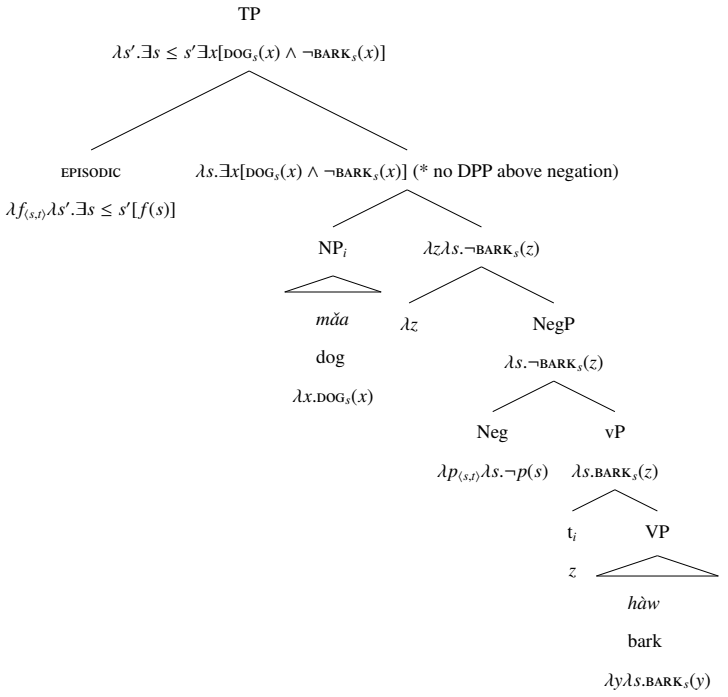


(87) EXISTENTIAL, EPISODIC, NEGATION



In order to get a high scope existential reading with respect to negation, it would be necessary for *mǎa* 'dog' in (88) to be interpreted at a position above negation. If the DPP operation (or existential closure) cannot take place above negation, it makes sense that the subject has to be interpreted below negation. This is why the reading in (88) is not possible with a bare noun, which cannot be interpreted as existential without DPP (or existential closure), but an expression with an indefinite determiner like *a* can be interpreted at this higher position since it does not rely on DPP for its indefinite interpretation.

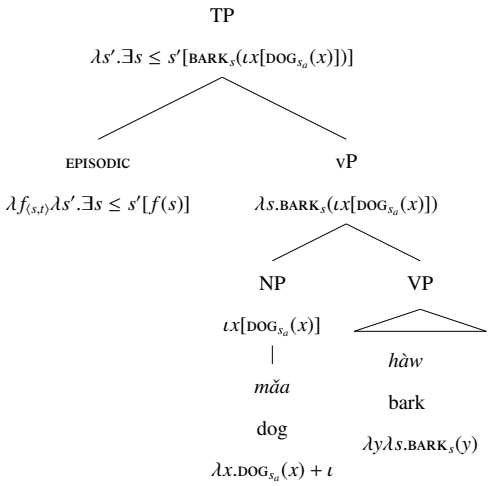
(88) EXISTENTIAL, EPISODIC, *∃ > ¬



Definite

An example with a definite interpretation of *măa* can be seen in (89). This would have a meaning similar to 'the dog barked'. If there is a familiar/salient dog in the context, it is possible to type-shift using ι to create something of type e that can combine with the verb. In this example, s_a specifies the reference situation for the specific dog that the speaker is referring to.

(89) EXISTENTIAL, DEFINITE



Kind

Then, I propose that a KIND interpretation is a world to individual mapping, rather than a situation to individual mapping.[9] For particular lexical items that take kinds as arguments, they take an argument of type $\langle s, e\rangle$. Then the situation variable is either universally quantified over or else evaluated with respect to the world. Characterizing sentences that seem to involve kinds have that interpretation because they involve universal/generic quantification over situations (Carlson & Pelletier 1995).

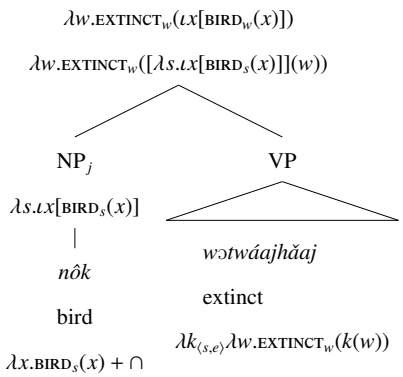
For a kind-verb like *extinct* in (90), it would take a kind as an argument, and then the world variable w associated with that verb would be applied to the kind to get the extension of the kind at a given world w. It seems like the underlying meaning of extinct should be related to a world and a reference time t, and say that at some time before t, this kind existed in the world w, but at the reference time the extension of the kind is empty. This possible semantics is given to the right of the $\rightsquigarrow$, but I will use $\lambda k\lambda w.\text{EXTINCT}_w(k(w))$ in the following derivations.

[9]This does not distinguish the role times versus worlds. It may well be that a kind should be a world/time pair to individual mapping.

(90) $[\![extinct]\!]^t = \lambda k \lambda w.\text{EXTINCT}_w(k(w)) \rightsquigarrow \lambda k \lambda w.\neg\text{EXIST}_w(k(w), t) \wedge \exists t' < t[\text{EXIST}_w(k(w), t')]$

For a kind interpretation, the $\langle s, \langle e, t\rangle\rangle$ noun type-shifts as it combines with the kind-verb. The derivation of a kind interpretation proceeds as in (91).

(91) KIND



This account is intended to combine accounts of generic semantics with a situation semantic account of domain restriction. This way we can characterize generic sentences as involving generic quantification over situations in the world and episodic sentences as existential quantification over a situation in the world.[10] If we are evaluating nominal expressions and sentences more generally with respect to some topic situation, we do not want to rule out the possibility of referring to a kind which includes reference to members of the kind that are not within the topic situation.

Things like attitude verbs would still combine with propositions. For example, *believe* in (93) is modeled after (92), abstracting away from how time is used.

(92) $[\![believe]\!]^{w,t} = \lambda p_{\langle s,t\rangle} \lambda x.p(w', t')$, for all worlds w' and t' such that for all that x can tell in w at t, x might be located in w' at t'. (Von Fintel & Heim 2011: (155))

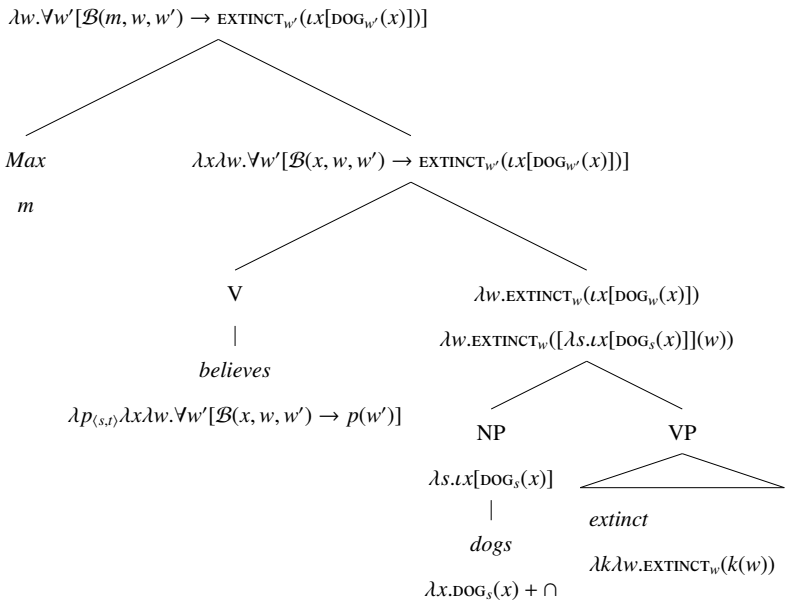
[10] It would probably be worthwhile to test whether this type of analysis can be integrated with topic situations as described by Schwarz (2009), among others.

(93) $[\![believe]\!] = \lambda p_{\langle s,t\rangle}\lambda x\lambda w.\forall w'[\mathcal{B}(x,w,w') \rightarrow p(w')]$

The worlds w such that for all w' compatible with x's beliefs in w (written $\mathcal{B}(x,w,w')$), the proposition p is true.

With a semantics for *extinct* it is possible to show how the kind interpretation interacts with attitude verbs like *believe*. This is shown in (94). Here, the kind *dogs* is evaluated with respect to Max's belief worlds rather than the worlds that define the proposition.

(94) EMBEDDED, KIND (*Max believes that dogs are extinct*)



2.3.5 Sub-kinds

(95) repeats example (26), which demonstrates that bare nouns in Shan cannot be used to refer anaphorically to a sub-kind of the noun. For example, the bare noun *nôk* 'bird' in (95a) cannot be used to refer anaphorically to the sub-kind *kâw* 'owl' mentioned in (95).

(91) above shows the interpretation of *nôk* 'bird' when it shifts using ∩. This meaning is incompatible with an anaphoric interpretation (Dayal 2011). It can only refer to all of bird kind.

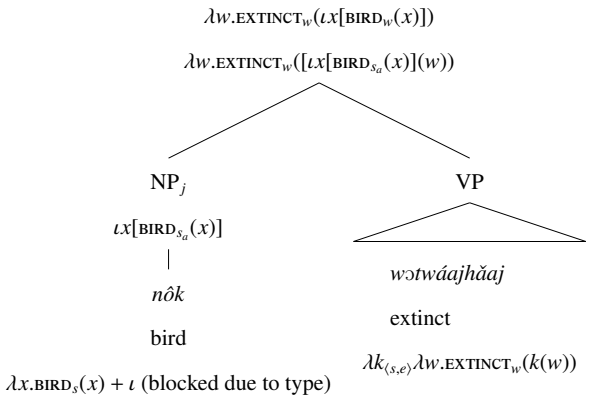
For this reason, (26a), repeated in (95a) is unacceptable. (96) shows the unacceptable derivation of a bare noun being used to refer anaphorically to a previously mentioned sub-kind. The verbal expression meaning 'extinct' is looking for an argument of type $\langle s, e \rangle$, so the bare noun will only shift using $\cap$. This is why the noun that names the sub-kind can be used—it can still type-shift using $\cap$. This can be seen in (26b) repeated in (95b).

(95) laat nɛ lɔŋ kâw lɛ sɤ̌-láajkàajkɔ̌n tsɔ̀ŋ laj khaa.
say show about owl and big.cat-bengal POLAR.Q can HUMB
'Could you tell me about owls and Bengal tigers?'

a. #sɤ̌ lɛp tě wɔtwáajhǎajkwàa jâw kójkaa nôk tɛ̂ jáŋ mí năm nàa náj lumfâa
big.cat likely IRR extinct PFV but bird CONTR still have many very in world
intended: 'The big cat will likely go extinct soon but the bird is still common in the world.'

b. sɤ̌-láajkàajkɔ̌n lɛp tě wɔtwáajhǎajkwàa jâw kójkaa kâw tɛ̂ jáŋ mí năm nàa náj lumfâa
big.cat-bengal likely IRR extinct PFV but owl CONTR still have many very in world
'Bengal tigers will likely go extinct soon but owls are still common in the world.'

(96) ANAPHORIC SUB-KIND; *NÔK



This type mismatch driven type-shifting also supports the idea that there is no null ι present in the syntax. We will see in section 2.4 that it is possible to refer to sub-kinds using N DEM

expressions, which is presumably type e. If there were a null ι in the syntax, we would expect the same process of allowing an e expression to refer to a sub-kind in N DEM expressions would be able to take place with (96).

Generic

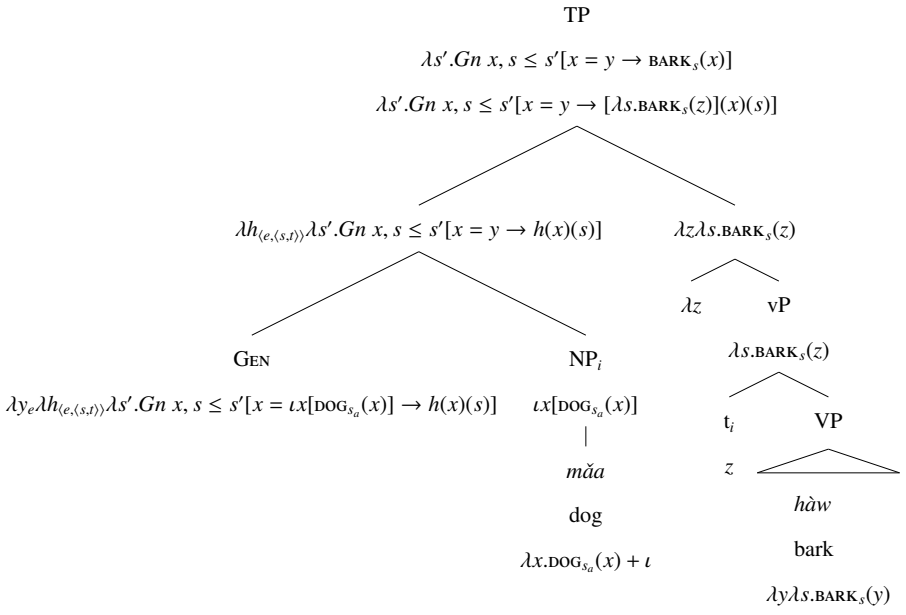
Generic quantification can take place for bare nouns above the level of existential closure. If we assume that generics involve generic quantification over situations and/or individuals, then characterizing sentences do not necessarily involve a kind interpretation at any point in the derivation. I am going to assume that generic interpretations involve either a definite or kind denoting nominal. By making this assumption, it may be easier to incorporate the semantics of determiner/demonstrative expressions which also can have both generic and kind interpretations. A possible semantics of generics is given in (97)–(98). There are generic sentences about individuals and about kinds, so the generic semantics would need to be compatible with either. (97) combines with a definite expression of type e and (98) combines with a kind of type $\langle s, e\rangle$.

(97) $[\![\text{GEN}_{entity}]\!] = \lambda y_e \lambda h_{\langle e,\langle s,t\rangle\rangle} \lambda s'.Gn\ x, s \leq s'[x = y \rightarrow h(x)(s)]$

(98) $[\![\text{GEN}_{kind}]\!] = \lambda k_{\langle s,e\rangle} \lambda h_{\langle e,\langle s,t\rangle\rangle} \lambda s' Gn\ x, s \leq s'.[x \leq k(s) \rightarrow h(x)(s)]$

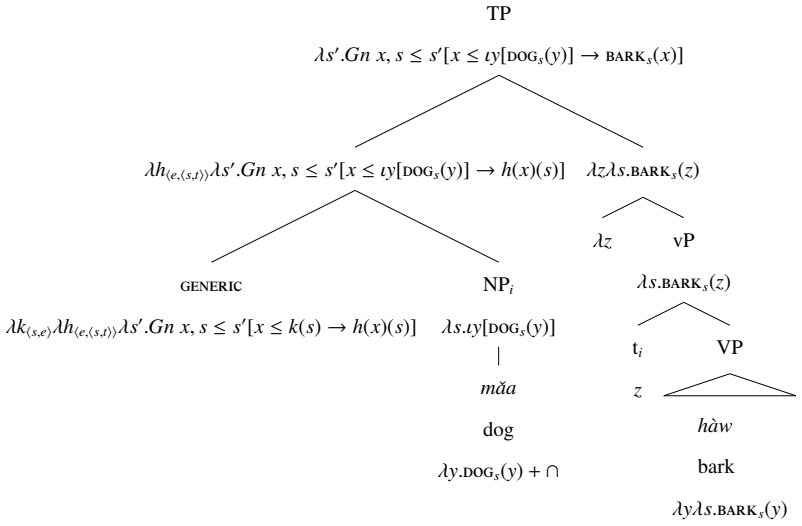
(99) is the derivation involving a definite using (97).

(99) GENERIC DEFINITE, *the dog barks*



(100) gives an example derivation involving a kind using (98). This sentence would have a meaning like 'dogs bark'. Here, instead of having existential quantification over situations, there is generic quantification over individuals and situations. The generic semantics requires a kind or definite. The $\langle s, \langle e, t\rangle\rangle$ expressions can type-shift to a kind. The generic operator quantifies over members of the kind, which at a given situation is the maximal extension of individuals within that kind.

(100) GENERIC KIND, *dogs bark*



Looking back at the two possible interpretations of the English sentence in (58), repeated in (101), it is possible to show how the two interpretations are derived. The derivation of (101a) is given in (102) and the derivation of (101b) is given in (103).

(101) Typhoons arise in this part of the Pacific. (Milsark 1974; Carlson 1989; Mendia 2020)

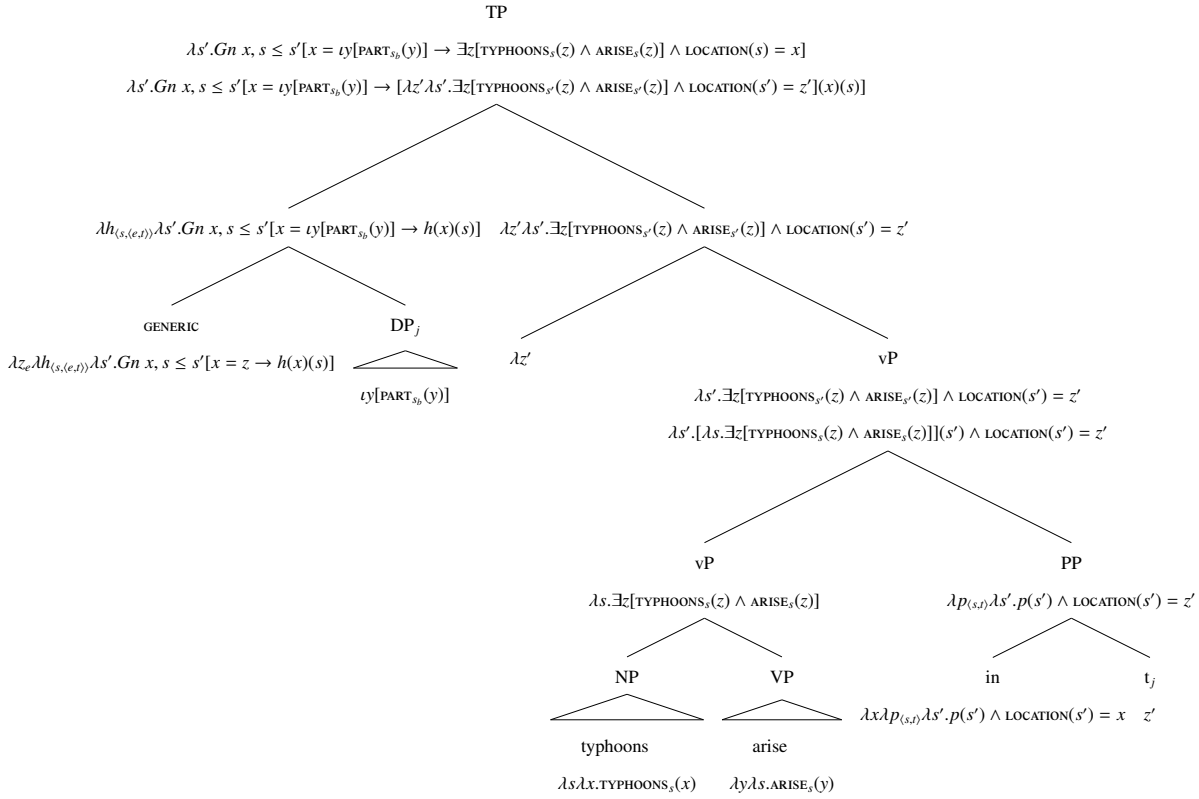
a. Object-referring interpretation:

   This part of the Pacific has the property that some typhoons arise in it.

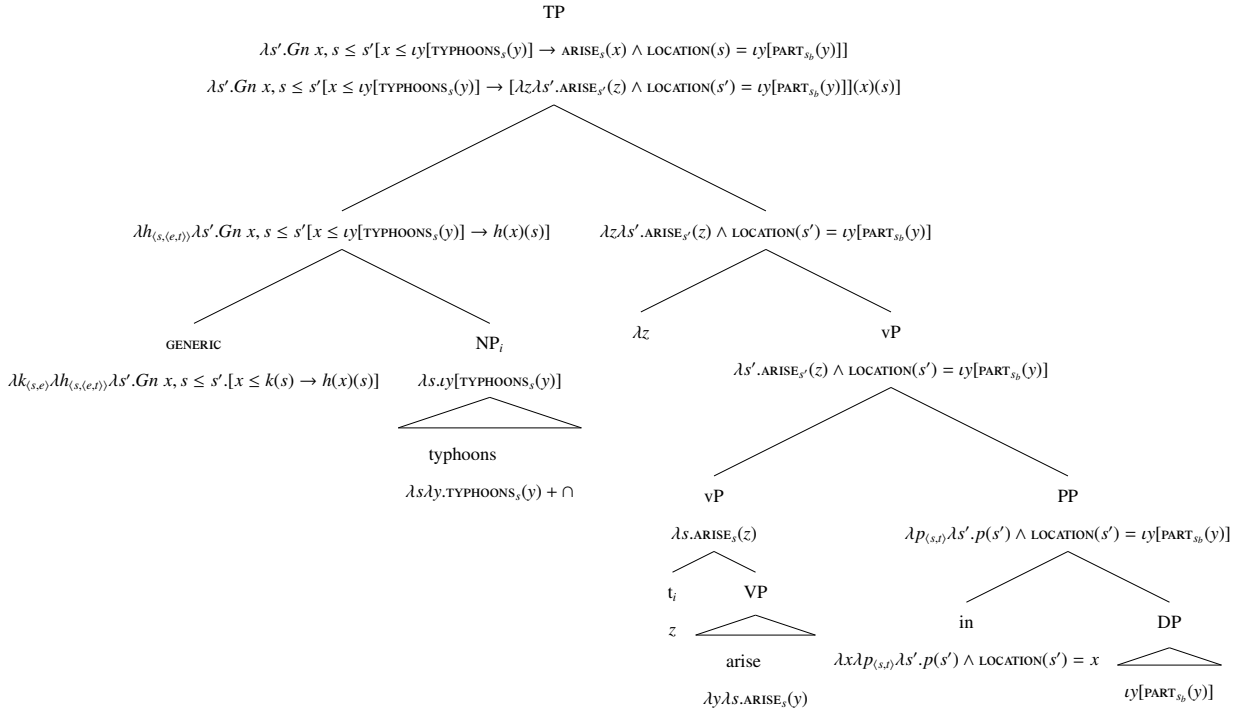
b. Kind-referring interpretation:

   It is a property of typhoons that they arise in this part of the Pacific.

(102) GENERIC: (101A)



(103) GENERIC: (101B)



2.3.6 Summary

This section has demonstrated how the bare noun interpretations that are evident in Shan and cross-linguistically could be derived from a noun of type $\langle s, \langle e, t \rangle \rangle$. I have argued that it is not necessary to assume an analysis where the base interpretation of a bare noun is the kind interpretation. Additionally, it would argue that it is simpler as it does not require multiple type-shifting steps to go from a kind interpretation to a predicate to an individual.

2.4 Demonstrative

Thus far, I have proposed a semantics of bare nouns. In this section, I take a preliminary look at the role of the demonstrative in generating the possible nominal interpretations. The demonstrative can appear in expressions with definite, kind, and generic interpretations.

2.4.1 Shan: N Dem

Definite

We have seen that N Clf Dem expressions, in addition to bare nouns, can have an anaphoric interpretation. It is also possible for N Dem expressions to be anaphoric, as (106) shows. (104)–(106) come from a story about a beekeeper. (104) is the second sentence of the story. (104)–(105) use a bare noun to refer back to the beekeeper that was introduced in the first sentence. (106) comes a few sentences later and has the N Dem expression *kón-lêŋ-phɯŋ nân* 'that beekeeper'.

(104) ʔɔ̀n-jíŋ ʔan nuŋ khó-tsùm sǐ-sɔ̀m nàn pěn luk-jíŋ kón-lêŋ-phɯŋ
small-female comp wear clothes-set color-blue that cop child-female person-raise-bees
The young girl who is wearing a blue dress is the daughter of the beekeeper.

(105) kón-lêŋ-phɯŋ ʔăw nâm-phɯŋ ti phɯŋ ʔăn mán-tsáaj lêŋ nân máa hét
person-raise-bees take water-bees from bees that 3-MASC raise that come make

khawmún-khêk
bread-cake

The beekeeper took honey from the bees which he takes care of and made cake.

(106) kón-lêŋ-phɯŋ nân laat waa jɔ̂npɤwaa mán thaw jâw
person-raise-bees that say SAY because 3 old PERF
The beekeeper said, 'Because he is old.'

In (107), the N DEM expression *khaw nâj* 'that rice' in the second sentence is used to anaphorically refer back to the rice mentioned in the first sentence. This example shows that mass nouns can have a definite interpretation when combined with a demonstrative.

(107) mɤnâj háw sûɯ khaw lɛ thò.
today 1 buy rice and bean

khaw nâj mí kaakhăn jàj nàa. thò tê̂ kaakhăn ʔàm pɔ́ jàj
rice this have price big very bean CONTR price NEG quite big

'Today I bought rice and beans. The rice was very expensive, but the beans were not expensive.'

Kind

When looking at kind and generic interpretations, there is an apparent difference between the proximal demonstrative *nâj* 'this' and the distal demonstrative *nân* 'that'. I would like to emphasize that this observed distinction needs further verification.

The noun-demonstrative expression *nôk nân* 'that bird'/'those birds' in (108) can refer to a previously mentioned sub-kind of birds. In contrast, *nôk tŏ nân* cannot refer to a kind, only to an individual, so it is infelicitous in (109). The classifier *tŏ* is only compatible with an individual interpretation. Classifiers will be discussed further in chapter 3.

(108) nôk nân wɔt.wáaj.hăaj kwàa jâw.
bird that disappear go finish
'Those birds are extinct.' KIND: ?, ANAPHORIC, SUB-KIND: ✓

(109) #nôk tŏ nân wɔt.wáaj.hăaj kwàa jâw.
bird CLF.ANML that disappear go finish
intended: 'That bird is extinct.' KIND: ✗, ANAPHORIC, SUB-KIND: ✗

Examples (110), repeated from (24), and (111) demonstrate the types of nominal expressions that can refer to kinds in object position. The bare noun *káa* 'car' in (24) can refer to a kind. However, when the noun is combined with a demonstrative, as in (111–(111), the nominal expression *káa nân* can only refer to a sub-type of car that is visible or was previously mentioned.

(110) Karl Benz pĕn kô khŏŋsaaŋ káa ʔɔ̆ntáaŋ pɤn
Karl Benz COP CLF.HUM invent car before others
'Karl Benz is the first person who invented the car.' KIND: ✓, SUB-KIND: ✗

(111) Karl Benz pĕn kô khŏŋsaaŋ káa nân ʔɔ̆ntáaŋ pɤn
Karl Benz COP CLF.HUM invent car that before others
'Karl Benz is the first person who invented that car.' KIND: ✗, SUB-KIND: ✓

There is some evidence that noun-demonstrative expressions in Shan can refer to kinds in some contexts. This seems to be limited to the proximal demonstrative *nâj* 'this'. For example, the first sentence in a written story about ants is given in (112). This is clearly a kind interpretation since it is talking about ants as a species. In this example, the nominal expression *môt nâj* is in a left-located position, and an overt pronoun *khăw* co-refers with that nominal expression.

(112) môt nâj, khăw mí khɤ́ năm têtê.
ant this 3.PL have species many really
'Ants, they include very many species.'

It is also possible to have N DEM expressions in object position with a kind interpretation, as (113) shows. Here, *măa (nâj)* 'this dog' refers to dogs as a species.

(113) mɤ nɯŋ mùn pǐ pôn máa nân kón ʔɔ̌nkăn lêŋ măa (nâj) ʔɔ̌ntáaŋ
when one ten.thousand year past come that people together raise dog (this) first

pɤn
others

'10,000 years ago, humans first took care of dogs.'

There is some speaker variation in whether the N with *nâj* 'this' is acceptable with a kind interpretation in all positions. For one speaker, the [N *nâj*] expression can only have a kind interpretation in a left-located position. For another speaker, [N *nâj*] can be in object position as well. They agree that [N *nâj*] can have a kind interpretation in some contexts, but there is more work to be done to determine precisely why they are accepted in certain positions and not others.

(114)–(115) set up a context for anaphoric reference to a sub-kind. From the acceptability of the possible follow ups in (116), it is possible to see which expressions can be used to refer to kinds/sub-kinds. (116a) shows that it is not possible to refer back to 'radio telegraph' as a sub-kind using *radio nâj* 'this radio' which includes the proximal demonstrative. This expression can only mean radio-kind, not a sub-type of radio. If there is the classifier *méw* used for kinds/sorts, as in (116b), it is possible to refer to the sub-kind with a N CLF DEM expression. It is also possible to use *radio nân* 'that radio' with the distal demonstrative to refer anaphorically to the sub-kind.

(114) Teacher: laat nɛ lɔŋ pʊ́n radio sĕméwméw tsɔ̀ŋ laj
say show about history radio some.sort POLAR.Q can
'Can you tell me about the history of some sort of radio.'

(115) Student A: mɤ tsâw pǐ hĕŋ kaw pàak nâj kón khŏŋsaaŋ radio
when early year thousand nine hundred this people invent radio

telegraph ʔɔ̌ntáaŋ sút
telegraph first completely

'In the early 1900s, people first invented the radio telegraph.'

(116) Student B:

a. #Marconi pĕn kɔ̂ khŏŋsaaŋ radio nâj ʔɔ̌ntáaŋ pɤn
Marconi COP CLF.HUM invent radio this first others
#'Marconi invented the radio before others.' (# if he just invented one kind of radio)

b. Marconi pěn kô khŏŋsaaŋ radio méw nâj ʔɔ̆ntáaŋ pɤn
Marconi COP CLF.HUM invent radio kind this first others
'Marconi invented this kind of radio before others.'

c. Marconi pěn kô khŏŋsaaŋ radio nân ʔɔ̆ntáaŋ pɤn
Marconi COP CLF.HUM invent radio that first others
'Marconi invented that radio before others.'

(117) includes a mass noun with a demonstrative, *khaw nâj* 'this rice', which still has a kind interpretation. (118) and (119) show that *khám nâj* 'that gold' can be used to refer to a kind in either subject or object position.

(117) mɤ̌ lěw khaw nâj hăa jàap nàa
now rice this find difficult very
'Now it is difficult to find rice.'

(118) khám (nâj) mí kaakhăn jàj tɛ̂tɛ̂
gold (this) have value big really
'Gold is really valuable.'

(119) haa hěŋ pǐ pôn máa nân kón tsâj khám (nâj) ʔɔ̆ntáaŋ sút
five thousand year past come that people use gold (this) first absolutely
'Humans first used gold about 5000 years ago.'

Sub-kinds

(120) sets up possible anaphoric reference to the sub-kind 'dodo bird'. (120a) shows that a bare noun describing the kind cannot be used to refer anaphorically to a previously mentioned sub-kind. Using the name of the sub-kind and a demonstrative, as in (120b), it is possible to refer back to the sub-kind. It is also possible to use a classifier for sorts/kinds *méw* with the noun and demonstrative, as in (120c), or a noun-demonstrative expression that includes the plural classifier, as in (120d). (121) similarly shows that *nôk tsɤ́ nân* with a plural morpheme can be used to refer to a sub-kind. The discourse referent it sets up in the first clause is compatible with the predicate meaning 'go extinct' in the second clause.

(120) háw hûtsák nôk dodo lɛ sɤ̌ kójkaa
1 know bird dodo and tiger but
'I know about dodo birds and tigers, but'

a. #nôk laj wɔt.wáaj.hăaj kwàa jâw.
bird ACH disappear go finish
intended: 'The bird is extinct.'

b. nôk dodo nân laj wɔt.wáaj.hăaj kwàa jâw.
bird dodo that ACH disappear go finish
'The dodo bird is extinct.'

c. nôk méw nân laj wɔt.wáaj.hăaj kwàa jâw.
bird kind that ACH disappear go finish
'That type of bird is extinct.'

d. nôk tsɤ́ nân laj wɔt.wáaj.hăaj kwàa jâw.
bird PL that ACH disappear go finish
Those birds are extinct.'

(121) nôk tsɤ́ nân láklɛ̆m nàa. kójkaa lɛp tĕ wɔt.wáaj.hăaj kwàa jâw.
bird PL that clever very but probably IRR disappear go finish
'Those birds are very clever but will likely go extinct.'

For expressions that have an overt demonstrative as in (120b), (120c), and (120d), we would expect to get a denotation of type e. Since the verb expects something of type $\langle s, e\rangle$, it must be possible to type-shift the expression to allow for a kind interpretation.

The licit derivations using overt demonstratives are shown in (120b)–(120d).[11] I follow the argument by Despić (2019) that says type-shifting using ι is blocked because type-shifting with $\cap$ is available.[12] Looking at (120b)–(120d), we see that type e expressions can be arguments of kind-denoting verbs. However, the type-shift that the type mismatch between a kind-denoting verb and its argument is $\cap$ not ι, so in the absence of an overt determiner or demonstrative, we would not expect a bare noun to have an anaphoric sub-kind interpretation.

[11] I have abstracted over the difference between the demonstrative and determiner her for simplicity, but see section 2.4.3 for discussion of that difference.

[12] This is why languages with singular bare nouns, which do not allow shifting with $\cap$, can refer anaphorically to sub-kinds. See Despić 2019 for detailed discussion.

Generic

(122) and (123) give examples of generic expressions that I would argue involve a kind interpretation. In (122), the subject *măa nâj* 'this dog' can either refer to a specific dog or to dogs in general. In a context where there is no specific dog (or dogs) that the speaker could be referring to, a generic interpretation is possible. In a context where the speaker could be referring to a specific dog, the individual reading is possible. It is also possible to have the non-specific interpretation when the N DEM expression is in object position, as with *môt nâj* 'this ant' in (123).[13]

(122) măa nâj kâj lɯp-lám hăaŋ tŏ-tsawkàw mán khɯ́n
dog this often chase tail body-self 3 back
'Dogs often chase their own tails.' KIND: GENERIC ✓

'This dog often chases and bites its own tail.' INDIVIDUAL (SG/PL): GENERIC ✓

(123) háw ʔàm lajtsăj môt (nâj)
1 NEG like ant (this)
'I do not like ants.'

(125)–(126) give examples of characterizing sentences, referring both to specific dogs and dogs in general. These include adverbs like *tɯ́n* and *tàasè* 'always' to indicate that this is an event that happens over and over. (124) sets up the context where someone hears one or more dogs barking and wonders why. (125)–(126) are possible responses to that, which depend on how many dogs are barking and whether they are discussing specific dogs or dogs in general. (125) refers to a specific group of dogs barking (perhaps the neighborhood dogs). This is a generic sentence about specific individuals.

In contrast, (126) refers to characteristics of dogs in general.[14] In (125), the anaphoric expression *măa tsɤ́ nân* 'those dogs' includes the distal demonstrative. The distal demonstrative is less likely to be associated with a kind interpretation since it tends to be used in anaphoric contexts.

[13]This sentence could presumably have an individual interpretation as well in the right context.

[14]It is possible for *nâj* to be used deictically in another context. Here, speakers only hear the barking, so the dogs are not close enough for the deictic usage.

The proximal demonstrative with *măa nâj* 'this dog' in (126) is more likely to be interpreted as a kind.

(124) măa hàw săŋ. tsɔ̀ŋ mí săŋ kɤ̀t khɯn
dog bark what POLAR.Q have what happen up
'Why are dogs barking. Did something happen?'

(125) ʔàm mí săŋ kɤ̀t khɯn. măa tsɤ́ nân kɔ tɯ́n hàw hét năn tàasè jâw
NEG have what happen up dog PL that PRT always bark do like.that always PFV
'Nothing happened. Those dogs always bark like that.'

(126) ʔàm mí săŋ kɤ̀t khɯn. măa nâj khăw tɯ́n hàw tsɤ̂ŋ-năn
NEG have what happen up dog this 3.PL always bark how-like.that
'Nothing happened. Dogs always bark like that.'

N DEM expressions can have be used to express generalizations about either a kind or an individual. There seems to be a tendency for the proximal demonstrative *nâj* 'this' to be compatible with kind interpretations in N DEM expressions. In contrast, the distal demonstrative *nân* 'that' tends to be limited to anaphoric interpretations, which can include anaphora to a sub-kind.

2.4.2 English: N DET

Having discussed the interpretations possible with N DEM expressions in Shan, I will briefly compare this distribution to the NOUN DETERMINER (N DET) expressions in English.

English bare nouns—either bare plurals or bare kinds and furniture nouns—can have a kind interpretation. However, when there is an overt definite determiner, it is quite unnatural in some cases, such as in (127).[15]

[15]Interestingly, anaphoric reference to a sub-kind is felicitous with a determiner plus mass noun as (1) demonstrates. Here, *the tea* refers anaphorically to the (sub-)kind *green tea*.

(1) I really like both green tea and dark-roast coffee, but the/that tea is difficult to find.

In (127), the mass noun *rice* cannot have a kind interpretation when it appears with a definite determiner. This is also true of *furniture* nouns, as in (128).

(127) #The rice was first discovered in 12,000BCE.

(128) #The furniture was first invented in 12,000BCE.

With count nouns, the data is more complicated. It is possible to have a kind interpretation with a singular N DET expression, as with *the dinosaur* in (129). However, there is some variation in whether a plural N DET expression can have a kind interpretation, as the contrast between (130) and (131) shows. The bare singular *the dog* is fine with a kind interpretation, as in (132).

(129) The dinosaur is extinct. (KIND: ✓, SUB-KIND: ✗)

(130) The dinosaurs are extinct.

(131) #The dogs were first domesticated more than 10,000 years ago.

(132) The dog was first domesticated more than 10,000 years ago.

This contrast might be due to the plural marking that is possible for English. Perhaps the grammaticality of (55) comes about as a result of the interpretation that the plural *the dinosaurs* refers to all the sub-kinds of dinosaurs (see Dayal 2004). Then, *the dinosaur* refers to dinosaur-kind. Then, the mass nouns and *furniture* nouns would not have this interpretation available because they can shift to a kind reading without the need for a determiner. I will come back to this distinction in the discussion of generics in the next section.

Generic

The following examples are all generic/characterizing sentences. All of the examples in this section with definite articles have a possible definite interpretation. For example, (133a) has a determiner

expression *the tigers* which could refer to specific tigers that the speaker has in mind. All the examples without definite articles cannot have a definite interpretation. The examples marked with '#' are those that cannot be used to describe properties of members of a kind. For example, (133a) cannot be used to say that members of tiger-kind generally eat meat.

The distribution seems to be that *the* with a plural or mass noun is not used to refer to members of a kind in these characterizing sentences, these examples are shown in (133a), (133c), (134a), and (135b).

(133) GENERIC: COUNT

a. #The tigers eat meat. (KIND: ✗, DEFINITE: ✓)

b. The tiger eats meat. (KIND: ✓, DEFINITE: ✓)

c. #The dogs often chase their own tails. (KIND: ✗, DEFINITE: ✓)

d. Dogs often chase their own tails. (KIND: ✓, DEFINITE: ✗)

(134) GENERIC: MASS

a. #The rice grows in paddy fields. (KIND: ✗, DEFINITE: ✓)

(135) GENERIC: *furniture* NOUNS

a. Furniture is expensive. (KIND: ✓, DEFINITE: ✗)

b. #The furniture is expensive. (KIND: ✗, DEFINITE: ✓)

Chierchia (1998) noted this distinction for plural nouns in English. However, he had said that the determiner + plural was bad for kinds and generics, but there are at least some kind interpretations that are available for the definite plural, as (130) shows. By contrast, Italian can have a determiner with both generic and kind examples, as shown in (136)–(137). Chierchia (1998) argued that in Italian nouns must project D, which is either overt or licensed by a lexical head. His explanation for the distinction between English and Italian is that English can type-shift at

the NP level, and Italian cannot. Then, there is an economy constraint against projecting D when type-shifting is available.

(136) I cani abbaiano.
the dogs bark
'Dogs bark.'

(Italian, Chierchia 1998: (90a))

(137) I cani sono diffusi.
the dogs are widespread
'Dogs are widespread.'

(Italian, Chierchia 1998: (91a))

This explanation is somewhat similar to the proposal by Despić (2019) that says type-shifting using ι is blocked because type-shifting with $\cap$ is available. Both of these proposals in combination would account for the English N DET data. However, it does not account for the fact that in Shan it is possible to have a kind interpretation with a demonstrative, even when the possibility of type-shifting is available. We would expect that a bare noun in Shan could have a kind interpretation by type-shifting, so it would not be expected that a demonstrative phrase would be possible since it is not necessary. It is worth looking at the English N DEM expressions for a clearer comparison with Shan.

With the proximal demonstratives *this* or *these*, generic kind readings are not available, but generic sub-kind readings are, as (138a)–(139b). These are different from the Shan data in that a whole-kind reading is unavailable. In the English cases, it seems that the taxonomic reading is what allows for the sub-kind reading of these examples.

(138) GENERIC: COUNT, DEM

a. These tigers eat meat. (SUB-KIND: ✓, DEFINITE: ✓)

b. This tiger eats meat. (SUB-KIND: ✓, DEFINITE: ✓)

c. These dogs often chase their own tails. (SUB-KIND: ✓, DEFINITE: ✓)

(139) GENERIC: MASS/*furniture* NOUNS

a. This rice grows in paddy fields. (SUB-KIND: ✓, DEFINITE: ✓)

b. This furniture is expensive. (SUB-KIND: ✓, DEFINITE: ✓)

The same is true of N DEM expressions in English when they appear with verbs that take kind denoting arguments, as in (140)–(145). Here the individual definite interpretation is not available because they are incompatible with this type of verb.

(140) This rice was first discovered in 12,000BCE. (SUB-KIND: ✓, DEFINITE: ✗)

(141) This furniture was first invented in 12,000BCE. (SUB-KIND: ✓, DEFINITE: ✗)

(142) This dinosaur is extinct. (SUB-KIND: ✓, DEFINITE: ✗)

(143) These dinosaurs are extinct. (SUB-KIND: ✓, DEFINITE: ✗)

(144) These dogs were first domesticated more than 10,000 years ago. (SUB-KIND: ✓, DEFINITE: ✗)

(145) This dog was first domesticated more than 10,000 years ago. (SUB-KIND: ✓, DEFINITE: ✗)

Looking at the distribution of possible interpretations of bare nouns, N DET expressions, and N DEM expressions in table 2.6, there is essentially complementary distribution between the possible bare noun and N DET interpretations.[16] The exception is with the sub-kind reading, which has a more limited distribution. In comparison, the N DEM expressions are much more freely available and do not depend on the interpretive possibilities of bare noun or N DET expressions.

Given that we are looking at the N DEM expressions in Shan, it is not surprising that their interpretative possibilities do not depend on the interpretations of bare nouns. The possible interpretations for these categories are given in table 2.7. This table only includes information about the proximal demonstrative *nâj* 'this', but it would be worthwhile to do a more extensive investigation of both the proximal and distal demonstratives.

[16] I have used the generic examples in (138a)–(139b) as the examples of both the definite and generic interpretations for N DEM expressions since these can be expressing generalities about an individual if it has a definite interpretation or about a kind, which I have categorized as a generic interpretation.

Table 2.6: English bare noun, N DET, and N DEM expressions

	English								
	N			**N Det**			**N Dem**		
	mass	count		mass	count		mass	count	
	mass	SG	PL	mass	SG	PL	mass	SG	PL
Def	✗	✗	✗	✓	✓	✓	✓	✓	✓
Ex.	(49)	*(51)	(52)	(50)	(53)	(54)	(139a)	(138b)	(138c)
Kind	✓	✗	✓	✗	✓	✗	✗	✗	✗
Ex.	(55)	*	(56)	(127)	(132)	(132)	(140)	(142)	(143)
Sub-kind	✗	✗	✗	✗	✗	✓	✓	✓	✓
Ex.	(55)	*	(56)	(127)	(129)	(131)	(140)	(142)	(143)
Generic	✓	✗	✓	✗	✓	✗	✓	✓	✓
Ex.	(59)	*	(60)	(134a)	(133b)	(133a)	(139a)	(138b)	(138c)

Table 2.7: Shan bare noun and N DEM (*nâj*) expressions

	Shan			
	N		**N Dem**	
	mass	count	mass	count
Def	✓	✓	✓	✓
Ex.	(33)	(105)	(107)	(106)
Kind	✓	✓	✓	✓
Ex.	(34)	(23)	(117)	(112)
Generic	✓	✓	✓	✓
Ex.	(36)	(27)	(118)	(122)

Another difference between English and Shan is that demonstrative-noun expressions in Shan can refer to a kind, but only English singular count nouns can refer to kinds when they combine with *the*. This can be explained if we assume that the determiner + singular kind interpretation generates a taxonomic reading in English, as proposed by Dayal (2004). The sub-kind interpretation is not available for mass and furniture nouns when they combine with determiners because they can already have a kind interpretation through type-shifting to a kind, negating the need for projecting a determiner in these contexts in English.

2.4.3 The role of the demonstrative

Dayal & Jiang (to appear) proposed the following semantics for demonstratives and definite determiners. (146a) has a presupposition of non-uniqueness in a wider domain. This does seem to be desirable for a N CLF DEM expression, which seems to presume non-uniqueness of the noun.[17] However, N DEM is not necessarily associated with non-uniqueness. To account for the fact that both count and mass nouns combined with demonstratives can refer to kinds, we do not want to include non-uniqueness in a widened domain as part of the demonstrative semantics. Therefore, I would propose that the demonstrative has an enriched version of the meanings of the definites in (146b)–(146c). This presupposes that the referent is unique in a given situation. If an individual with the salient property has been introduced previously, the strong definite (146b) is used; otherwise, (146c) is used. The enriched meaning is whether or not the intended referent is close to or far from the speaker.

(146) DEMONSTRATIVES AND DEFINITES (Dayal & Jiang to appear: (20))

a. $[\![\text{DEM}]\!] = \lambda s \lambda P : \exists s \leq s' |P_{s'}| > 1 . \iota x[P_s(x) \wedge x = y]$ *demonstrative*

b. $[\![\text{THE}_{\text{STRONG}}]\!] = \lambda s \lambda P : |P_s \cap \lambda x[x = y]| = 1 . \iota x[P_s(x) \wedge x = y]$ *strong definite*

c. $[\![\text{THE}_{\text{WEAK}}]\!] = \lambda s \lambda P : |P_s| = 1 . \iota x[P_s(x)]$ *weak definite*

There are two main demonstratives in Shan *nâj* 'this' and *nân* 'that'. I would argue that the demonstratives include an extra semantic component that says whether the referent is close to or far away from the speaker (this may be physical or a more abstract 'closeness'). This added semantic component is why the demonstrative in Shan does not function as a determiner, despite having a similar semantics to the English definite determiner. The semantics of the two Shan demonstratives given in (147) and (148).

(147) $[\![nâj]\!] = \lambda P : |P_s| = 1 . \iota x[P_s(x) \wedge \text{CLOSE.TO.SPEAKER}(x)]$

[17] This has not been demonstrated at this point, but it is mentioned in chapter 4.

(148) $[\![n\hat{a}n]\!] = \lambda P : |P_s| = 1.\iota x[P_s(x) \wedge \textsc{far.from.speaker}(x)]$

2.5 Chapter summary: Nouns

Bare nouns in Shan can have a variety of different interpretations depending on the context in which they appear. These nominals can be interpreted with definite, indefinite, kind, or generic semantics. I have found that Shan shares several properties with bare nouns in Dënesųłiné as described by Wilhelm (2008), (2014). Many of these semantic properties are also shared with English bare nouns.

If we assume that the bare noun structure is the reason for these possible interpretations—the generic and existential interpretations though existential closure and the definite and kind interpretations through type-shifting, we can account for the interpretations that are available in English and Shan (and likely other languages with bare nouns). The semantics of bare nouns in both English and Shan can be explained through existential/generic closure and type-shifting, where type-shifting is constrained by the determiners available in a language, while closure operations are not so constrained. If we separate these two possibilities we can account for many of the discrepancies between these two interpretations. We find that the existential and generic interpretations are quite ubiquitous with bare nouns, but the definite interpretations are far more limited.

From this we can see that there is a great deal that can be explained through a cross-linguistic view of nominal semantics. Certain interpretations seem to be cross-linguistically available, but certain interpretations are constrained by nominal morphology in a language. For example, we can explain the impossibility of having a sub-kind interpretation for a bare noun through the lack of plural morphology in Shan. We can also say that English bare nouns cannot be interpreted as definite expressions because English has overt definite determiners that block type-shifting.

I have proposed that it is not necessary to assume that bare nouns in languages like Shan are

kinds, contra Chierchia (1998). Instead, I argue that the kind interpretation comes about as a result of a nominal being interpretative with respect to the broadest possible situation: the world in which the situation described by the nominal takes place. It also can be used to explain why nominals in argument position in episodic sentences that do not involve a kind-denoting verb cannot be interpreted as a kind.

These themes will be developed further in the following chapters. The next chapter adds a level of complication to the interpretation of bare nouns by incorporating classifiers. The classifier is connected to the number neutrality of languages like Shan.

CHAPTER 3

CLASSIFIERS

DEFINITENESS AND QUANTIFICATION: EVIDENCE FROM SHAN

Mary Rose Moroney, Ph.D.

Cornell University 2021

This chapter examines classifier constructions in Shan, which includes classifiers for individuals, groups, measure expressions, and containers. The classifiers form a syntactically unified class—i.e. they have the same distribution. However, classifiers vary in the degree to which each kind has a separate noun function: some have an independent noun use, some can be head nouns in N-N compounds, and some cannot be either. It is also possible to have classifiers appear as part of predicates, particularly stubbornly distributive predicates. In these contexts, the classifiers specify the level to which the adjective distributes. This is especially notable when stubbornly distributive predicates combine with collectives. The semantics proposed for the classifier is unified across the categories of classifier and measure/count semantics, contra Rothstein (2016); although the particular semantics for a classifier varies based on the noun it is derived from. Instead, I argue that the property that is essential to counting and measuring is identifiable atomic parts. The classifier serves to quantize a noun either by making use of the count noun's inherent atomic parts or by imposing atomic parts on mass nouns by means of a container or measure classifier. Furthermore, evidence involving plural and group classifiers demonstrates that there is a separate syntactic position that is used for classifiers that denote non-singular terms.

3.1 Introduction

This chapter explores classifier constructions and proposes a syntactic and semantic analysis for numeral-classifier expressions in Shan that can give insight into these constructions cross-linguistically. The closest correlates that English has for classifiers can be seen in (149)–(152). In (149)–(152) *pieces* and *head* are functioning as 'numeral-classifiers'.

(149) I bought three pieces of furniture.

(150) *I bought three furniture.

(151) I bought three head of cattle.

(152) *I bought three cattle.

English does not allow words like *furniture* and *cattle* to combine with numerals unless there is another noun intervening. This property is connected to the noun property of CUMULATIVITY, defined in (76) in chapter 2, repeated in (153). The property of cumulativity is one that English bare plurals, mass, and *furniture* nouns and Shan bare nouns all share. Essentially, cumulative predicates are ones that include individuals that consist of sums of other individuals with that property.

(153) CUMULATIVITY (adapted from Deal 2017: (6))

A predicate *P* is cumulative iff any sum of parts that are *P* is also *P*.

For some languages, such as Mandarin Chinese, all nouns are like *furniture* and *cattle* in that a classifier is required when the noun combines with a numeral (Krifka 1995; Chierchia 1998; a.o.). I will call these 'numeral-classifier languages'. Shan is a numeral-classifier language. As (65)—repeated in (154) below—shows, the classifier *tŏ* is obligatory in the presence of a numeral.

(154) măa săam *(tŏ)
dog three *(CLF.ANML)
'three dogs'

In this chapter, I argue that classifiers in Shan and classifier-like nouns, like we see in (149)–(152) in English, are obligatory with numeral expressions because the cumulative predicates lack

identifiable countable units. Therefore, the role of classifiers is to provide a unit by which the noun can be counted. This is consistent with previous analyses of classifiers, such as Nomoto 2013. Shan, Mandarin Chinese, and English provide examples of languages with very grammaticalized (Shan) to un-grammaticalized (English) classifier expressions.

Section 3.2 gives background on the different types of numeral-classifiers that this chapter focuses on. Section 3.3 proposes a structure for classifier expressions in Shan. Section 3.4 provides a semantic analysis of classifiers. Section 3.5 proposes a unified analysis for both counting and measuring constructions. Section 3.6 concludes this chapter.

3.2 Shan classifiers

Before analyzing Shan classifiers, I will provide background on the main types of classifiers found in Shan. 'Numeral-classifiers' are not a homogeneous class, even just within Shan. Table 3.1 gives some of the categories of classifiers.[1] The category 'individual classifiers' is likely the most unfamiliar to English speakers. The others can be found in constructions similar to (149) and (151).

Table 3.1: Based on Chao's (1968) classification of nominal measures as cited by Li (2011)

	Measures	Mandarin Examples
A	Individual	*ge* (general classifier), *ben* 'volume', *tou* 'head'...
B	Group	*qun* 'group', *lie* 'series', *kun* 'bundle', *zhong* 'kind' ...
C	Partition	*pian* 'piece', *jie* 'section', *si* 'slice', *di* 'drop' ...
D	Container	*he* 'box', *hu* 'kettle', *wan* 'bowl', *shao* 'spoon' ...
E	Temporary	*shen* 'body', *lian* 'face', *di* 'floor'...
F	Standard	*mi* 'meter', *shen* 'liter' ...

While classifiers in Thai, a related Tai language, have been investigated in detail (e.g., Iwasaki & Ingkaphirom 2005; Piriyawiboon 2010; Jenks 2011), there have been very few descriptions or analyses of Shan numeral classifiers. Cushing (1887) first identified classifiers in Shan, calling

[1] See also Rothstein 2017: ch. 8 p. 195 for more categories of classifiers.

them ‘numeral auxiliaries’, which denotes “some rank of being, some form of object or some quality in the noun to which it belongs”.

In some languages of Southeast Asia, ‘noun classifiers’, which appear in nominal expressions that do not necessarily include numericals, are considered a type of numeral classifier (Aikhenvald 2000). This is relevant to Shan, since the same classifiers that appear in numerical expressions also appear in contexts that do not include numericals. In fact, individual classifiers, measure expressions, containers, and groups in Shan can appear in all of the positions shown in (155). The classifier is only obligatory in (155a) with the numeral, but the classifier does affect the semantics of the nominal expressions in all cases.

(155) a. N Num Clf
b. N (Clf) Dem
c. N (Clf) Adj
d. N (Clf) Rel-CP

The following sections will introduce several of the measure constructions from table 3.1 as they appear in Shan. The distinct properties that these expressions have can give us insight into the function of classifiers.

3.2.1 Individual classifiers

This sub-section discusses individual classifiers and introduces the basic properties of classifiers. Cushing (1887) only discusses the appearance of classifiers with numerals, but as shown below, classifiers also can appear in a variety of constructions, including with quantifiers (156), demonstratives (157), and relative clauses (158). It is also possible for a classifier to appear with certain adjectives, as shown in (159).[2]

[2]The word for ‘big’ can be pronounced *j*à*j* or *jà*ɯ.

(156) mǎa ku tǒ
dog every CLF
'every dog' (N QUANT CLF)

(157) mǎa tǒ nâj
dog CLF DEM
'this dog' (N CLF DEM)

(158) mǎa tǒ [RC ʔǎn nɔ́n jù]
dog CLF COMP sleep IPFV
'the dog that is sleeping' ' (N CLF RC)

(159) mǎa tǒ jàɯ
dog CLF big
'big dog' (N CLF ADJ)

(160) mǎa sǎam *(tǒ)
dog three *(CLF)
'three dogs' (N NUM CLF)

Shan numericals must appear either with a measure term or a classifier, as shown in (154) repeated in (160). In many languages, classifiers semantics are associated with singularity (Nomoto 2013, a.o.). In Shan, N CLF DEM constructions like (157) is obligatorily singular at the unit described by the classifier. Classifiers can appear with certain types of adjectives, as in (159). Typically, these are adjectives that pertain to the physical body and do not significantly affect the interpretation of the nominal other than to specify what 'unit' is big. Classifiers that appear with relative clauses like (158) either induce a singular interpretation or have little effect on the nominal interpretation, depending on the context.

The classifiers in (156)–(160) can appear in the same nominal but with a much more restricted distribution. A more complex example is given in in (161), schematized in (162). The individual classifier can appear in two positions simultaneously, with the adjective and with one of the other structures. It will be argued in section 3.3.3, that this is because the classifier-adjective construction has different properties than the other constructions. For the other three constructions, the individual classifier cannot appear in more than one place at the same time.

(161) mǎa tǒ jàj sǎam tǒ tsɤ́ ʔǎn lâk kǐn kàj (*tsɤ́) nân ...
dog CLF.ANML big three CLF.ANML PL COMP steal eat chicken PL that
'The three big dogs that stole and ate the chicken...'

(162) N [Clf Adj] [Num Clf] [Clf.Pl Rel-CP] Dem

As (161) shows, there is a confounding aspect to this generalization: the morpheme *tsɤ́*, which I am calling a plural classifier, can co-occur with the numeral classifier *tǒ*. (163) shows an example where a noun is found with the numeral-classifier structure *săam tǒ* from (160), which is then followed by the classifier-relative clause structure with the plural classifier, *tsɤ́*, which is ended by the demonstrative *nân* 'that'.

(163) mɤnâj háw hăn măa săam tǒ tsɤ́ ʔăn khóp mɛ́w mɤwàa nân
today 1 see dog three CLF.ANML PL COMP bite cat yesterday that
'Today, I saw those three dogs that bit cats yesterday.'

(164) includes the classifier-adjective construction *tǒ jàj*, the numeral-classifier construction *săam tǒ*, and the classifier demonstrative *tsɤ́ nân*, which includes the plural morpheme.

(164) mɤnâj háw hăn măa tǒ jàj săam tǒ tsɤ́ nân
today 1 see dog CLF.ANML big three CLF.ANML PL that
'Today, I saw those three big dogs.'

I will argue that the plural classifier is essentially a plural version of a classifier with a distinct syntactic projection. The plural classifier will be discussed in section 3.4.4.

Nominal properties of classifiers

The function of a classifier that I will propose essentially turns a cumulative noun, which contains both individuals and sums of individuals in its denotation, to a quantized noun, which consist of individuals that do not contain proper individual sub-parts in its denotation. This means that the function of the classifier is to generate an atomic set from a non-atomic set, essentially creating a singular noun from a plural noun. If we think about the role of the classifier-like expressions that exist in English, this is the same thing that the quantized count noun is doing to the cumulative mass or *furniture* noun, (165)–(166). Therefore, it is not surprising that there is some overlap

between classifiers in Shan and Chinese and expressions like *cups* and *pieces* in English when they function as classifiers.

(165) I drank three cups of water.

(166) I bought three pieces of furniture.

According to Li (2011), the classifiers in Chinese are largely borrowed from nouns, but some have lost nominal properties as they have been grammaticalized as classifiers. On the basis of this, Li (2011) claims that classifiers in Chinese are not nouns but have a separate function projection Classifier Phrase; whereas classifiers in non-classifier languages like English have the status of nouns. Shan slightly complicates this picture since Shan classifiers seem to be even more noun-like than those in Chinese. Even frequently used classifiers like *tǒ* for animals and *kɔ̂* for humans can be used as a morpheme in noun compounds. Shan uses repeater classifiers, where a noun can function both as a noun and a classifier. Nouns that can be used as classifiers often use repeater classifiers when the noun is combined with a numeral. Repeater classifiers are also found in Thai, Burmese and Lahu (Simpson 2005: p. 20). The clear presence of separate classifier and noun morphemes in these structures still supports the presence of a separate Classifier Phrase in the syntax of some languages with more nominal-like classifiers.

Shan classifiers appear to be derived from nominal elements. For some nouns that are typically found as compounds, there is a connection between the form of the classifier and one part of the compound, as in (167a). Here the noun compound *ton-mâj* 'tree' and the classifier *ton* both contain *ton*. I have not surveyed all of the classifiers extensively on this feature, but those that I have checked can either stand as nouns independently or are found as a morpheme in compound nouns. As mentioned before, classifiers obligatorily appear in the presence of a numeral, as shown in (167b). Which classifier appears depends on properties of the nominal referent. Table 3.2 gives some examples of properties that determine which classifier is used.

(167) a. ton-mâj săam *(ton)
plant-tree three *(plant)
'three trees'

b. măa săam *(tǒ)
dog three *(CLF)
'three dogs'

Table 3.2: Basic numeral classifiers in Shan

Form	Used to count	Examples	Translation
ʔǎn	inanimate things	*tsɔ̂ sǎam ʔǎn*	'three forks'
tǒ	animals	*mɛ́w sǎam tǒ*	'three cats'
kɔ̂	people	*kón sǎam kɔ̂*	'three people'
hòj	round objects	*màak-khɔ̌ sǎam hòj*	'three jujube'
ton	plants, trees	*ton-mâj sǎam ton*	'three trees'
lǎŋ	roofs	*hɤ́n sǎam lǎŋ*	'three houses'

The classifier used for animals, *tǒ*, also means 'body'. (168) is used in a context where you are talking to someone on a radio call and the voice is located in one place and the person in another.

(168) tǒ sǔ jù lǎj
body 2 stay which
'Where is your body?' (i.e., Where are you located?)

In Shan the classifier includes important semantic content, such as the type of fire being built in (169) and (170).

(169) tsû fáj sɔ̌ŋ tǎw
build fire two stove
'Build two fires.' (for cooking on a stove)

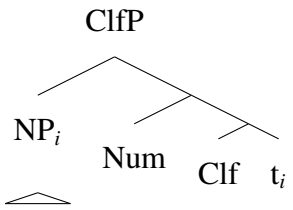
(170) tsû fáj sɔ̌ŋ mûk
build fire two fireplace
'Build two fires.' (for heating the house)

Li (2011) also argues that Mandarin Chinese classifiers are not nouns because they cannot be modified by adjectives outside of a few exceptional cases. Shan classifiers can be modified in constructions such as (159) above, but as I will argue later, this is a different use than the classifier that appears with numerals and demonstratives.

Structure

The basic word order of the extended nominal domain in Shan is Noun-Numeral-Classifier. Simpson (2005) proposed that this word order in Thai, another Southwestern Tai language, comes as a result of the noun moving from its base position to a position above the numeral and classifier. This is the same structure argued for by Jenks (2011) for Thai. I assume the same movement happens in Shan, as schematized in (171).

(171)



Word order can, however, vary depending on what combines with the noun. When quantification is involved, such as when the noun combines with a numeral or quantifier, the classifier appears after the noun and quantificational material, as in (172a). When the classifier appears with a demonstrative or relative clause, the classifier comes between the noun and other material, as shown in (172b)-(172d).

(172) a. N NUM CLF

b. N (CLF) DEM

c. N (CLF) ADJ

d. N (CLF) REL-CP

One issue related to the constituent structure of phrases with numerals and classifiers is that the numeral and classifier can be separated from their associated noun, as in (173). Here the noun *phujíŋ* 'women' is separated from the numeral and classifier *sǎam kɔ̂*. The numeral-classifier expression is one of the few (if not only) things that can be separated from the noun. This will be discussed more in section 3.3.2.

(173) phujíŋ hɔtphɛ̌w kéŋmàj jâw sǎam kô
woman arrive Chiang_Mai PFV three CLF.HUM
'Three of the women arrived in Chiang Mai.'

This will be our starting point for the syntax of classifiers, but it may be revised later on.

3.2.2 Measure terms

Measure terms are words that describe standard units used in measurement, such as the ones in table 3.3. (174) gives a typical construction with measure terms. They appear in the same position as individual classifiers in numeral-classifier constructions.

Table 3.3: Basic measure terms in Shan

Form	Used to count	Examples	Translation
lît	liters	*nâm sǎam lît*	'three liters of water'
kìlò	kilograms	*màak-khɔ̌ sǎam kìlò*	'three kilos of jujube'
pè	measures of grain	*khaw sǎam pè*	'three measures of rice'
wàat	money	(ŋʊ́*n*) *sǎam wàat*	'three baht' (Thai money)

(174) mɤnâj háw kǐn màakkhɔ̌ sǎam kìlò
today 1 eat jujube three kilo
'Today, I ate three kilos of jujube.'

Normally, measure terms are not compatible with referential interpretations. According to Chierchia (1998) and Li (2011), measure words do not impose 'atomic structure'. This, I believe, is due to the fact that measures do not tend to have a physical existence in the world. As (175)–(177) show, it is possible grammatically to use measure terms in referential constructions as long as the context supports the use. In this case, having the jujube separated into 1 kilogram portions allows for this referential interpretation.

(175) hàw tě sʊ̂ʊ̀ màakkhɔ̌ kìlò nâj
1 IRR buy jujube kilo this
'I will buy this kilo of jujube.'

(176) hàw tě sûɪ màakkhɔ̌ ku kìlò ti lâan
1 IRR buy jujube every kilo at store
'I will buy every kilo of jujube in the store.'

(177) hàw tě sûɪ màakkhɔ̌ kìlò ʔǎn súk jâw nân kój
1 IRR buy jujube kilo COMP ripe PFV that only
'I will buy only the kilos of jujube that are already ripe.'

The measure terms function differently than other classifier terms in that the measure term cannot appear as part of a noun compound in a N Num Clf phrase, as shown in the contrast between (178) and (179). Measure terms cannot refer to a particular object, except in denoting a unit of something as describe above.

(178) háw laj sûɪ khaw sǎam pè
1 ACC buy rice three pe
'I bought three *pe* of rice'

(179) *háw laj sûɪ pè khaw sǎam pè
1 ACC buy pe rice three pe
intended: 'I bought three *pe* of rice'

Measure terms can otherwise appear in the same syntactic positions as individual classifiers. The measure terms are different in that, unlike other classifiers, they do not have a distinct nominal use: they cannot be used to refer.

3.2.3 Group terms

Chao (1968) and Li (2011) separate the classifier/measure terms into several categories. I will initially follow Chao's (1968) classification and describe the group terms separately, but later we will revisit how it patterns with respect to other categories. Two of the main group terms used in Shan are given in table 3.4.

Table 3.4: Basic group terms in Shan

Form	Used to count	Examples	Translation
phŭŋ	group (animal)	*măa săam phŭŋ*	'three groups of dogs'
tsúm	group	*kón săam tsúm*	'three groups of people'

As (180) demonstrates, the word *tsúm*, used for groups of humans, can also function as part of a noun compound, *tsúm lukhén* 'group of students' or 'student group'. The numeral-classifier use of *tsúm* is with the question-numerical *lăaj tsúm* 'how many groups'? (181) also shows both the noun and numeral-classifier uses of *tsúm*. Here, the noun and numeral-classifier both appear after the verb *mí* 'have'. This is common for existential constructions.

(180) ʔó nâp tój lú tsúm lukhén qăn khaj laat tɔ́ háw nân mí lăaj tsúm
Oh! count see PRT group student COMP want talk to 1 that have how_many group
'Oh, count how many groups of students there are that want to talk to me.'

(181) ti tsântsɔ̆m háw nâj mí tsúm lukhén lɛ tsúm khúsɔ̆n lăajlăaj tsúm
at university 1 this have group student and group teacher many group
'At our university there are many student groups and teacher groups.'

(182)–(184) demonstrate that with the construction *phɛ̆ pĕn* 'split into' the correct form of the constituent following the construction is Num Group (*sɔ̆ŋ tsúm*) rather than N Num Group (*tsúm sɔ̆ŋ tsúm*).

(182) háw ʔăw tsúm nâj phɛ̆ pĕn sɔ̆ŋ tsúm lɨ̌ hâa
1 take group this split COP two group good POLAR.Q
'I split the group into two groups, is that good?'

(183) tsúm háw lò phɛ̆ pĕn sɔ̆ŋ tsúm
group 1 need split COP two group
'Our group needs to split into two groups.'

(184) *tsúm háw lò phɛ̆ pĕn tsúm sɔ̆ŋ tsúm
group 1 need split COP group two group
intended: 'Our group needs to split into two groups'

However, the data patterns in (183)–(184) may instead be due, in part, to a feature of certain classifiers. The word meaning 'group' and the classifier for counting groups is the same word.

Simpson (2005) identified these as 'self-classifiers' or 'repeaters' in Thai, Burmese, and Lahu. I will call such words 'repeater classifiers'. As the contrast between (185) and (186) shows, it is not natural to include repeater classifiers twice in numeral-classifier expressions. If 'R' represents the homophonous noun and classifier, the structure Num R is used rather than R Num R.

(185) #mɤnâj háw laj laat tɔ̀ tsúm sɔ̌ŋ tsúm
today 1 ACH speak to group two group
intended: Today, I spoke with two groups.

(186) mɤnâj háw laj laat tɔ̀ sɔ̌ŋ tsúm
today 1 ACH speak to two group
'Today, I spoke with two groups.'

These classifiers can be used with other nouns, as shown in (187), where the noun is *kón* 'person'.

(187) mɤnâj háw laj laat tɔ̀ kón sɔ̌ŋ tsúm
today 1 ACH speak to people two group
'Today, I spoke with two groups of people.'

If the repeater noun is combined with an adjective or another noun, the noun appears in the usual position. This is shown in (188), where the compound noun *tsúm lukhén* 'student group' is followed by the numeral-classifier expression.

(188) mɤnâj háw laj laat tɔ̀ tsúm lukhén sɔ̌ŋ tsúm
today 1 ACH speak to group student two group
'Today, I spoke with two groups of students.'

Group terms can be used in all the same positions as individual classifiers represented in (155) above. Group terms are more like individual classifiers than measure terms in having a nominal use. They are unlike individual classifiers in that they can appear in a repeater classifier construction.

3.2.4 Containers

Shan container terms, like group terms, are homophonous with the noun they are related to. They appear in the same position as other numeral classifiers, as shown in (189). Table 3.5 lists a few container terms.[3]

(189) nâm sǎam kɔ́k
water three cup
'three cups of water' (Shan)

Table 3.5: Basic container terms in Shan

Form	Used to count	Examples	Translation
kók	cups	*nâm sǎam kók*	'three cups of water'
taw	bottles	*nâm sǎam taw*	'three bottles of water'
thǒŋ	bags	*khaw sǎam thǒŋ*	'three bags of rice'
jɔ̀t	drop	*nâm sǎam jɔ̀t*	'three drops of water'
kǒŋ	piles	*màak-khɔ̌ sǎam kǒŋ*	'three piles of jujube'

An important question that arises for container constructions is whether the container or the contents are being referred to in a given construction. This is similar to the group terms in the previous section except that for group terms the group consists of the individual parts and for container terms the container and the contents are separate entities.

For example, (189) describes a measure of water, rather than bottles of water because the head noun is *nâm* 'water' not *taw* 'bottle'. This can be seen in the infelicity of (190) in out-of-the-blue contexts.[4] It is strange to refer to two-bottles worth of water when talking about possessing something because they would normally be in bottles or another container.

[3] I've included *kǒŋ* 'pile' and *jɔ̀t* 'drop' in this table, but it is possible that they more properly belong in Chao's (1968) PARTITION category.

[4] These distinctions can be somewhat difficult to tease apart. For the moment, I am assuming that when the initial word in the string is a noun referring to the container, reference is to the container and when the initial word is the contents, reference is to the contents. There might be a more nuanced CONTAINER+CONTENTS reading but that requires further investigation.

(190) #mí nâm sɔ̌ŋ taw
have water two bottle
intended: ‘(I) have two bottles of water.’

(191) mí taw nâm sɔ̌ŋ taw
have bottle water two bottle
‘(I) have two bottles of water.’

(192) demonstrates that the measure use of water is felicitous in certain contexts, such as when describing the amount of water that one has drunk. When using numerals to describe containers themselves, the same word is used in the position of the noun and the classifier. This can be seen in (193). With container expressions, the word in the typical noun position is often optional as it was for group terms. This is discussed further in section 3.3.1 in reference to repeater classifiers.

(192) háw kǐn nâm sɔ̌ŋ taw
1 drink water two bottle
‘I drank two bottles of water.’

(193) háw mí taw sɔ̌ŋ taw
1 have water two bottle
‘I have two (empty) bottles.’

As with other classifier terms, it is possible to separate the noun from the numeral-classifier expression, as shown in (194), where *khawmún* ‘cake’ the contents of the bag, is separate from Num Clf sɔ̌ŋ thŏŋ ‘two bags’.

(194) mɤnâj háw sûı khawmún máa sɔ̌ŋ thŏŋ
today 1 buy cake come two bag
‘I bought two bags of sweets today.’

When the action pertaining to the noun-numeral-classifier expression affects the container, the container noun also appears to form a compound with the contents as with *thŏŋ khawmún* ‘bags of cake’ in (195).

(195) mɤnâj háw ʔàm laj táŋtsǎj sějâwlɛ ʔǎw thŏŋ khawmún tɛ̀k pɛt sɔ̌ŋ thŏŋ
today 1 NEG ACH intend CONJ take bag cake break DETR two bag
'Today, I accidentally crushed two bags of sweets.'

Container terms allow overt repeater classifier constructions, as shown in (196).

(196) ʔó mɤnâj háw tsǎjtsâa nàa. háw lúm ʔǎw thŏŋ sàj khɔ̌ŋ máa kɔpnǎn lɛ tě laj
Oh! today 1 be_angry very 1 forget take bag put thing come so and IRR ACH

sûɪ thŏŋ sɔ̌ŋ thŏŋ jâw
buy bag two bag PFV

'Oh, today I am angry. I forgot to bring my bags for putting things in and I will have to buy two bags.'

The container terms are much like the group terms discussed in the previous section. They can function as nouns or classifiers, and the fact that the same word is used for the noun and classifier in some contexts has an effect on the syntax. This will be discussed more in the following section.

3.2.5 Summary of classifiers

Table 3.6 summarizes the data presented in this section. Based on this data, there appear to be two or three distinct distributions of classifiers. The most restricted of those is the measure expression, which I would argue does not have a noun function in addition to its classifier function. The second category is classifiers that can function as bound morphemes in N-N compounds. These can be head nouns in addition to being classifiers. They cannot be repeater classifiers because the noun and classifier will never be entirely identical. The third category is classifiers that have an independent noun function. Since the noun and classifier can be identical in certain constructions, the noun can be deleted.

Table 3.6: Summary of classifier distribution

Classifier Type	Classifier dist.	Head Noun	Repeater Clf
Individual	✓	✓	✗/✓
-distinct Clf and N	✓	✓	✗
-same Clf and N	✓	✓	✓
Measure	✓	✗	✗
Group	✓	✓	✓
Container	✓	✓	✓

3.3 Structure of numeral-classifier expressions

3.3.1 Repeater classifiers

There is a set of nouns in Shan that have identical nouns and classifiers. I call these Repeater Classifiers. Simpson (2005) has previously identified a set of 'self-classifiers' or 'repeaters' in Thai, Burmese, and Lahu, where the classifier and noun are identical.[5] The group terms discussed in section 3.2.3 and the container terms discussed in section 3.2.4 are examples of repeater classifiers.

These are different from the measure expressions described in the previous section (§3.2.2) because they can have the form A_n Num A_{clf} (where A_n and A_{clf} are homophonous) whereas the measure expression can only have the form A_n Num B_{clf} (where A_n and B_{clf} are not homophonous). While these constructions can have the typical N Num Clf (A_n Num A_{clf}), they often have the form Num Clf/N (Num A_{clf}), which I call the Reduced Repeater form. A non-exhaustive list of categories with the Num Clf/N form includes time expressions (197), container terms (198), and political bodies (199).

(197) sɔ̌ŋ pǐ
two year
'two years'

(198) sɔ̌ŋ tɯk
two box
two boxes

(199) sɔ̌ŋ tsɯŋmɤ́ŋ
two country
'two countries'

[5]Hundius & Kölver (1983) and Simpson (2005), have noted that Thai has a set of measure terms which appear with no head noun.

For a similar construction in Mandarin Chinese, Huang (2020) calls the CLF/N forms that follow the numerals 'post-numeral lexical items' (PNLs), and I will do the same. (200a) and (200b) show the Chinese PNL is not based on a repeater classifier construction. This is evident from the fact that the noun and classifier in (200b) are not the same.

(200) MANDARIN CHINESE, adapted from Huang 2020: (20b)

a. yi hua
one stroke
'one stroke' (PNL)

b. yi ge bihua
one CLF stroke
'one brushstroke' (NUM CLF N)

In Shan, some repeater classifiers are ungrammatical with the N Num Clf structure (201), and need the reduced structure (202). Others are fine (203) reduced or un-reduced.

(201) *wán sǎam wán
day three day
'three days'

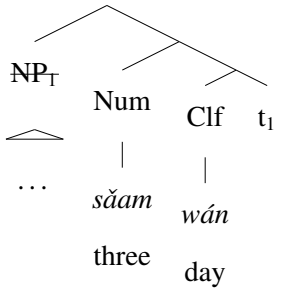
(202) sǎam wán
three day
'three days'

(203) (taw) sŏŋ taw
(bottle) two bottle
'two bottles'

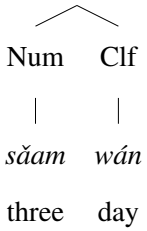
I propose that the repeater classifier constructions in Shan show evidence to support the structure in (204), which involves deletion of the NP when it is the complement of an identical Clf.[6] However, beyond the particular analysis used to account for this data, I will show that all the possible accounts of Shan repeater classifiers support a structure where the classifier and noun form a constituent before they combine with the numeral [Num [Clf N]] rather than one where the classifier and numeral form a constituent before they combine with the noun [[Num Clf] N]. For example, another possible structure is (205), but I will argue that both are most compatible with the [Num [Clf N]] structure.

[6]The deleted NP is at the surface position, but I am not making any claims about whether the deleted NP would in fact move.

(204) Num Clf (deleted NP)



(205) Num Clf



PNL is a classifier

As (206) shows, the classifier can be reduplicated in reduced repeater constructions. This is different from the homophonous forms of the noun and classifier in repeater classifier constructions. The quantifier *ku* 'every' typically appears in the same position as a numeral between the noun and classifier, suggesting that (206) is a reduced repeater construction. (207) shows a similar reduplication using *tǒ* the classifier for the noun *nôk* 'bird'. Assuming this is the same construction, as it appears to be, what is being reduplicated is the classifier. The reduplication here may be associated with distributivity.

(206) ku wán wán nâj háw ǩǐn wáaj sɔ̌ŋ kɔ́k ʔɔ
each day day this 1 drink wine two cup PRT
'Each day, I drink two cups of wine.'

(207) nôk ku tǒ tǒ nâj pěn tǒ mí tsǎj sâttǎwàa
bird every CLF.ANML CLF.ANML this COP body have mind rational
'Each bird is a living creature.'

Reduced repeater is referential

(208) and (209) show that even in reduced repeater form, the word *ts*ɯŋ*mɤ́*ŋ 'country' can be interpreted referentially, meaning that it can be used to refer to the countries themselves rather than simply to the amount of countries (MEASURE).

(208) náj lumfâa nâj mí jù sɔ̌ŋ tsɯŋmɤ́ŋ ʔǎn pěn pǎaŋtúk páaj-maakmí
in world this have IPFV two country COMP COP fight field-wealth

tɔ̀ kǎn jù
toward each_other IPFV

'In the world, there are two countries competing over economics.'

(209) náaŋ_ʔɔ̀n jaam ʔɛ̀wlɛ̀ sɔ̌ŋ tsɯŋmɤ́ŋ nân
Nang_Orn have_ever visit two country that
'Nang Orn has visited those two countries.'

(208) and (209) also demonstrate that the reduced repeater construction can be part of a larger nominal structure, so the reduced repeater constituent likely is the same grammatical category as the un-reduced version. (210) includes another referential reduced repeater expression with a demonstrative and the noun *wán* 'day'.

(210) sǎam wán nân pěn wán lǎj/sǎŋ
three day that COP day which/what
'Those three days are which/what days?'

Repeater classifiers are just like other classifiers in their distribution and interpretation. Where they differ is in that the fact that the classifier is identical to the noun. This allows the noun to be deleted as long as it is not being modified.

PNL is not a noun

Simpson & Ngo (2018) offer two possible explanations for Thai and Burmese repeater classifiers, the second of which they adopt.

1. A noun is merged both in the N and Clf position and then the identical N is deleted through haplology.
2. A noun is merged in the N position and moved to Clf.

Head movement

If the PNL were in N or moved from N to Clf, we would expect that adjectives modifying the nouns would be visible in this construction. This is not what we find.

See the contrast between the three examples in (211)–(213). The Num N structure (202), repeated in (211), is preferred except when the noun is modified by something, as in (212). For repeater constructions, it tends to be better to use the Num N structure rather than the N Num Clf structure. When the noun is modified in (212), we can see that the modified noun *wán lɯ* 'rest day' has to be in the typical noun position. It is not possible to have the structure in (213) where the repeater noun has been deleted and the modifier is on either side of the numeral and classifier.

(211) sǎam wán
three day
'three days'

(212) wán lɯ sǎam wán
day rest three day
'three rest days'

(213) (*lɯ) sǎam wán (*lɯ)
rest three day rest
'three *rest days'

The same is true for container terms like *tɯk* 'box' in (214) and (215) below. In (214) the Num N form is used—or N Num for 'one'. In (215), the *tɯk màakkhɔ̌* in the N position specifies what is in the box and the following Num Clf says how many there are. When the modifier is present, the noun must be present also.

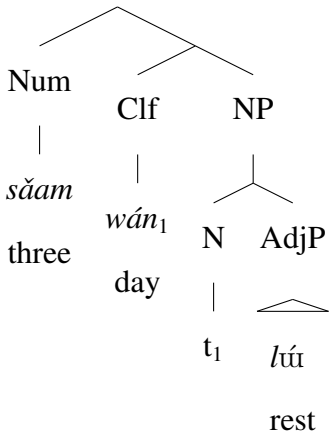
(214) háw mí sɔ̌ŋ tɯk jâw kójkaa lò thɛŋ tɯk nɯŋ
1 have two box PFV but need another box one
'I have two boxes but I need one more.'

(215) tɯk màakkhɔ̌ sɔ̌ŋ tɯk
box jujube two box
'Two boxes of jujube'

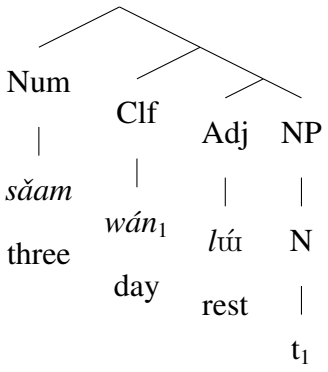
(216) and (217) show two possible structures for a head movement account of repeater constructions. In (216), the adjective is adjoined to the noun. In this case, we would expect to see [Num Clf Adj] constructions, which are not available, shown in (213). In (217), the adjective takes the noun as a complement. If we were to adopt this account, we could say that the adjective head

blocks movement of N to Clf.[7] If movement is blocked in the case of (217), it would need to be possible for the noun and classifier to be inserted separately to get the licit example in (212). Note that the head movement account requires that the base structure be [Num [Clf N]] where the noun and classifier form a constituent before they combine with the numeral.

(216) Head movement with adjoined Adj



(217) Head movement with Adj head



If only the head N moves to Clf, we would expect the adjective to still be possible in one of the positions in (216), contrary to fact. Additionally, it is possible to see both the N and Clf present when the noun is modified, so the PNL must not be a noun syntactically. If the structure should be (217), it supports an analysis where the underlying structure of the numeral-classifier expression is [Num [Clf N]].

Haplology

Dong & Wong 2020, referencing Chao (1968), notes that in Mandarin Chinese, it is possible to have haplology of functional morphemes. Two examples are given in (218)–(219). In (218), there is haplology of one instance of *le*. One is functioning as a perfective suffix, which would attach to the predicate verb or adjective. The other is a sentence final particle for new information. In (219), there is haplology of one instance of *de*. One instance is a nominalizer and one is a marker

[7] See Hankamer & Mikkelsen 2002 for a comparison of the adjoined versus head analysis of adjectives in analyzing definite nouns in Danish.

of modification or possession. In both cases, the form with both instances of the morpheme is not allowed, so haplology is obligatory here.

(218) yǐ.jīng zǒu le
already go LE
'(Someone) has already gone.' (Dong & Wong 2020: (5); citing Chao 1968: 247)

(219) shì nàge mài cài de
be that sell vegetable DE
'It's that vegetable vendor's. (Dong & Wong 2020: (6); citing Chao 1968: 298)

The haplology here seems to be motivated by the identical, adjacent phonological forms of these morphemes. These cases are interesting because even though the morphemes are not semantically identical, haplology still takes place.

Would it be possible to argue for a [[Num Clf] N] structure for numeral-classifier constructions in Shan using a haplology account? I would argue that such an account is not possible. If the structure were [[Num Clf] N], there would be no movement in the numeral-classifier constructions. The numeral would intervene between the noun and classifier, so the haplology could not be restricted to direct adjacency. First, it would need to be possible for haplology to take place even when only one word intervenes between identical words. The argument would go as follows:

1. Assume the deletion rule is [~~X~~ Y X]
2. Given the string *wán sǎam wán*, we would get ~~*wán*~~ *sǎam wán* 'three days'
3. Given the string *wán lɯ sǎam wán* 'three rest days', there would be no deletion

Then, we would expect that deletion would be blocked anytime there were two words intervening between the noun and classifier. However, when something intervenes modifying the [Num Clf] expression rather than the noun, deletion still takes place, as shown in the contrast between (220) and (221).

(220) *wán mɔk săam wán
day about three day
intended: ‘about three days’

(221) mɔk săam wán
about three day
‘about three days’

(222) [[$_{NP}$ ~~wán~~]$_i$ [$_{ClfP}$ săam [$_{ClfP}$ wán t$_i$]]]
day three day
‘three days’

(223) [[$_{NP}$ wán lɯ]$_i$ [$_{ClfP}$ săam [$_{ClfP}$ wán t$_i$]]]
day rest three day
‘three rest days’

(224) [[$_{NP}$ ~~wán~~]$_i$ [$_{ClfP}$ [$_{AdvP}$ mɔk] [$_{ClfP}$ săam [$_{Clf'}$ wán t$_i$]]]]
day about three day
‘about three days’

Thus, it is not possible to use a haplology account argue for a [[Num Clf] N] structure, where the numeral and classifier form a constituent first.

The reduced repeater form is licensed by having identical classifier and nominal constituents. It is often not good to have both N and Clf in repeater classifier expressions (185).

(185) mɤnâj háw laj laat tɔ̀ (#tsúm) sɔ̌ŋ tsúm
today 1 ACH speak to (#group) two group
intended: Today, I spoke with two groups.

These Clf’s can be used with other nouns, as shown in (187) with noun *kón* ‘person’. In these cases, deletion cannot take place.

(187) mɤnâj háw laj laat tɔ̀ kón sɔ̌ŋ tsúm
today 1 ACH speak to people two group
‘Today, I spoke with two groups of people.’

Even when the classifier is identical to one morpheme in a N-N compound, as in (167a), repeated in (225), and (188), repeated in (226), the noun and classifier positions must be overt. However, it is relevant whether the morphemes in expressions like *ton-mâj* ‘tree’ can be syntactically independent of one another. Barrie & Mathieu (2016), citing Baker (2009), posit that English

compounding takes place in the lexicon, so compound morphemes are not syntactically independent.[8] It is likely that the same in true in Shan. If this is the case, we cannot rule out the head movement analysis based on evidence from compounds.

(225) ton-mâj sǎam ton
plant-tree three plant
'three trees'

(226) mɤnâj háw laj laat tɔ̀ tsúm lukhén sɔ̌ŋ tsúm
today 1 ACH speak to group student two group
'Today, I spoke with two groups of students.'

Thus, complete overlap between the NP and Clf licenses the reduced form.

Analysis

I argue that the data in (201)–(212), repeated below, are due to a type of NP deletion, caused by the identical form in Clf.

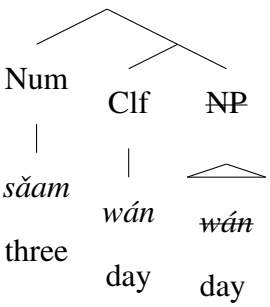
(201) *wán sǎam wán
day three day
'three days'

(202) sǎam wán
three day
'three days'

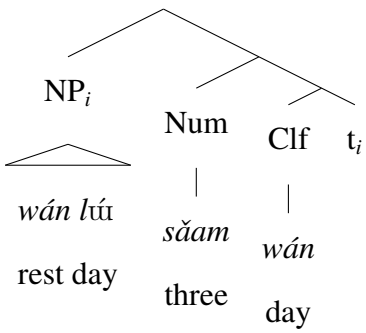
(212) wán lɯ sǎam wán
day rest three day
'three rest days'

Therefore, I propose the structure in (204), repeated in (227), for the data in (202) and (228) for the data in (212).

(227) Num Clf (deleted NP)



(228) N Adj Num Clf



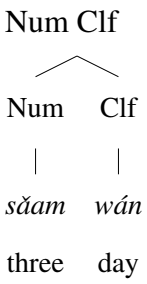
[8] See Harley (2009) for an alternative approach.

If we assume the structure [Num [Clf N]] as in (204), then the classifier c-commands the noun at some point in the derivation. Then, the identical form of the classifier and noun licenses deletion of the noun complement. In (228) the complement of the classifier is not identical to the classifier so deletion does not take place.

Semantics of reduced repeater constructions

While, the structure in (204) better accounts for the variation between the reduced and un-reduced repeater classifier constructions, the structure in (205) could account for the reduced forms. Huang (2020) argues for the structure [[Num Clf] N] in (205) for PNLs in Mandarin Chinese. However, the semantics of the classifier in (205) must be a fusion of the semantics of a noun and classifier in the structure [Num [Clf N]], rather than a fusion of the noun and classifier in the structure [[Num Clf] N] .

(205)



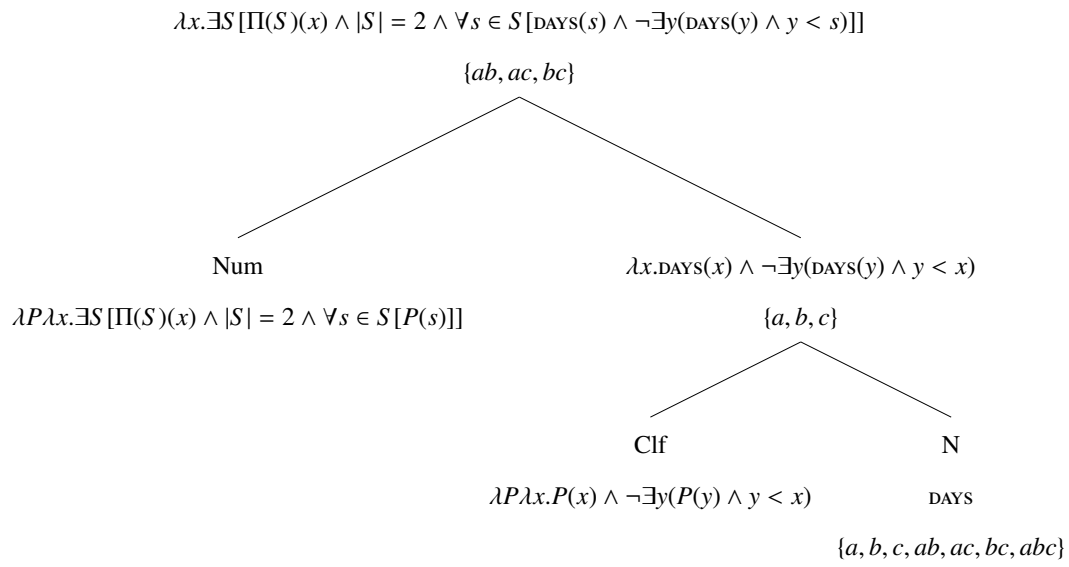
(229) MANDARIN CHINESE, adapted from Huang 2020: (20b)

a. yi hua
one stroke
'one stroke'
(PNL)

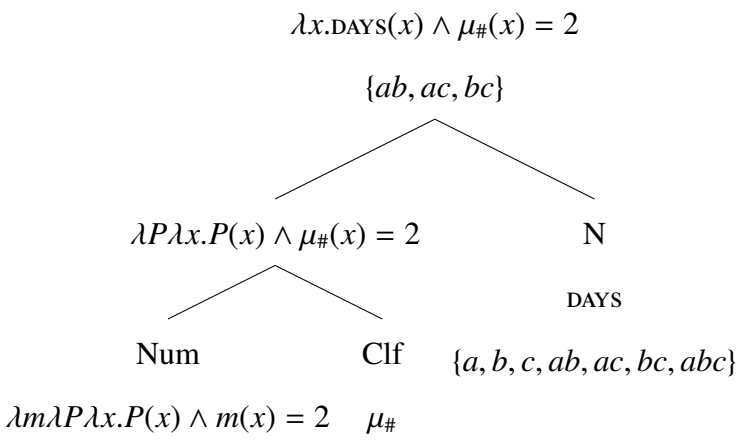
b. yi ge bihua
one CLF stroke
'one brushstroke'
(NUM CLF N)

(230) shows a possible semantic derivation based on the structure [Num [Clf N]], and (231) shows a derivation based on the structure [[Num Clf] N] . These use the semantics proposed by Little et al. (to appear) for numeral-classifier expressions with the structures [Num [Clf N]] and [[Num Clf] N]. The detailed semantics of each component, which will be discussed more in section 3.4, is not as relevant here as the order in which the N, Num, and Clf combine. [Num [Clf N]] assumes the noun and classifier combine first, then the numeral, and [[Num Clf] N] assumes the numeral and classifier combine first, then the noun.

(230) [Num [Clf N]] (adapted from Little et al. to appear: (11))



(231) [[Num Clf] N] (adapted from Little et al. to appear: (8))



As a result, there is a node in the semantic tree for [Num [Clf N]] that corresponds to a combined classifier and noun semantics, which would be an appropriate semantics for the classifier in (205). However, there is no such node in tree in (231), since the noun and classifier never form a constituent. Therefore, I propose that the structure in [Num [Clf N]] is most compatible with the repeater classifiers in Shan. This is compatible with the classifier typology in Little et al. to appear. NP deletion is licensed when the classifier complement noun is identical to the classifier and cannot take place when something is modifying the noun.

3.3.2 Quantification

As has been mentioned previously, the numeral-classifier phrase can be separated from the noun. This is a phenomenon that has been called 'quantifier float' (Thai, Jenks 2013), but is very different from quantifier float in English. In Shan, the floated quantifier appears (i) following a verb in what looks like a serial verb construction or (ii) at the end of the clause, even following aspect markers. The 'quantifier' that floats is typically a numeral-classifier expression or a quantifier-classifier expression.

The floated numeral-classifier phrases appear to the right of the aspect markers *jù* and *jâw* in (232) and (233), repeated from (173) above. Thus, the numeral-classifier phrase must be adjoined at the VP level (if not higher). Floated quantifiers cannot appear to the right of the sentence final polar question marker *hâa* in (234). These constructions cannot be analyzed in the same way as quantifier float in English since the floated material is in a position where the associated nominal argument has never been.

(232) măa háw tŏ sǐ-khăaw nân hàw jù sɔ̆ŋ tŏ
dog 1 CLF.ANML color-white that bark IPFV two CLF.ANML
'Two of my white dogs are barking.'

(233) phujíŋ hɔtphɛ̆w kéŋmàj jâw săam kô
woman arrive Chiang_Mai PFV three CLF.HUM
'Three of the women arrived in Chiang Mai.'

(234) *phujíŋ hɔtphɛ̆w kéŋmàj jâw hâa săam kô
woman arrive Chiang_Mai PFV POLAR.Q three CLF.HUM
intended: 'Have three of the women arrived in Chiang Mai?'

Simpson (2011) and Jenks (2013) analyze quantifier float as a kind of movement in Thai. For Jenks (2013) this involves quantifier raising and for Simpson (2011) this involves rightward movement of the quantificational material. An adverbial analysis of quantifier float was considered and rejected by both on the basis of the fact that this construction is sensitive to islands for movement.

Jenks (2013) noted that floated quantifiers in Thai affect their semantic scope. This seems to be the case in Shan as well. When the quantifier is floated from the subject, as in (235), it scopes under negation. When the quantifier is not floated, it scopes over negation, as in (236). Investigating the full relationship between floated quantifiers and scope in Shan is left to future work.

(235) lukhén ʔăn mí jù náj hɔŋhén ɳaj ʔàm mí lɔŋ-mantsăj nàj tŏtsawkàw ku kɔ̂
student COMP have stay in class this NEG have NOM-be.sure in self each CLF.HUM
'Not every student in the class has confidence in themself.'

¬ > ∀

(236) lukhén ku kɔ̂ ʔăn mí jù náj hɔŋhén ɳaj ʔàm mí lɔŋ-mantsăj nàj tŏtsawkàw
student each CLF.HUM COMP have stay in class this NEG have NOM-be.sure in self
'Every student in the class doesn't have confidence in themself.'

∀ > ¬

As the following examples show, there are adverbial uses of the distributive construction.

(237) hàw kwàa thóp mɔ̆jăa nân nɯŋ lɤ̆n lăj nɯŋ pɔk
1 go meet doctor that one moon which.Q one occurrence
'I go see that doctor once a month.'

Often in distributive quantificational contexts, quantifier float is used. An example of a distributive quantificational construction can be seen in (238). Here the quantificational construction appears as to be in an argument position following *păn* 'give'. The numeral-classifier expression *n*ɯŋ *k*ɔ̂ 'one person' is connected to another numeral-classifier expression *n*ɯŋ *k*ɔ̆ŋ 'one pile' by a morpheme *lăj*, which normally combines with a classifier to express the question 'which one'.

(238) háw mí khawmún. háw khaj phɛ̆ păn tɔntàa săam kɔ̂. háw máa hét pĕn săam
1 have cake 1 want split give for three CLF.HUM 1 come do COP three

kɔ̆ŋ sĕ păn nɯŋ kɔ̂ lăj nɯŋ kɔ̆ŋ
pile and give one CLF.HUM which.Q one pile

'I have cake. I want to split it up to give to three people. I make it into three piles and give one pile per person.'

(239) provides a clear example of a floated numeral-classifier in a distributive context.[9] Such a construction is also the only place where a classifier can appear without either its associated numeral or noun. Here, *kɔ̂ sɔ̌ŋ ʔǎn* means 'two (assignments) per person'. The classifier *kɔ̂* has with it neither its associated numeral 'one' nor its noun. In the parallel construction in English, 'two per person', the numeral expression is assumed to have an elided noun 'assignments'. The Shan construction would seem to allow more elements be elided. Given that *kɔ̂* is the classifier for humans and there are two sets of humans described in this sentences (i.e., teachers and students), this can be ambiguous (though the more prominent reading is 'two per student'). Both nominal arguments already have an associated universal quantifier, so the distributive construction could not have formed a constituent with them before movement.

(239) khúsɔ̌n ku kɔ̂ náj hóŋhén ʔàan kǎanwaan ʔǎn lukhén ku kɔ̂ hét sòŋ
teach every CLF.HUM in school read assignment COMP student every CLF.HUM make send

pǎn kɔ̂ sɔ̌ŋ ʔǎn
give CLF.HUM TWO CLF.GN

'Every teacher in the school read assignments that their students had completed and turned in, two per person.'

It seems likely that it is necessary to have an adverbial account of the floated quantifiers that cannot be associated with an argument position.

For cases like (240) which involve a left-located constituent, it might be possible to argue for NP-ellipsis. Here, the left-located constituent is the prepositional phrase *náj phujíŋ sǎam kɔ̂ nân* 'among those three women' and the floated expression is *kɔ̂ nɯ̌ŋ* 'one of them', which refers to one of the three women.

[9]This is different from Thai, where the distributive construction cannot float (Jenks 2013). Notably, the equivalent of the distributive operator, tɛ̀ɛ*lá*ʔ-CLF, that Jenks (2013) mentions for Thai is not found in Shan. However, the construction CLF_1 láʔ NUM CLF_2 is similar to the construction discussed here for Shan.

(240) [náj phujíŋ sǎam kô nân]$_i^{PP}$ tsáaj lǎawkhám tsaaŋ kwàa hǎa [náaŋ lǎawŋúun]$_{k<i}$/
in women three CLF.HUM that Jai Lao_Kham can go find Nang Lao_Nguen/

[kô nɯɯŋ]$_{k<i}$
CLF.HUM one

'Among [those three women]$_i$, Jai Lao Kham could find Nang Lao Nguen$_{k<i}$/[one of them]$_{k<i}$.' (PP)

Aikhenvald (2000) noted that numeral classifiers can be used anaphorically without an overt noun, such as in Minangkabau in (241).

(241) a. bara pisang sa-sikek Mak
how.much banana one-NUM.CL:BUNCH 2nd.FEM.HON
'How much does a bunch of banana cost?' (Minangkabau, Aikhenvald 2000: (12.15a)

b. agiah duo sikek
give two NUM.CL:BUNCH
'Give (me) two (bunches).' (Minangkabau, Aikhenvald 2000: (12.l5b))

These are cases where arguments of the same or previous clauses can be anaphorically linked to the floated quantificational material. Since the classifier expresses morphological agreement with its associated noun, it supports NP-ellipsis, in the same way Corver & van Koppen (2011) argued that morphological agreement on adjectives licenses NP-ellipsis in Dutch. Following Saito & Murasugi's (1990) and Lobeck's (1995) proposal that NP ellipsis, VP ellipsis, and sluicing are licensed when the specifier of the relevant functional head is filled, we can explain Shan NP-ellipsis by the fact that the functional classifier phrase specifier is filled by a numeral.

The facts related to quantifier float in Shan require a more in depth look. There are some clear cases of adverbial quantificational expressions and some cases that look like NP-ellipsis, but this does not give a full account of such expressions.

3.3.3 Structure

Evidence from measure adverbs

Li (2011) uses the position of certain morphemes in Mandarin Chinese classifier/measure phrases to argue for separate syntactic structures for count versus measure interpretations. Based on an observation initially by Lü (1980/1990), the morpheme *duo* 'more than' can either precede or follow a classifier, as shown in (242a)–(242b), Li (2011) argues that when *duo* follows the classifier, only the measure interpretation is available. In (243), where *duo* follows the classifier, only the measure interpretation in (243a) is available not the count interpretation in (243b).

(242) a. ta ling le shi duo ping hongijiu.
he carry PERF ten more CLF$_{bottle}$ red_wine
'He carried more than ten bottles of red wine.'

(Mandarin Chinese, Li 2011: p. 114, (25a); [counting])

b. ta zhishao neng he-xia shi duo ping hongijiu.
his at_least can drink-down ten more CLF$_{bottle}$ red_wine
'He can at least drink more than ten bottles of red wine.'

(Mandarin Chinese, Li 2011: p. 114, (25b); [measure])

(243) a. ta jinwan zonggong he le you san ping duo hongjiu.
he tonight altogether drink PERF have three CLF$_{bottle}$ more red_wine
'Tonight, he drank more than three bottles of red wine.'

(Mandarin Chinese, Li 2011: p. 114, (26a); [measure])

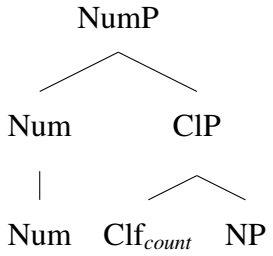
b. #ta ling le shi ping duo hongjiu.
he carry PERF ten CLF$_{bottle}$ more red_wine
Intended: 'He carried more than ten bottles of red wine.

(Mandarin Chinese, Li 2011: p. 114, (26b; [counting])

Based on this data, Li 2011 proposes the following structures for counting (244) and measure (245). Under this analysis *duo* 'more than' is adjoined to Num in both counting and measuring contexts, and the constituency of these constructions can be deduced from the position of *duo* with respect to other words.

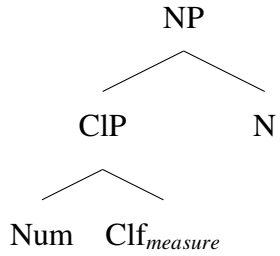
(244) Counting reading

(Li 2011: p. 118, (31a))



(245) Measure reading

(Li 2011: p. 118, (31b))



It is not clear that Shan displays such differences in counting and measuring contexts. I will be using the phrases *khɯnnɤ̌* 'more than', *mɔk* 'approximately', and *ʔàm pôn* 'not more than' to probe the constituency of these classifier/measure phrases.

The phrase *khɯnnɤ̌* 'more than' typically follows the numeral and classifier expression. Since the noun always appears before the numeral and classifier, *khɯnnɤ̌* never intervenes between the noun and numeral-classifier. In (246), the Num N expression is followed by *khɯnnɤ̌*. In the analysis of repeater classifiers like *tsɯŋmɤ́ŋ* 'country', the noun—which would otherwise appear in a position to the left of the numeral—has been deleted, but no other difference is expected since *khɯnnɤ̌* modifies the classifier phrase rather than the noun.

(246) náj lumfâa nâj mí jù sɔ̌ŋ tsɯŋmɤ́ŋ khɯnnɤ̌ ʔǎn mí pǎaŋtúk páajmaakmí
in world this have IPFV two country more_than COMP have battle economy
tɔ̀ kǎn
against each_other
'In the world there are more than two countries fighting over economics.'

As (247)–(249) show, the expression *khɯnnɤ̌* only appears in the position after the Num Clf.

(247) thɛŋ sàj nâm sǎam kɔ́k khɯnnɤ̌ ti náj mɔ nân
add put water three cup more_than at in pot that
'Add more than three cups of water to that pot.'

(248) * thɛŋ sàj nâm khɯnnɤ̌ sǎam kɔ́k ti náj mɔ nân
add put water more_than three cup at in pot that
intended: 'Add more than three cups of water to that pot.'

(249) * thɛŋ sàj khɯnnɤ̌ nâm sǎam kɔ́k ti náj mɔ nân
add put more_than water three cup at in pot that
intended: 'Add more than three cups of water to that pot.'

All such expressions seem to appear immediately before or after the Num Clf expression. Another such amount-related adverb is *mɔk* 'approximately'. This element precedes rather than follows the numeral-classifier, as shown in (250)–(251). It also follows the noun obligatorily as (252) demonstrates.

(250) ʔǎw nâm thɛŋ mɔk sǎam kɔ́k pǎaj sàj ti náj mɔ nân
take water add approximately three cup exceed put at in pot that
'Add approximately three cups of water to that pot.'

(251) ʔǎw nâm mɔk sǎam kɔ́k thɛŋ pǎaj sàj ti náj mɔ nân
take water approximately three cup exceed add put at in pot that
'Add approximately three cups of water to that pot.'

(252) *ʔǎw mɔk nâm sǎam kɔ́k thɛŋ pǎaj sàj ti náj mɔ nân
take approximately water three cup exceed add put at in pot that
intended: 'Add approximately three cups of water to that pot.'

Examples (253) and (254) show the measure construction and count construction involving containers. The position of *mɔk* 'approximately' does not vary with the different constructions.

(253) mán-tsáaj ʔǎw wáaj lɛ̌ŋ máa mɔk síp taw
3-MASC take wine red come approximately ten bottle
'He brought approximately ten bottles of red wine.'

(254) mán-tsáaj ʔǎw taw wáaj lɛ̌ŋ máa mɔk síp taw
3-MASC take bottle wine red come approximately ten bottle
'He brought approximately ten bottles of red wine.'

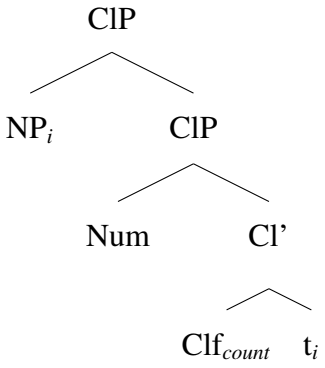
The same fixed positioning is true of expressions like *ʔàm pôn* 'no more than', which always appear before the numeral-classifier. This is true even with the Num N construction, the *ʔàm pôn* still appears before the numeral, as in (255) and (256).

(255) ʔàmpôn sì tsɯŋmv́ŋ ʔǎn háw laj hɔtphɛ̌w
no_more_than four countries COMP 1 ACH visit
'(There are) no more than four countries that I have been to.'

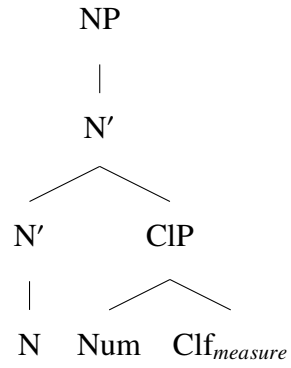
(256) háw ʔǎw tɛ̀k ʔàmpôn sǎam kɔ́k
1 take break no_more_than four cup
'I broke no more than three cups.'

It seems that the likely structures for Shan numeral classifier are either (257) or (258).

(257) Counting reading

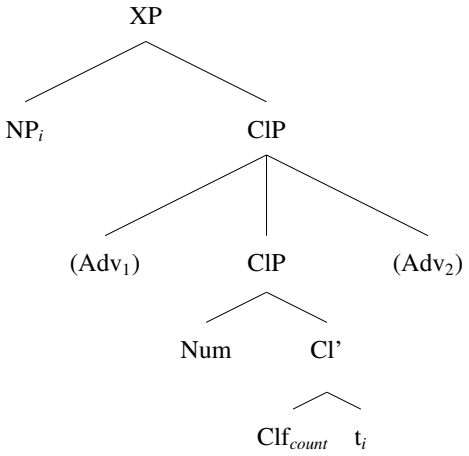


(258) Measure reading

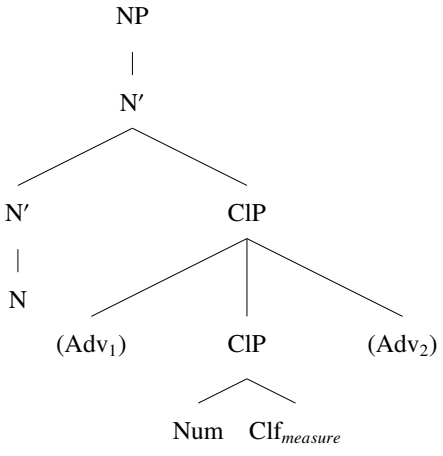


However, if adjunction of these phrases takes place in the way laid out in (259) and(260),[10] it is still not clear that any distinctions can be drawn between the two structures based on this data. As I argued in section 3.3.1, the repeater classifier data is most compatible with the analysis in (257).

(259) Counting reading



(260) Measure reading



[10]Note that I am not claiming that the structure would have Adv$_1$ and Adv$_2$ type adverbs at the same time.

Evidence from order and selection

Constituent order and patterns of selection support analyzing numeral-classifiers as something non-predicative.

(261)–(262) shows that the order of elements in the nominal is N Clf Adj Num Clf since the adjective cannot appear after the numeral-classifier unless it is made into a separate predicate.

(261) háw lɤ̆k laj măa tŏ haaŋtsâa nɯŋ tŏ
1 choose ACH dog CLF.ANML appearance-bad one CLF.ANML
'I chose one ugly dog.'

(262) háw lɤ̆k laj măa săam tŏ *(pĕn măa) haaŋtsâa
1 choose ACH dog three CLF.ANML *(COP dog) appearance-bad
'I chose three dogs *(that are) ugly.'

(263) shows that the position for the demonstrative is after the numeral-classifier.

(263) háw lɤ̆k laj măa tŏ haaŋtsâa săam tŏ nâj
1 choose ACH dog CLF.ANML appearance-bad three CLF.ANML this
'I chose these three ugly dogs.'

As (264) demonstrates, relative clauses follow the adjective and numeral-classifier, but precede the demonstrative. When a numeral-classifier follows the relative clause and demonstrative as in (265) there is a partitive interpretation. A numeral-classifier can follow a relative clause without leading to a partitive interpretation if there is not a demonstrative between the noun and the numeral-classifier, as in (266).

(264) háw sûɯ măa (tŏ) haaŋtsâa săam tŏ [ʔăn háw laj lɤ̆k wâj mɤ
1 buy dog CLF.ANML appearance-bad three CLF.ANML COMP 1 ACH choose stay when

qɔ̆ntáaŋ] (nân)
before that

'I bought three ugly dogs that I had chosen earlier.'

(265) háw sûɯ măa [ʔăn háw laj lɣk wâj mɣ qɔ̆ntáaŋ] nân săam tŏ
1 buy dog COMP 1 ACH choose stay when before that three CLF.ANML
'I bought three of the dogs that I had chosen earlier.'

(266) háw sûɯ măa ʔăn háw laj lɣk wâj mɣ qɔ̆ntáaŋ (*nân) săam tŏ tsɣ́ nân
1 buy dog COMP 1 ACH choose stay when before (*that) three CLF.ANML PL that
'I bought those three dogs that I had chosen earlier.'

Shan adjectives generally can be predicated of nouns, as with *món* 'round' in (267). This is not true of numerals, as in (268). The way to have a predicate with a numeral is to use the verb *mí* 'have', as shown in (269). This suggests that the numeral-classifier expression is not able to function as a predicate in the way that adjectives and verbs are. This could indicate that the numeral-classifier expression is not a predicate.

(267) màaknăŋ tsɣ́ nâj món
ball PL this round
'These balls are round.'

(268) *măa tsɣ́ nâj săam tŏ
dog PL this three CLF.ANML
intended: 'These dogs are three.'

(269) măa tsɣ́ nâj mí săam tŏ
dog PL this have three CLF.ANML
'These dogs are three.'

The copula *pĕn* similarly takes arguments that do not function as predicates by themselves. In (270), the copula take a relative clause version of the predicate *món* 'round'. One difference between *mí* and *pĕn* can be seen in (271)–(272). The copula indicates there is identity between 'my dogs' and the group of sleeping ones, while *mí* indicates that there is overlap between those two groups.[11] The copula and *mí* both can select for relative clauses and other nominal expressions.

(270) màaknăŋ tsɣ́ nâj pĕn ʔăn món
ball PL this COP COMP round
'These balls are round ones.

[11]This is compatible with an analysis of *mí* as an existential quantifier.

(271) măa háw pěn ʔăn nɔ́n jù
dog 1 cop comp sleep ipfv
'My dogs are the sleeping ones.'

(272) măa háw mí ʔăn nɔ́n jù
dog 1 have comp sleep ipfv
'My dogs have some that are sleeping.'

Numeral-classifier expressions can also follow the copula as in (273). In this case, the copula marks a change of state. The verb *mí* means something like 'have', as (274) demonstrates. If all other elements that are selected by *mí* and *pěn* are nouns, we might propose that numeral-classifier expressions are nouns too, perhaps with an elided NP as I proposed above.

(273) háw ʔăw tsúm nâj phɛ̌ pěn sɔ̌ŋ tsúm lɨ̌ hâa
1 take group this split cop two group good polar.q
I split the group into two groups, is that good?

(274) háw mí tájkɔ̂ lɨ̌ sɔ̌ŋ kɔ̂
1 have friend good two clf.hum
'I have two close friends.'

Looking at (275), we can see that *tǒ jàj* has the status of a predicate, rather than a noun. It can combine directly with a nominal subject without the need of a verb like *mí* or *pěn*. This raises the question about the function of the noun/classifier *tǒ* in (275). Jenks (2012) argued that in Thai these constructions are complex predicates and not in the same syntactic position as the classifiers in other constructions. This argument is based on the fact that (i) it is possible to omit the classifier without a change in meaning, (ii) different classifiers are used than in other classifier uses, and (iii) the Clf-Adj construction can be used predicatively. This appears to be true for Shan as well. However, point (i) is not entirely true since the classifier can specify the unit that is being distributed to in cases where there could be ambiguity, as will be discussed in section 3.4.1.

(275) măa tsɤ́ nân (tǒ) jàj nàa
dog pl that (clf.anml) big very
'Those dogs, they are very big.'

Looking at (276), it is possible to coordinate two of such adjective phrases even when the classifier is distinct. Here, *hòj jàj* 'large (fruth)' and *kɔ̌ŋ jàj* 'large pile' are being coordinated. This is, however, not possible for all classifiers, as (277) shows. This difference could be due to the distinction between the container and contents in (277) that does not exist for (276). (277) uses predicate describe both the container and contents.

(276) màakkhɔ̌ kɔ̌ŋ nâj hòj jàj nàa lɛ kɔ̌ŋ jàj (nàa)
jujube pile this CLF.RND big very and pile big very
'This pile of jujube, the jujube are big and the pile is big.'

(277) *khaw thǒŋ nân mêt jàj lɛ thǒŋ lêk
rice bag that seed big and bag small
intended: 'That bag of rice, the grains are big and the bag is small.'

A better way to express the meaning of (277) is given in (278). Here, the preposition *náj* helps more clearly specify that the rice is being described rather than the bag. The pronouns in the matrix clause cross-reference the rice in possessive constructions in order to describe the grains and the bag. The expression *mêt mán* 'its grains' is the subject of the predicate *jàj* 'big.' Instead of coordinating predicates, (278) conjoins clauses.

(278) khaw náj thǒŋ nân mêt mán jàj lǐ nàa kójkaa thǒŋ mán lêk nàa
rice in bag that seed 3 big good very but bag 3 small very
'The rice in the bag, the grains are nice and big but the bag is very small.'

(279) uses a different kind of coordination that relies on the left-dislocation construction. Here, the left-dislocated *màakkhɔ̌ kɔ̌ŋ nâj* is referred back in parts that are followed by the focus particle *kɔ*. The first part *hòj*, which notably does not require an overt possessive pronoun, combines with the predicate *hòj jàj* and the second part *kɔ̌ŋ* combines withe the predicate *kɔ̌ŋ jàj*.[12]

(279) màakkhɔ̌ kɔ̌ŋ nâj, hòj kɔ hòj jàj, kɔ̌ŋ kɔ kɔ̌ŋ jàj
jujube pile this CLF.RND PRT CLF.RND big pile PRT pile big
'This pile of jujube, the jujube are big and the pile is big.'

[12]It is likely that the two freestanding 'parts' *hòj* and *kɔ̌ŋ* are functioning as nouns here.

Looking at (280) in comparison to (281), the numeral-classifier expression can be separated from the noun. As (282) shows, the adjective cannot be similarly separated.

(280) wéŋ jàj sɔ̌ŋ wéŋ thùknjáa hét hâaj
city big two city PASS do injure
'Two big cities were destroyed.'

(281) wéŋ jàj thùknjáa hét hâaj sɔ̌ŋ wéŋ
city big PASS do injure two city
'Two big cities were destroyed.'

(282) *wéŋ thùknjáa hét hâaj sɔ̌ŋ wéŋ jàj
city PASS do injure two city big
intended: 'Two big cities were destroyed.'

Adjectives cannot be directly selected by *mí* or *pěn*. There must be an overtly represented noun that the adjective modifies. I take this to indicate that adjectives are properly predicate adjuncts of nouns, as such, it is not possible to elide the modified head. The grammatical relationship between the numeral-classifier expression and the noun must be different. If the structure was identical to adjective adjunction, why would it not pattern the same in terms of eliding the noun?

A comparison with Ch'ol numeral-classifiers

Bale et al. (2019) propose the complement of the analysis proposed here: an analysis of Ch'ol classifiers where the syntactic structure involves the classifier forming a constituent with the numeral no matter whether a 'measure' or 'count' reading is being expressed.

Evidence that NUM+CLF form a constituent The evidence they use to support this proposal are the following:

1. Complex numeral structure
2. Juxtaposition (coordination)

3. A′-movement

Ch'ol forms complex numerals with the structure in (283). Here the 'classifier' meaning twenty is marked with agreement that is usually found on possessed or relational nouns.

(283) cha'-p'ej i-cha'-k'al
two-CLF A3-two-CLF.20
'twenty-two' (lit. two of the group of two-20s)

(Ch'ol, Bale et al. 2019: (46a); citing (Warkentin & Scott 1980: 108))

In Ch'ol [NUM+CLF] can be juxtaposed to express an indefinite quantity, as in (284).

(284) wajali am-bi li [juñ-tyikil cha'-tyikil] la-k-pi'äl.
back.then EXT-REP DET one-CLF two-CLF PL-A1-friend
'It is said that back then we had some friends.' (Ch'ol, Bale et al. 2019: (47a))

In A′-movement constructions, the [NUM+CLF] constituent can move away from the noun, but the numeral cannot move away from the classifier and noun, as shown in (285)–(286).

(285) Ux-tyikil$_i$ ta' jul-i-y-ob [_$_i$ x'ixik].
three-CLF PFV arrive-ITV-EP-PL woman
'*Three* women arrived.' (Ch'ol, Bale et al. 2019: (49a))

(286) *Ux$_i$ ta' jul-i-y-ob [_$_i$ tyikil x'ixik].
three PFV arrive-ITV-EP-PL CLF woman
'*Three* women arrived.' (Ch'ol, Bale et al. 2019: (49b))

There is no evidence of complex numeral structures like (283) in Shan. Examples such as (284) would be expressed by juxtaposing two numerals rather than numeral-classifier phrases. It is possible to have structures like (173), repeated in (287) in Shan. This could be taken as evidence that the numeral and classifier form a constituent to the exclusion of the noun. However, the numeral-classifier expression must appear at the right side of the clause. It cannot be moved to a higher position than the VP.

(287) phujíŋ hɔtphɛ̌w kéŋmàj jâw sǎam kɔ̂
woman arrive Chiang_Mai PFV three CLF.HUM
'Three of the women arrived in Chiang Mai.'

Additionally, the numeral-classifier expression is the only things that can appear displaced from the noun in Shan, while both adjectives and numeral-classifier expressions can move away from the noun in Ch'ol (Bale et al. 2019). Shan numeral-classifier expressions are different from adjuncts like adjectives. Ch'ol numeral-classifier expressions and adjectives have similar distributions.

Evidence that the measure phrase is an adjunct A modifier, such as an adjective, can appear on either side of the noun in Ch'ol, which Bale et al. (2019) take to indicate that the noun is the head and the [NUM+CLF] adjoin higher up. The Ch'ol measure phrase shares certain distributions with other modifiers. For example, just like other predicates like (288a)–(288b) it can appear as a predicate in the 'non-verbal predicate' constructions, as in (288c).

(288) a. Kolem jiñi xiñich'.
big DET ant
'The ant is big.'

b. Buch-ul jiñi x'ixik.
seated-STAT DET woman
'The woman is seated.'

c. Ux-tyikil jiñi x'ixik-ob.
Three-CLF DET woman-PL
'The women are (in a group of) three.'

(Ch'ol, Bale et al. 2019: (61))

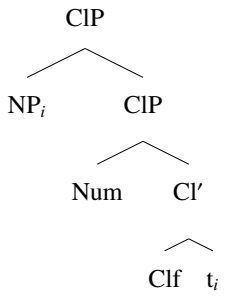
In contrast, the Shan numeral-classifier expressions have a different distribution than adjectives, as was established in the previous section. It cannot function as a predicate in the same way that adjectives can, and the adjective always appears closer to the noun than the numeral-classifier does.

There is no evidence along the lines of that proposed by Bale et al. (2019) for Ch'ol to support analyzing the Shan numeral-classifier expression as an adjunct modifier. To the contrary, the numeral-classifier expression does not seem to be able to have the status of a predicate.

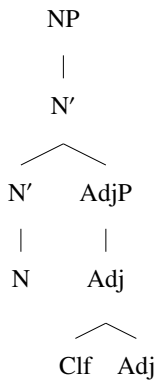
Proposed structures

I have argued that the structure for the [N Num Clf] construction should be the one in (289). I have also argued that the Clf-Adj construction is a compound predicate, which can function as a noun adjunct with the structure in (290). This explains why the classifier can appear in both the Clf-Adj and another classifier position in the same noun structure. To avoid violating anti-locality, the roll-up movement could involve movement to agreement phrases between the labeled functional categories here in the sense of Cinque 2005.

(289) N NUM CLF



(290)



3.3.4 Summary

This section has discussed classifier expressions found in Shan. All classifiers can appear in a similar range of syntactic positions, but not at the same time. This supports proposing a single functional classifier head. Most kinds of classifiers, with the exception of measure terms, function as nouns in some contexts. A set of classifiers are 'repeater classifiers' where the noun and its associated classifier have the same form. This results in making available a type of NP-deletion, licensed by the homophonous classifier. More common anaphoric NP-ellipsis can be seen with floated quantifiers and numeral-classifier expressions. Evidence from constituent order and selection also support treating the classifier as a functional head that selects its associated NP. The following section will propose a semantic analysis of these numeral-classifier constructions.

3.4 Classifier Semantics

The previous section demonstrated that classifiers, collectives, and measure terms function similarly syntactically. This section will argue that they also have a shared semantic function: to provide a unit of measurement by which a noun can be individuated. Although there is a semantic distinction between mass and count nouns in this numeral-classifier, number-neutral language, there is no real syntactic distinction between how those are measured/counted. Since there is a unified semantic function and no significant syntactic difference, I provide a unified semantics for the individuating use of classifiers, collectives, and measure terms. This will facilitate cross-linguistic comparison of count/measure expressions.

In some languages plurality is not morphologically expressed. There is reason to think that Shan is a language that does not need to morphologically mark plurality, as discussed in chapter 2. If you just say *măa*, as in (291), it could describe one or more dogs.[13] Whether you are talking about one dog (292) or two dogs (293), the noun *măa* does not change.

(291) măa hàw jù.
dog bark IPFV
'A dog is barking.'

'Dogs are barking.'

(292) háw mí măa nɯŋ tǒ
1 have dog one CLF
'I have one dog.'

(293) háw mí măa sɔ̌ŋ tǒ
1 have dog two CLF
'I have two dogs.'

The contrast between (294) and (295) show that the individual classifier *tǒ* in (295) forces a singular interpretation of the noun. (294) does not include a classifier, so the interpretation could be singular or plural.

[13] As discussed in another chapter, these nouns are also under-specified for definiteness, so definite interpretations are also available here.

(294) măa nân hàw jù.
dog that bark IPFV
'That dog is barking.'

'Those dogs are barking.'

(295) măa tŏ nân hàw jù.
dog CLF.ANML that bark IPFV
'That dog is barking.'

~~'Those dogs are barking.'~~

A noun in a classifier language is often thought to denote a set of atoms and sums of atoms, depicted in figure 3.1, and the classifier functions to create a set of the atomic parts (the lowest level of the figure). The semantic denotation is given in (296).[14] A common analysis of classifiers is that they function to atomize or quantize a noun (Jenks 2011; Dayal (2012); Nomoto (2013)). Nomoto's (2013) semantics for individual classifiers can be seen in (297). This is consistent with the data in (294)–(295). The result of combining the noun denotation in (296) with the classifier denotation in (297) would be the quantized noun denotation in (298).

(296) $[\![\text{N}]\!] = \lambda x.{}^{*}\text{NOUN}(x) = \{a, b, c, d, ab, ac, ad, bc, bd, cd, abc, bcd, acd, abd, abcd\}$

(297) $[\![\text{CLF}\ ]\!] = \lambda P \lambda x.P(x) \wedge \neg\exists y[P(y) \wedge y \leq x]$

(298) $[\![\ \text{N}_{\text{SG}}]\!] = \lambda x.\text{NOUN}(x) = \{a, b, c, d\}$

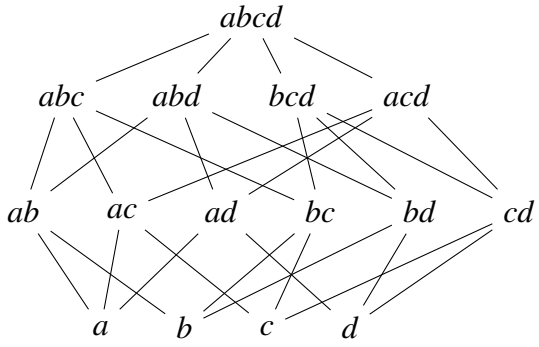


Figure 3.1: Complete atomic join semi-lattice

I propose that the semantics of classifiers, group terms, and measure terms in classifier constructions be unified. While the semantics of each kind of classifier might differ in meaningful ways, I would like to argue that the basic notion of counting and measuring are the same. A key

[14] Another common assumption for classifier languages is that nouns denote kinds. Presumably, this analysis could be adapted to use that assumption.

aspect of this is assuming that atomicity is at the core of all counting and measuring. The intuition behind this is that in order to count something, you have be able to identify a item or unit by which to count it. This can be formalized by having the basic measure function presuppose the atomicity of the component parts of an entity, which could be written semantically as (299) or (300).

(299) $[\![\mu]\!] = \lambda x : \exists y[y \leq x \wedge \neg\exists z[z \leq y \wedge z \neq y]].|y : y \leq x \wedge \neg\exists z[z \leq y \wedge z \neq y|$

This presupposes that there is some y that is a sub-part of the input x such that there is no z that is a sub-part of y that is not equal to y. Then it returns the number of such y in x with that property.

(300) $[\![\mu]\!] = \lambda x : \exists y[y \leq x \wedge AT(y)].|y : y \leq x \wedge AT(y)|$

This is a short-hand way of writing (299) that just presupposes that there are y that are sub-parts of x that are atoms and returns the number of those atoms.

This ties into the discussion of the distinction between count and mass nouns. Deal (2017) claims that a count/mass distinction of some type exists in all languages. Following Deal's (2017) notation, I assume that in Shan, like in Mandarin Chinese, all nouns are cumulative, meaning that there is no equivalent of the quantized singular count noun in English. The distinction in mass and count nouns in such languages comes in 'the nature of their minimal parts' (Deal 2017: p. 9).

The proposed semantics of the other lexical items are given in (301)–(303).[15]

(301) $[\![m\check{a}a]\!] = \lambda x.{}^{*}\textsc{dog}(x)$

(302) $[\![t\check{o}]\!] = \lambda P\lambda x.P(x) \wedge \mu_{\textsc{body}}(x) = 1$

[15]The semantics I proposed for the demonstrative in (148) in chapter 2 is repeated in (1). The semantics for nouns is given in (2). I am assuming that the situation variable has already saturated when it combines with a classifier. For simplicity, I am using a simplified version of these in this chapter.

(1) $[\![n\hat{a}n]\!] = \lambda P : |P_s| = 1.\iota x[P_s(x) \wedge \textsc{far.from.speaker}(x)]$

(2) $\lambda w\lambda x.\textsc{dog}_w(x)$

.

(303) $[\![nân]\!] = \lambda P.\iota x[P(x)]$

Table 3.7 shows the interpretations of N (Clf) Dem. The classifier creates a singular interpretation.

Table 3.7: N Clf Dem interpretation

Shan	Translation	CLF	Interpretation
măa tŏ nân	'That dog'	individual	$\iota x[^*\text{DOG}(x) \wedge \mu_{\text{BODY}}(x) = 1]$
măa nân	'That/Those dog(s)'	∅	$\iota x[^*\text{DOG}(x)]$

The semantics of (294) is given in (304), and the semantics of (295) is shown in (305).

(304) $[\![(294)]\!] = \text{BARK}(\iota x[^*\text{DOG}(x)])$

(305) $[\![(295)]\!] = \text{BARK}(\iota x[^*\text{DOG}(x) \wedge \mu_{\text{BODY}}(x) = 1])$

3.4.1 Classifiers and collectives

Stubbornly distributive predicates (Schwarzschild 2007; Schwarzschild 2011), also identified by Moltmann (2004), are expressions like *big* which tell us something about what counts as singular or 'atomic' in a particular context. Thus they can be used to identify the size of the units that Shan classifier and group constructions generate. When a predicate applies to a plurality, the predicate might be true of the collective plurality or the individuals that make up the plurality, as with *heavy* in (306a). Stubbornly distributive predicates, on the other hand, are only true of the individuals, not the collective, as with *large* in (306b).

These can be distinguished if we consider whether the predicates can apply to individual boxes as in (307) for a distributive reading or to a collection of boxes as in (308) for a collective reading.

(306) Predicates and stubbornly distributive predicates

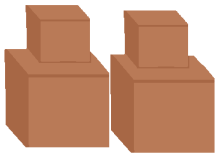
a. The boxes are heavy. (Schwarzschild 2011: (1))

What is heavy? the collection of boxes ✓, the individual of boxes ✓

b. The boxes are large. (Schwarzschild 2011: (8))

What is large? the collection of boxes ✗, the individual of boxes ✓

(307) 

(308) 

In Shan, the predicates that correspond to typical stubbornly distributive predicates in English tend to have the complex Clf Adj structure described in section 3.3.3. I would argue that this is not by chance, but rather demonstrates the connection between classifiers and distributivity. The classifier here overtly represents the unit to which the distributive adjective applies.

The adjectives that appear in these constructions represent physical properties, such as length as in (309), size as in (310), or shape as in (311). Moltmann (2004) noted that certain predicates related to size must have a distributive interpretation with a noun that expresses a plurality. Similarly, stubbornly distributive predicates are identified as predicates that relate to 'size, shape and duration' (Schwarzschild 2011: 4).

(309) sǎajkhɔ́ (sen) jáaw nɯŋ sen
necklace string long one string
'one long necklace'

(310) ʔǎn-nâj pěn mǎa (tǒ) jàj. ʔǎn-nân pěn mǎa (tǒ) lêk.
CLF.GN-this COP dog (CLF.ANML) big CLF.GN-that COP dog (CLF.ANML) small
'This is a big dog. This is a small dog.'

(311) háw mí màakkhɔ̌ hòj món sǎam hòj lɛ màakkhɔ̌ hòj laaw sɔ̌ŋ
1 have jujube CLF.RND round three CLF.RND and jujube CLF.RND egg-shaped two
hòj
CLF.RND

'I have three round jujube and two egg-shaped jujube.'

In (309), the classifier *sen*, which is used for necklaces and other string-shaped things, appears before the adjective *jáaw* 'long'. This is in addition to the classifier *sen* that is obligatory with the numeral. In (310), the classifier *tǒ* for animals optionally appears between the noun and the size-related predicates *jàj* 'big' and *lêk* 'small'. In (311), the classifier for round objects, *hòj* appears before the adjectives *món* 'round' and *laaw* 'egg-shaped' that describe the shape of the noun.

It is unlikely to be a coincidence that the types of predicates that tend to be stubbornly distributive are also the ones that have this optional morpheme, which is associated with distributivity and singularity in other contexts. Therefore, I propose that the classifier in this construction overtly specifies the level of distributivity that is required by this category of adjectives. We might expect to find classifiers used in these contexts in other languages, as well. It is possible to have Clf Adj constructions in Thai, as noted by Hundius & Kölver (1983), Piriyawiboon (2010), Jenks (2011), among others. However, the usage in Thai seems to much more generalized in comparison to Shan, since classifiers can appear with a wide variety of adjectives. Jenks (2012) posits two distinct categories [Clf Adj] constructions. One category is connected to 'physical' adjectives, and seems parallel to the Shan [Clf Adj] cases discussed here. He analyses these constructions as complex predicates. The other category, Jenks (2012) calls 'classifier modifier constructions', and this includes [Clf Adj] constructions that consist of a broader category of adjectives. Shan only seems to have the first category.

(312) mă̆a tua jàj
do CLF big
'(a/the) big dog(s)'

(313) náam khùat lék
water bottle small
'a/the small bottle of water'

(Thai, Jenks (2012): Type(1)a,b)

(314)–(317) all have subjects with structure N Clf Dem and predicates with the structure (Clf) Adj. The Clf is either (i) the individual classifier—*tǒ* for animals (CLF.ANML) in (314) or *hòj* for round objects (CLF.RND) in (315); or (ii) a group term—*phŭ̌ŋ* for groups of animals in (316) or *kɔ̌ŋ* 'pile' in (317). (161), repeated in (318) shows an example where the Clf Adj is modifying the

noun rather than acting as a predicate. In such constructions, the Clf Adj expression must appear immediately to the right of the noun, before any numeral-classifier expression or demonstrative.

(314) măa tŏ nân (tŏ) jàj nàa
dog CLF.ANML that (CLF.ANML) big very
'That dog is very big.'

(315) màak.khŏ hòj nân (hòj) jàj nàa
jujube CLF.RND that CLF.RND big very
'That jujube is very big'

(316) măa phŭŋ nân jàj nàa
dog GROUP.ANML that big very
'That group of dogs is very big.'

(317) màak.khŏ kŏŋ nân jàj nàa
jujube pile that big very
'That pile of jujube is very big'

(318) măa tŏ jàj săam tŏ tsɤ́ ʔăn lâk kǐn kàj (*tsɤ́) nân ...
dog CLF.ANML big three CLF.ANML PL COMP steal eat chicken PL that
'The three big dogs that stole and ate the chicken...'

(314) and (315) have an individual classifier in the Clf position of the [N Clf Dem] subject, so the predicate *jàj* 'big' describes the individual jujube or dog. In (316) and (317) the Clf position is filled by a group term, so what is big is the group. In the N Clf Dem configuration, the word in the classifier position can determine what unit is being described as 'big'.

Notice that the predicate adjective can also have an associated classifier in (314) and (315). As has been noted in section 3.3.3, the Clf Adj construction has a different syntax than the other classifier constructions, which is why a classifier can appear in that construction even when another classifier is already present in the nominal structure, as in (318). The classifier is optional in (314) and (315) because the unit of measurement for the distributive predicate has already been specified and the basic noun denotations of these count nouns already includes atoms. However, if the subject refers to a group or collective, an individual classifier with the predicate adjective can still specify the unit of measurement for the adjective.

(319) and (320) have the same collective subjects as (316) and (317), but in the predicate there are the individual classifiers *hòj* and *tŏ* that appear before the predicate *jàj*. These classifiers specify that the predicate applies to the individual components of the collective rather than the group. When a classifier appears with the distributive adjective in a predicate position, the classifier that

immediately precedes the adjective determines what unit the adjective applies to. This suggests that, while the Clf Adj syntax is different than the other classifier structures, the classifier semantics is still connected to specifying a unit for measurement.

(319) màak.khɔ̌ kɔ̌ŋ nân hòj jàj nàa
jujube pile that CLF.RND big very
'That pile of jujube, the jujube are big.'

(320) mǎa phǔŋ nân tǒ jàj nàa
dog GROUP.ANML that CLF.ANML big very
'That group of dogs, the dogs are very big.'

Mass nouns like *nâm* 'water' in (322), unlike count nouns as in (321), cannot be modified by *jàj* unless they are associated with a container term as in (323). Classifiers, group terms, and container terms can specify the unit being measured by distributive predicates like *jàj* 'big'. Mass and count nouns differ in whether they have inherent atomic parts.

(321) mǎa nân jàj nàa
dog that big very
'That/Those dogs is/are big.'

(322) #nâm nân jàj nàa
water that big very
intended: 'That water is big.'

(323) nâm kɔ́k nân jàj nàa
water cup that big very
'That cup of water is big.'

The noun *khaw* 'rice' appears to be another noun that cannot appear with a stubbornly distributive predicate, as (324) shows. (325) has the same noun with a regular predicate, showing that it is the predicate choice causing the infelicity of (324).

(324) #khaw nân jàɯ nàa.
rice that big very
intended: 'That rice is very big.'

(325) khaw nân lǐ nàa.
rice that good very
'That rice is very good.'

Predicates like *jàj* are similar to numerals in Shan in that classifier/group/measure terms specify the relevant unit being measured. They are different in that distributive predicates can access inherent atomic parts even in the absence of a classifier. Otherwise, we might expect a classifier to be obligatory with all stubbornly distributive predicates, just as they are with numerals, which is not the case. Thus, the classifier is only obligatory in cases like (323) where the noun does not have inherent atomic parts.

Another difference between distributive predicates like *jàj* and numerals is that the predicate is associated with a physical property and therefore connected to the ability to refer. As has been mentioned previously, measure terms are not referential in the same way as other classifiers. As a result, it sounds unnatural to use the predicate *jàj* when the subject has a measure term in the classifier position for the N Clf Dem structure as in (326). It sounds better to use a predicate related to amounts, as in (327).

(326) #ʔó màakmâj kìlò nâj jàj nàa. tsɔ̀ŋ mɛn màakkhɔ̌ hâa
oh! fruit kilo this big very POLAR.Q be.right jujube POLAR.Q
intended: Oh, this kilo of fruit is very big. Is it jujube?

(327) ʔó màakmâj nɯŋ kìlò nâj laj nǎm nàa. tsɔ̀ŋ mɛn màakkhɔ̌ hâa
oh! fruit one kilo this get many very POLAR.Q be.right jujube POLAR.Q
'Oh, this kilo of fruit has a lot. Is it jujube?'

3.4.2 A semantic analysis of Shan classifiers

The presence of a classifier that specifies a group or container will also specify the level that the predicate distributes down to. Suppose there were two groups of dogs: one group, *j*, was made up of dogs *a* and *b*, and one group, *k*, was made up of *c* and *d*. This can be visualized as in figure 3.2.

This resembles Link's (1983) materialization function, which maps an individual to its material parts. For the individuals in figure 3.2, the materialization function *h* can describe the relationship

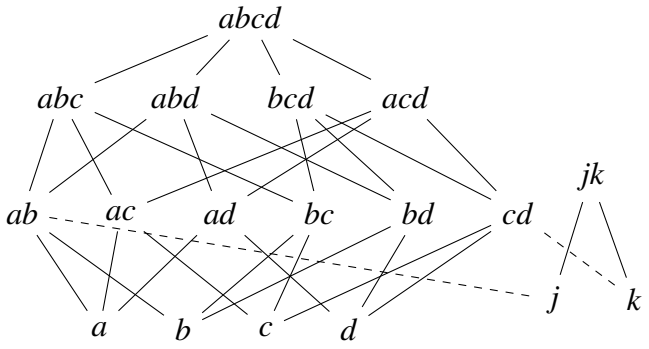


Figure 3.2: Group and parts of groups

between the two lattices as in (328).[16] Then, the function of the classifier is to individuate based on the classifier used, with an assumed material equivalency between the noun and the classifier unit. (328) just means that the *material/stuff* that comprises the plurality of dogs $a \bigoplus b$ is identical to the *material/stuff* that comprises the group j. This can be expressed as shown in (329).

(328) $h(a \bigoplus b) = h(j)$

The material that makes up the sum of individual a and b is equivalent to the material that makes up group j.

(329) $[\![phǔŋ_{\text{CLF}}]\!] = \lambda P \lambda x.\exists y[P(y) \wedge \mu_{\text{GROUP.A}}(x) = 1 \wedge h(y) = h(x)]$

The set of x that measure one group such that there exists some y with property P that is made up of the same material as x.

This formal semi-lattice homomorphism might be independently needed for group terms in general. The semantics of the group term could be modeled as in (330). Then, the function of the classifier position is still to quantize the interpretation.

(330) $[\![phǔŋ_{\text{N}}]\!] = \lambda P \lambda x.\exists y[P(y) \wedge {}^{*}\text{GROUP.A}(x) \wedge h(y) = h(x)]$

Then it is possible to distinguish between the examples in (331) and (332). The derivation for

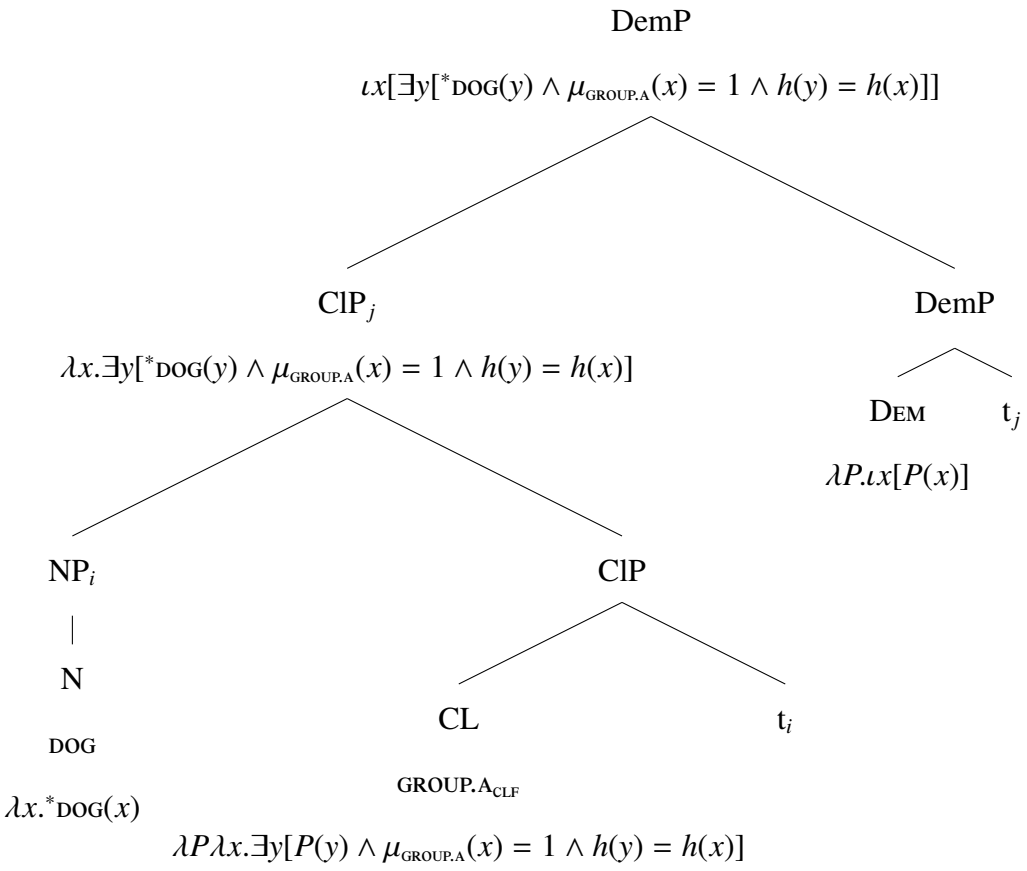
[16] In the figure I write $a \bigoplus b$ as ab to save space.

(331) is given in (333)[17] and the derivation for (332) is given in (334). The difference between the two is that (333) denotes a single group of dogs, while (334) could denote one or more groups of dogs.

(331) N CLF DEM
mǎa phǔŋ nân
dog group that
'that group of dogs'

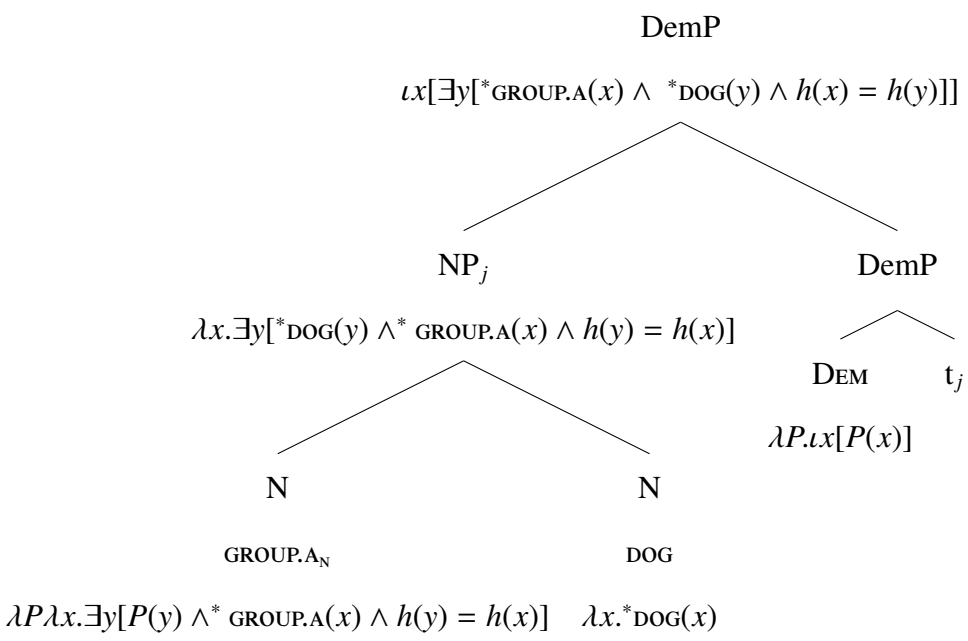
(332) N N DEM
phǔŋ mǎa nân
group dog that
'that/those group(s) of dogs'

(333)



[17]Here (and in the whole dissertation), I am ignoring how traces and lambda extraction would figure into this for simplicity and assuming that the classifier and noun combine at the base position of the noun.

(334)



With this modification, the interpretation of the N (Clf) Dem constructions can be seen in table 3.8. This includes the case where the Clf is the group classifier *phŭŋ*. For individual classifiers like *tŏ*, I am going to assume that they pick out atoms of the noun they combine with, rather than connecting to a separate lattice. Thus, μ_{BODY} is really μ_{DOG} in this case.

Table 3.8: N Clf Dem interpretation

Shan	Translation	CLF	Denotation
măa tŏ nân	'that dog'	individual	$\iota x[^*\text{DOG}(x) \wedge \mu_{\text{BODY}}(x) = 1]$
măa nân	'that/those dog(s)'	∅	$\iota x[^*\text{DOG}(x)]$
măa phŭŋ nân	'that group of dogs'	group	$\iota x[\exists y[^*\text{DOG}(y) \wedge \mu_{\text{GROUP.A}}(x) = 1 \wedge h(x) = h(y)]]$
phŭŋ măa nân	'that/those group(s) of dogs'	∅	$\iota x[\exists y[^*\text{GROUP.A}(x) \wedge {}^*\text{DOG}(y) \wedge h(x) = h(y)]]$

3.4.3 Clf Adj

To look at the contribution of the classifier in Clf Adj constructions, it is necessary to have a semantics for the distributive adjectives. As discussed in a previous section, the Clf Adj construction

is best analyzed as a complex predicate.[18] Semantically this could be handled in two ways, one is to assume a complex predicate interpretation like 'large-bodied' for *tǒ jàj*. Another way would be to assume a similar classifier semantics, such as in (335).[19] It is possible that both types of complex predicates are available in the language. One benefit of the classifier-based semantics is that it can rely on the 'unit' measure component of classifier semantics to specify the level for the distributive adjective to apply to. The denotations in (336) gives the semantics for the distributive adjective *jàj* 'big' where the unit specifying portion is overtly specified.

(335) $[\![tǒ]\!] = \lambda x.\mu_{\text{BODY}}(x) = 1$

(336) $[\![jàj]\!] = \lambda P \lambda x : \exists y \exists \mu[y \leq x \wedge \mu(y) = 1].\exists y[y \leq x \wedge P(y) \wedge \text{BIG}(y)]$

(336) takes a classifier argument P which specifies the unit that the distributive predicate describes. Then, it returns a predicate consisting of the x that are made up of parts y that are atoms of the measure term P, where y has the property of being big. Importantly, what is being described as 'big' is the unit of the classifier itself. For count nouns, this can be equivalent to the individual, but for mass nouns, it describes the container. The presupposition disallows this predicate from combining with a mass noun since it does not have identifiable atomic parts.

The derivation of (314), repeated in (337), is given below.[20]

(337) mǎa tǒ nân (tǒ) jàj nàa
dog CLF.ANML that (CLF.ANML) big very
'That dog is very big.'

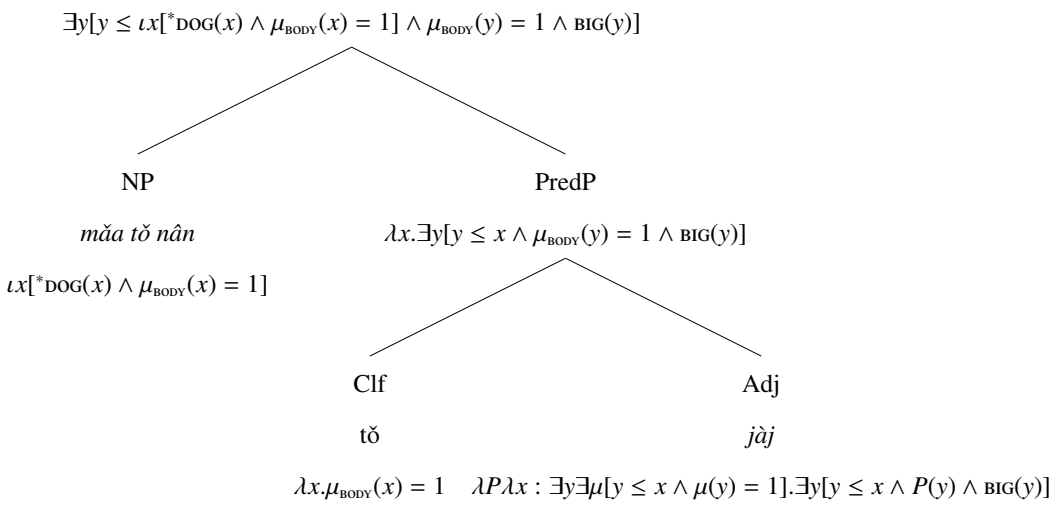
The semantic derivation of (314) proceeds as in (338), where the classifier appears before the adjective. Since the set with the property μ_{BODY} in this context is just the quantized set of dogs, the adjective applies to the individual dogs.

[18] I am entirely ignoring certain features of this predicate, such as it being scalar and involving contextual comparisons.

[19] This is a different syntactic position, so it might also have a slightly different semantics. Here, I just changed the type.

[20] The word *nàa* just means 'very' and will not be included in this semantics.

(338)



In contexts where a distributive adjective does not follow a classifier, the unit of distribution is inferred from the context. The semantics for this adjective can be seen in (339).

(339) $[\![jàj_2]\!] = \lambda x : \exists y \exists \mu[y \leq x \wedge \mu(y) = 1].\exists y[y \leq x \wedge \mu_C(y) = 1 \wedge \text{BIG}(y)]$

(320) and (316) have the group term in the classifier position of the subject. In (316), there is no compound adjective, and in (320), there is.

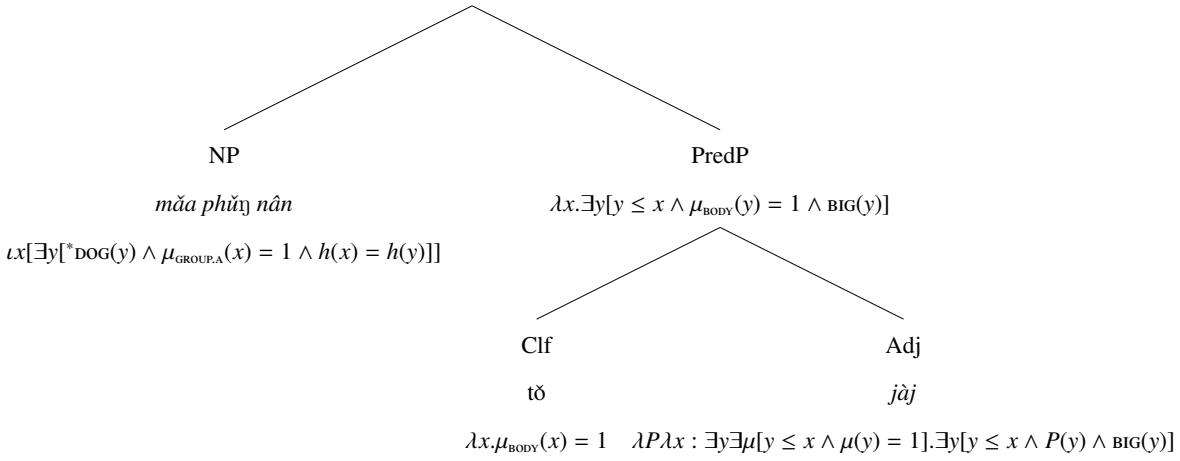
(320) măa phŭŋ nân tǒ jàj nàa
dog GROUP.ANML that CLF.ANML big very
'That group of dogs, the dogs are very big.'

(316) măa phŭŋ nân jàj nàa
dog GROUP.ANML that big very
(i) 'That group of dogs is very big.'

(ii) 'That group of dogs, the dogs are very big.'

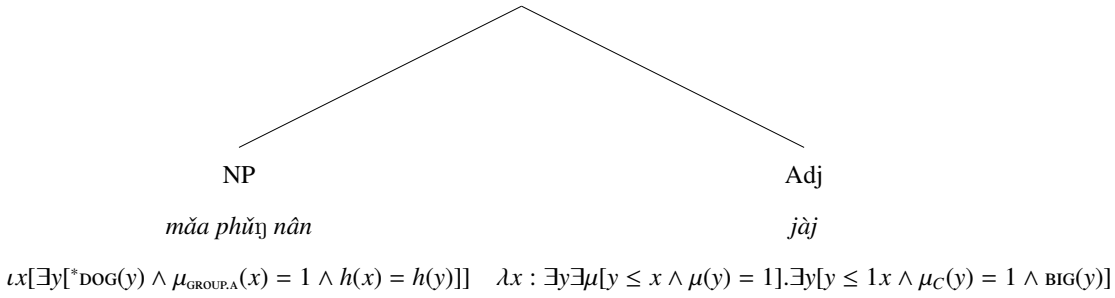
The semantic derivations of (320) and (316) proceed as in (340) and (341), respectively.

(340) $\exists y[y \leq \iota x[\exists y[^{*}\textsc{dog}(y) \wedge \mu_{\textsc{group.a}}(x) = 1 \wedge h(x) = h(y)]] \wedge \mu_{\textsc{body}}(y) = 1 \wedge \textsc{big}(y)]$



(340) works just like (338), except that there is a separate lattice specifying that the dogs comprise one group. The DOG lattice still needs to be accessible for the predicate $\lambda y.\mu_{\textsc{body}}(y) = 1 \wedge \textsc{big}(y)$ to apply. (341) lacks the classifier in the Clf Adj construction, so the unit that the distributive adjective applies to is inferred, represented by μ_C. When μ_C is interpreted as GROUP.A, meaning (i) of (316) results. When it is interpreted as BODY (DOG), meaning (ii) results.

(341) $\exists y[y \leq \iota x[\exists y[^{*}\textsc{dog}(y) \wedge \mu_{\textsc{group.a}}(x) = 1 \wedge h(x) = h(y)]] \wedge \mu_C(y) = 1 \wedge \textsc{big}(y)]$



The GROUP.A lattice is only accessible for μ_C if it has been overtly included. In contrast $\mu_{\textsc{body}}$ represents the atomic parts of DOG, so it is inherently accessible for count nouns.

3.4.4 Properties of the plural morpheme

The previous section focused on the semantics of classifiers and group terms, particularly looking at how they pattern with respect to stubbornly distributive predicates. This section compares group terms in Shan to the plural classifier morpheme and argues that they are semantically distinct expressions. I propose that group terms form collectives and the plural classifier forms a plurality.

Classifier languages often lack obligatory plural morphology on nouns (as noted by Greenberg 1972; Chierchia 1998; a.o.), but many of these languages have a lexical plural morpheme, such as Mandarin (e.g., Cheng & Sybesma 1999, Bošković & Hsieh 2012), Japanese (e.g., Ueda & Haraguchi 2008), and Korean (e.g., Lee 1992).

Shan has a lexical plural morpheme *tsɣ́*, which marks plurality in definite contexts or functions as a numeral classifier for groups. It can appear with humans, animals, and objects. This morpheme can appear in many of the same positions as numeral classifiers (e.g., *tŏ* for animals). Just like the individual classifier in (342), in (343) the plural morpheme can appear with a numeral. The plural morpheme in (343) is used for counting how many distinct pluralities of dogs there are. This is different from identifying sub-kinds, for example. The classifier *méw* 'sort'/'kind' is typically used to count sub-kinds. The pluralities described by (343) have some shared properties, but they might be transient properties, such as 'are sleeping', or properties related to size or color.

(342) măa săam tŏ
dog three CLF
'three dogs'

(343) măa săam tsɣ́
dog three *tsɣ́*
'three sorts of dogs'

In (345), the plural morpheme appears with a demonstrative in the same position that the individual classifier does in (344). In (347), it appears with a relative clause in the same position that the individual classifier does in (346). Everywhere that a classifier can appear, distinct from its use as a noun, the morpheme *tsɣ́* can also appear.

(344) măa tŏ nâj
dog CLF this
'this dog'

(345) măa tsɤ́ nâj
dog *tsɤ́* this
'these dogs'

(346) măa tŏ ʔăn nɔ́n jù
dog CLF COMP sleep IPFV
'the dog that is sleeping'

(347) măa tsɤ́ ʔăn nɔ́n jù
dog *tsɤ́* COMP sleep IPFV
'the dogs that are sleeping'

Based on whether the plural morpheme can (Japanese, Korean; Kim & Melchin 2018) or cannot (Armenian; Borer 2005) co-occur with numeral classifiers, previous work has discussed whether the plural morpheme appears in the same functional head as the classifier. On the surface, the Shan plural morpheme can co-occur with a numeral and classifier shown in (348). An individual classifier cannot appear in the position of *tsɤ́* when there is already another classifier in the nominal construction (with the exception of the Clf-Adj construction).

(348) màak-khɔ̆ síp hòj tsɤ́ nâj
fruit-jujube ten CLF.RND PL this
'these ten jujube.'

The classifier-associated syntactic positions are not the only places where *tsɤ́* can appear. The plural morpheme is something like a noun meaning 'group' that can form a compound with a noun that has already combined with a numeral. (349) gives an example where *tsɤ́* combines with the noun *luk-hén* 'students'. (350) shows that it can also appear before a pronoun.

(349) tsɤ́ luk-hén
tsɤ́ child-study
'(group of) students'

(350) tsɤ́ khăw
tsɤ́ 3.PL
'they'

The co-occurrence of the nominal and classifier version of *tsɤ́* in (351) shows that these are distinct uses.

(351) (tsɤ́) màak-khɔ̆ síp hòj tsɤ́ nâj
PL fruit-jujube ten CLF PL this
'these ten jujube.'

The rest of this chapter looks more closely at the relationship between the noun use and classifier use of the plural morpheme and other classifier terms.

A comparison with group terms

The morpheme *tsɤ́* is not the only classifier-like word that can co-occur with an individual classifier, as (352) shows. In (352), the first two words *kɔ̌ŋ màak*.khɔ̌ 'pile of jujube' form a compound noun and are found in the typical position for nouns. Then, *síp hòj* 'ten round ones' is a numeral-classifier expression containing the individual classifier *hòj* for round objects. Then, the word *kɔ̌ŋ* 'pile' appears again in the classifier position that is associated with the following demonstrative. This indicates that there is only one pile under discussion.

(352) (kɔ̌ŋ) màak.khɔ̌ síp hòj kɔ̌ŋ nâj
(pile) jujube ten CLF.RND pile this
'this pile of ten jujube.'

The word *kɔ̌ŋ* 'pile' can also be used in the same positions as individual classifiers, as (353)–(358) demonstrate. (353)–(354) show the N Num Clf construction. (355)–(356) show the N Clf Dem construction, and (357)–(358) show the N Clf relative clause construction. In (357)–(358), *ʔǎn* is the relative clause complementizer.

(353) màak.khɔ̌ sǎam hòj
jujube three CLF.RND
'three jujube'

(354) màak.khɔ̌ sǎam kɔ̌ŋ
jujube three pile
'three piles jujube'

(355) màak.khɔ̌ hòj nân
jujube CLF.RND that
'that jujube'

(356) màak.khɔ̌ kɔ̌ŋ nân
jujube pile that
'that pile of jujube'

(357) màak.khɔ̌ hòj ʔǎn súk jâw nân
jujube CLF.RND COMP ripe PFV that
'the jujube that is ripe'

(358) màak.khɔ̌ kɔ̌ŋ ʔǎn súk jâw nân
jujube pile COMP ripe PFV that
'the pile of jujube that is ripe'

Table 3.9 summaries the distribution of the classifier, group/container term, and plural morpheme. In terms of structure, the plural morpheme can appear in all the same positions as other classifiers, though it seems to have a slightly different meaning in numeral-classifier constructions—i.e., N Num tsɤ́ —where it means something like 'Num types of N'.

Structure	Classifier	Plural	Group
N Num _	✓	✓	✓
N _ Dem	✓	✓	✓
N _ RC	✓	✓	✓
N Num Clf_1 _ Dem	✗	✓	✓

Table 3.9: Classifies, the plural morpheme, and group terms

As demonstrated in (161) and (348), both repeated here, the plural classifier can co-occur with individual classifiers. However, there are restrictions. The morpheme *tsɤ́*, like individual classifiers, can only appear once in a classifier position within a nominal structure.

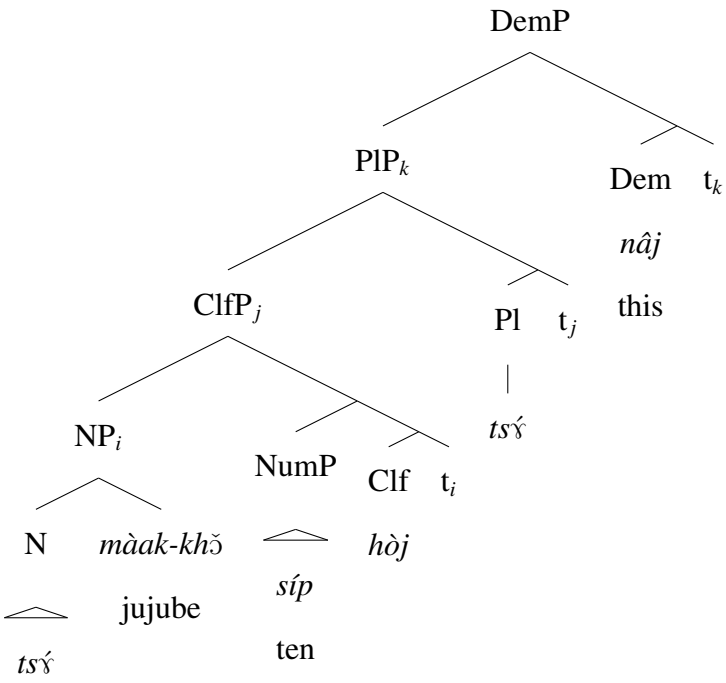
(161) mǎa tǒ jàj sǎam tǒ tsɤ́ ʔǎn lâk kǐn kàj (*tsɤ́) nân ...
dog CLF.ANML big three CLF.ANML PL COMP steal eat chicken (*PL) that
'The three big dogs that stole and ate the chicken...'

(348) (tsɤ́) màak-khɔ̌ síp hòj tsɤ́ nâj
(PL) fruit-jujube ten CLF PL this
'these ten jujube.' (Plural)

In (352), 'pile' is a measure term, and the initial noun 'pile' forms a compound with 'ten jujube'. In (342)–(346), *tsɤ́* has the function of a classifier or measure term and thus has the same distribution as one. Jenks (2011) also describes the Thai (Southwestern Tai) plural morpheme as a measure word.

(359) gives the structure of (348). The Pl position can be filled by any classifier terms that refer to pluralities.

(359)



For Armenian, Borer (2005) claimed that the lexical plural morpheme is in the same functional phrase as the classifier. Kim & Melchin (2018) argued that Borer's (2005) analysis cannot explain languages like Japanese and Korean where the plural appears with a numeral and classifier. It seems here that is is necessary to posit an extra functional phrase to be able to support the plural classifier. However, given that—aside from its co-occurrence with individual classifier —it patterns the same as individual classifiers, the functional phrase must be nearly identical except for differences in plurality.

Distributive plural morpheme

This section uses tests from Schwarzschild 2011 to determine the distributivity of nominal expressions in Shan with the plural morpheme. As was discussed in section 3.4.1, distributive predicates like *jàj* 'big' distribute to the of atomic units of the nouns they modify. The presence of a classifier that specifies a group or container will also specify the level that the predicate distributes down to. If *tsɤ́* is just like other group classifiers, it would specify that the distributive predicate

be predicated of the group rather than the individuals forming the group, unless the adjective was combined with an individual classifier.

The first step is to identify a set of predicates that are possibly obligatorily distributive. These include the Shan words for things like round, big, small, and long. (360)–(362) demonstrate some predicates that tend to be 'stubbornly distributive' as identified by Schwarzschild 2011.

(360) màaknăŋ nâj mán món.
ball this 3 round
'This ball is round.'

(361) màaknăŋ nâj mán jàɯ.
ball this 3 big
'This ball is big.'

(362) màaknăŋ nâj mán laaw.
ball this 3 egg-shaped
'This ball is egg-shaped.'

(363)–(364) demonstrates that mass nouns like *nâm* 'water' and *khaw* 'rice' do not combine as easily with predicates such as *jà*ɯ 'big'.

(363) #nâm nâj mán jàɯ.
water this 3 big
intended: 'This water is big.'

(364) #khaw nân mán jàɯ nàa.
rice that 3 big
intended: 'That rice is big.'

The N Clf Dem structure has the meaning 'this or that N' for individual classifiers or 'this or that GROUP of N' for group classifiers. When distributive predicate combines with a subject with the form N Clf Dem, where Clf is an individual classifier, the predicate distribute to the individual, shown in (314)–(315), repeated below.

(314) măa tŏ nân jàj nàa
dog CLF.ANML that big very
'That dog is very big.'

(315) màak.khŏ hòj nân jàj nàa
jujube CLF.RND that big very
'That jujube is very big'

When the N Clf Dem subject includes a group classifier, the distributive predicate distributes to the group or the individual, as in (316)–(317), repeated below.

(316) mǎa phŭŋ nân jàj nàa
dog GROUP.ANML that big very
'That group of dogs is very big.'

'That group of dogs, the dogs are very big.'

(317) màak.khŏ kŏŋ nân jàj nàa
jujube pile that big very
'That pile of jujube is very big.'

'That pile of jujube, the jujube are very big.'

Two structures the plural morpheme can appear in are (i) N Pl Dem, and (ii) Pl N Dem. For structure (i), (365)–(367) show that the addition of the plural morpheme *tsɤ́* to (360)–(362) does not hinder the distributive reading.

(365) màaknăŋ tsɤ́ nâj mán món.
ball PL this 3 round
'These balls are round.'

(366) màaknăŋ tsɤ́ nâj mán jàɯ.
ball PL this 3 big
'These balls are big.'

(367) màaknăŋ tsɤ́ nâj mán laaw.
ball PL this 3 egg-shaped
'These balls are egg-shaped.'

(365)–(367) include a left-located demonstrative-phrase and a co-referring pronoun *mán* 'he/she/it/they'. (275) demonstrates that even without the pronoun, the distributive interpretation to individual dogs is available. As discussed in section 3.5, the optional *tŏ* classifier in (275) would require that the dogs themselves be big. However, in the absence of that individual classifier, we would expect that the interpretation where the group is big would be available if *tsɤ́* were a group classifier. That interpretation is not available.

(275) mǎa tsɤ́ nân (tŏ) jàj nàa
dog PL that (CLF.ANML) big very
'Those dogs, they are very big.'

NOT: 'The group of dogs is big.'

For structure (ii) of the plural morpheme (Pl N Dem), the distributive reading is also available.

In (368), when *tsɤ́ màaknăŋ nâj* ‘these balls’ is repeated, it refers to different groups of balls, each with a different property

(368) tsɤ́ màaknăŋ nâj mán jàɯ. tsɤ́ màaknăŋ nâj mán haat. tsɤ́ màaknăŋ nâj mán lêk.
PL ball this 3 big PL ball this 3 medium PL ball this 3 small
‘These balls are big. These balls are medium. These balls are small.’

(referring to three separate groups of balls)

The predicate *tăaj* ‘die’ also demonstrates some interesting patterns with respect to distributivity. As (369)–(373) show, the predicate *tăaj* is compatible with both the (i) N Pl Dem and (ii) Pl N (Dem) structures.

(369) kón tsɤ́ nâj khăw tăaj jâw.
people PL this 3.PL die PFV
‘Those people died already.’

(370) tsɤ́ kón nâj tăaj jâw.
PL people this die PFV
‘Those people died already.’

(371) tsɤ́ khăw tăaj jâw.
PL 3.PL die PFV
‘They died already.’

(372) tsɤ́ măa tsɤ́ nâj tăaj jâw.
PL dog PL this die PFV
‘Those dogs died already.’

(373) măa tsɤ́ nâj tăaj jâw.
dog PL this die PFV
‘Those dogs died already.’

(374) shows that reference to the size of a group when the subject includes the plural morpheme *tsɤ́* requires bridging using *phŭŋ khăw* ‘their group’.

(374) măa tsɤ́ nân phŭŋ khăw lêk nàa
dog PL that group 3 small very
‘Those dogs, their group is very small.’

The data in (375)–(377) demonstrate how quantifiers that reference portions interact with collective and plural subjects. (375) and (376) include the morpheme *tsɤ́* in the head noun position in (375) and classifier demonstrative position in (376). With both of these, the portion word *kamphɔŋ* in a floated position can be used to indicate that some but not all of the plurality have died. In contrast, the collective word *tsúm* ‘group’ in the head noun position does not allow for only part of the group to have died when combined with *tăaj* ‘die’.

(375) tsɤ́-khăw tăaj kwàa kamphɔŋ jâw
PL-3 die go some PFV
'Some of them died already.'

(376) măa tsɤ́ nân lûtăaj kwàa kamphɔŋ jâw
dog PL that die go some PFV
'Those dogs died, some of them.'

(377) #tsúm nân tăaj kwàa kamphɔŋ jâw
group that die go some PFV
'Intended: The group already died, some of them.'

(376) demonstrates that when combined with the predicate *tăaj* 'die', the plural *măa tsɤ́ nân* 'those dogs' can have a partitive interpretation in combination with the word *kamphɔŋ* 'some', but the collective word *tsúm* 'group' cannot.

Plural Morpheme Analysis

It is clear from the preceding section that the plural morpheme, while perhaps syntactically similar to classifiers and collectives in Shan, is semantically distinct from both.

The way that it interacts with distributive predicates suggests that it does not create a collective reading but instead creates properly plural interpretations. Therefore, I propose that the plural classifier is unlike group classifiers in that it does not connect the individuals that make up the noun with a collective noun semantics. Instead it is more like the individual classifier. The semantics for the noun form and classifier form of *tsɤ́* are in (379) and (380).

(378) $[\![\text{CLF}]\!] = \lambda P \lambda x.P(x) \wedge \mu_{\text{CLF}}(x) = 1$

(379) $[\![\text{ts}\acute{\text{ɤ}}_{\text{N}}]\!] = \lambda x.\mu(x) \neq 1$

(380) $[\![\text{ts}\acute{\text{ɤ}}_{\text{CLF}}]\!] = \lambda P \lambda x.P(x) \wedge \mu(x) \neq 1$

(381) $[\![\text{GROUP}_{\text{N}}]\!] = \lambda x.{}^{*}\text{GROUP}(x)$

The interpretation of *măa tsɣ̌ nân* 'those dogs' with the plural morpheme can now be seen in (382). Table 3.10 shows the possible classifier combinations with *măa*.

(382)

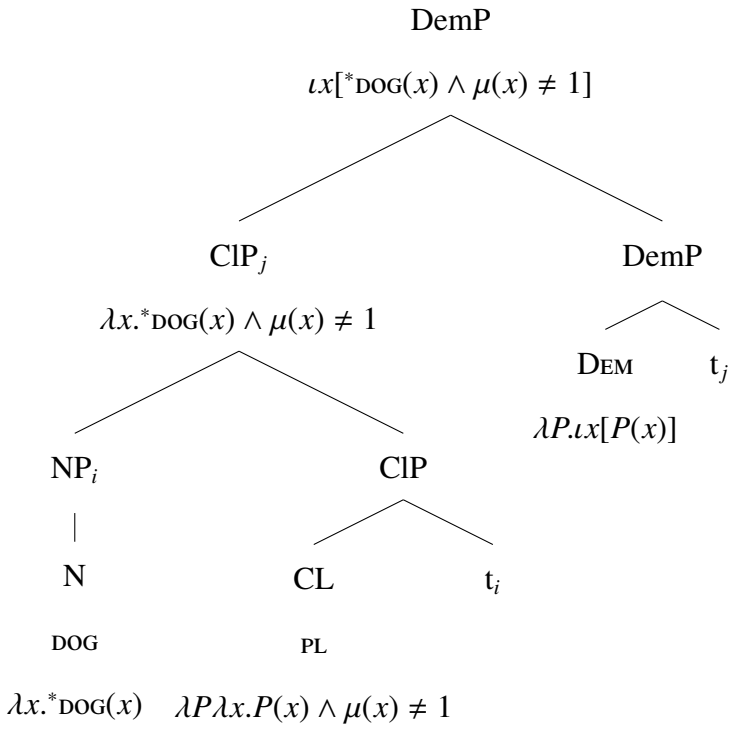


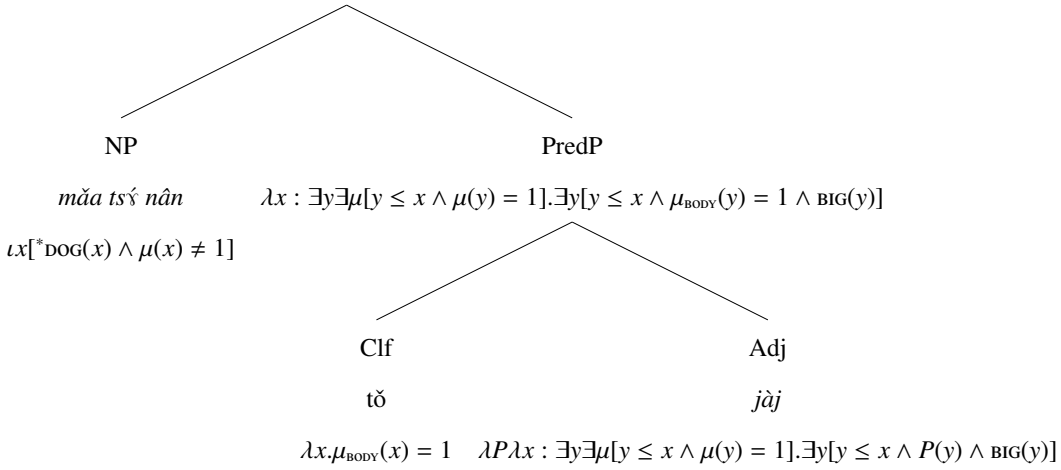
Table 3.10: N Clf Dem interpretation

Shan	Translation	CLF	Denotation
măa tŏ nân	'That dog'	individual	$\iota x[^{*}\text{DOG}(x) \wedge \mu_{\text{BODY}}(x) = 1]$
măa nân	'That/Those dog(s)'	∅	$\iota x[^{*}\text{DOG}(x)]$
măa phŭŋ nân	'That group of dogs'	group	$\iota x[\exists y[^{*}\text{DOG}(x) \wedge \mu_{\text{GROUP.A}}(y) = 1 \wedge h(x) = h(y)]]$
măa tsɣ̌ nân	'Those dogs'	plural	$\iota x[^{*}\text{DOG}(x) \wedge \mu(x) \neq 1]$

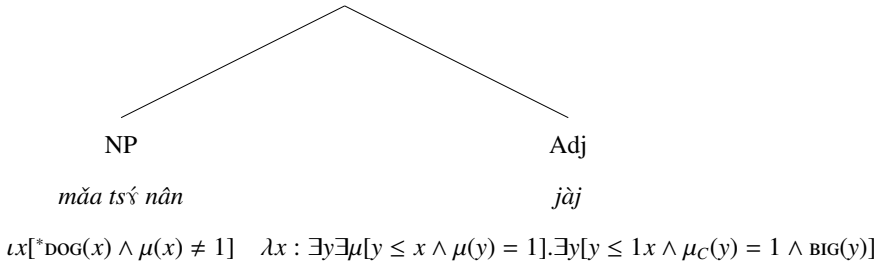
(383) and (384) show the derivations of (275), repeated below, where (383) shows the version that has an overt classifier in front of the adjective and (384) shows the example without the overt classifier. They end up with the same meaning because the plural morpheme does not add a μ distinct from the one for the noun.

(275) măa tsɣ̌ nân (tŏ) jàj nàa
dog PL that CLF.ANML big very
'Those dogs, they are very big.'

(383) $\exists y[y \leq \iota x[^{*}\text{DOG}(x) \wedge \mu(x) \neq 1] \wedge \mu_{\text{BODY}}(y) = 1 \wedge \text{BIG}(y)]$



(384) $\exists y[y \leq \iota x[^{*}\text{DOG}(x) \wedge \mu(x) \neq 1] \wedge \mu_C(y) = 1 \wedge \text{BIG}(y)]$

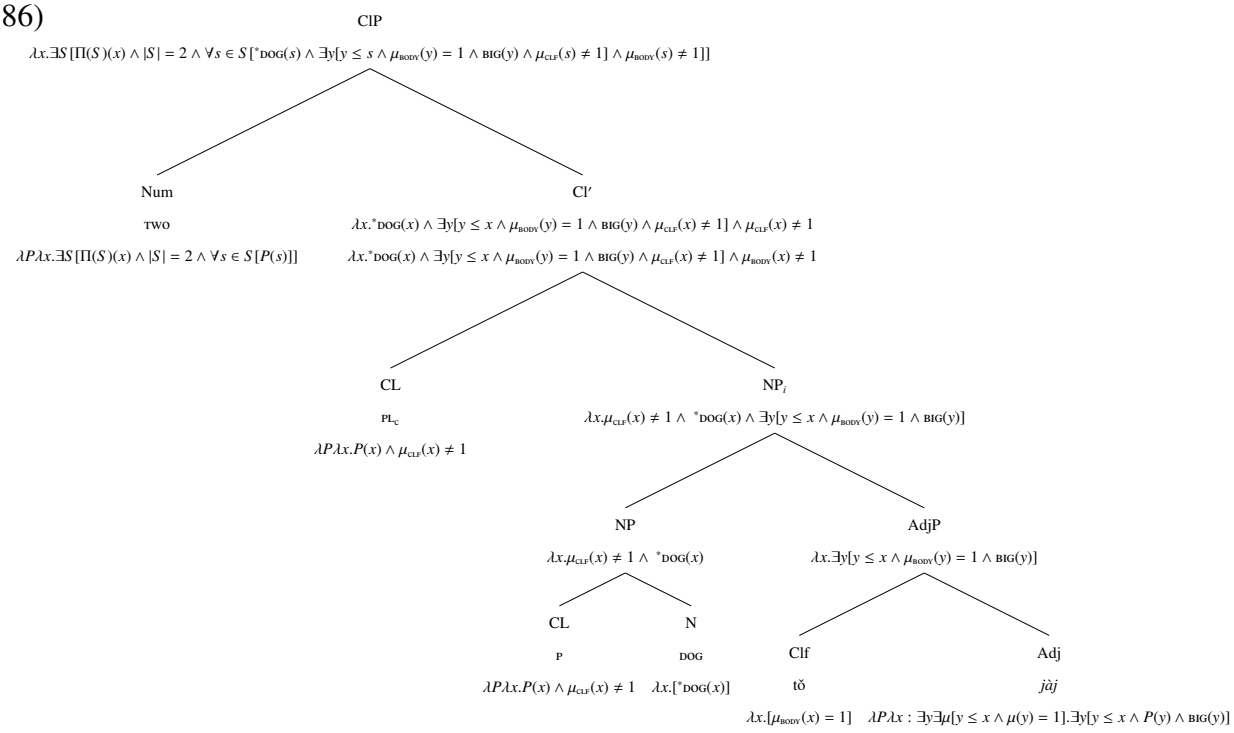


The derivation of (385) in (386) shows the N use of *tsɤ́* and the use of *tsɤ́* in numeral-classifier constructions.[21] This actually only uses the proposed classifier semantics for *tsɤ́*. The meaning is two pluralities of dogs that consist of individual big dogs. This is close to the desired interpretation.

(385) háw khaj sɯ̂ tsɤ́ măa tǒ jàj sɔ̌ŋ tsɤ́
1 want buy PL dog CLF.ANML big two PL
'I want to buy two types of big dogs.'

[21] The NP_i would be moved to adjoin to a position above the highest ClP, but I have left it in its original position for an easier derivation.

(386)



3.4.5 Conclusion

I have proposed that the basic semantic contribution of the classifier is to specify a unit. This function is apparent from the fact that the addition of a classifier creates a singular interpretation and from the interaction between stubbornly distributive predicates and classifiers. I have also proposed that group and plural morphemes are distinct. The group morpheme creates a collective interpretation, which includes semantics to indicate that the collective and the individuals that form the collective are made of the same 'stuff'. In contrast, the plural morpheme is used to create a plurality, not a collective. The plural morpheme *tsý* functions as a classifier or forms a compound with a noun. It can also co-occur with individual classifiers, showing that it is different from the regular classifier. Certain group/container terms can do that as well. I have proposed that the basic meaning of the plural morpheme is non-atomicity and that when it appears with an individual classifier, it appears in a separate plural functional layer that otherwise functions the same as the classifier phrase.

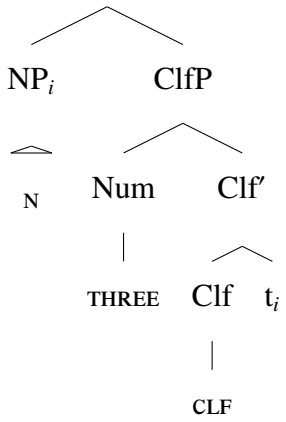
3.5 Count and measure: A unified analysis

3.5.1 Unified analysis

With a basic syntax and semantics of classifiers established, this section investigates the COUNT/MEASURE distinction in Shan and proposes a unified analysis of counting and measuring.

The proposed syntax for [N NUM CLF] structures is given in (387).

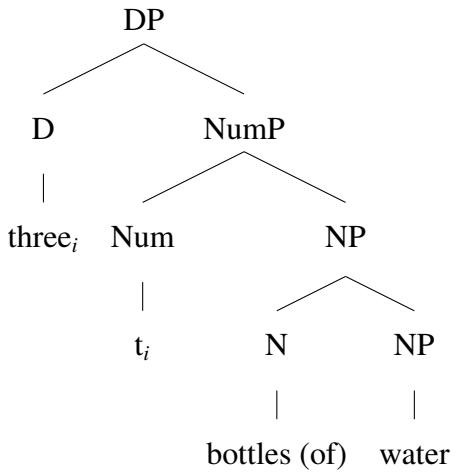
(387)



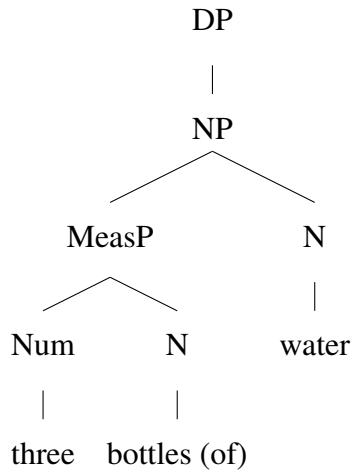
An important issue in the literature related to numerals and classifiers is the syntactic and semantic distinction between COUNT and MEASURE constructions. Li & Rothstein (2012) argued for a distinct syntax corresponding to a distinct semantics for [NUM CLF N] structures where [NUM [CLF N]], as in (388a), has a COUNT semantics and [[NUM CLF] N], as in (388b), has a MEASURE semantics. They especially rely on data from Mandarin Chinese and Hebrew to make this claim. Bale et al. (2019) argue that Ch'ol, a Mayan language, only has the syntactic structure in (388b) for both count and measure interpretations. Landman (2016a) argues for a unified syntax, as in (388a), for English and Dutch, but with a compositional semantics that corresponds to the COUNT and MEASURE syntax proposed by Li & Rothstein (2012).

(388) Li & Rothstein 2012: (47a,b)

a. Counting reading



b. Measure reading



I would argue that this problem comes up as a result of the possible bridging-like semantic shifts between containers and contents interpretations that is only possible when container and substance nouns are involved.

Some, such as Cheng & Sybesma (1999), have argued for separating classifiers from 'massifiers' (measure words). As I have already shown in the previous section Shan measure words are syntactically similar to other classifiers, but are semantically distinct in certain ways. Rothstein (2016) proposed that measure phrases can have different syntactic and semantic forms when they have a count or measure function, meaning that the same classifier can have a distinct syntax and semantics depending on whether it is counting or measuring.

Li & Rothstein (2012) largely discuss this possible count/measure distinction with respect to container terms in Mandarin, so I will also focus on container terms here. For many nouns that denote containers, such as *kɔ́k* 'cup' in (389)–(390), the noun and the classifier term have the same form.[22] For example 'three cups' looks like CUP$_N$ THREE CUP$_{\text{CLF}}$. When there is an action pertaining to the container noun *kɔ́k* 'cup' in (389), the container head noun is obligatory. When the sentence is about the contents of the container, the container head noun is disallowed as in (390).

[22]See more discussion of these repeater classifiers in §3.3.1.

(389)–(390) also show that, aside from the form of the noun, count and measure constructions in Shan have the same structure. The numeral-classifier/measure construction is *sǎam kɔ́k* for both.

(389) háw ʔǎw #(kɔ́k) nâm sǎam kɔ́k tɛ̀k
1 take #(cup) water three cup break
'I broke three water cups/#cups of water.' (COUNT)

(390) thɛŋ pǎn (#kɔ́k) nâm sǎam kɔ́k sàɯ ti náɯ mɔ
add give (#cup) water three cup put at in pot
'Add three cups of water/#water cups to the pot.' (MEASURE)

When the distinction between contents and container is not relevant (see Partee & Borschev 2012), the container head noun is optional, as shown in (391). The only syntactic differences between the COUNT (389) and MEASURE (390) constructions in Shan is that in counting contexts, the container word appears as the head noun of the N-N compound.

(391) háw ʔǎw (kɔ́k) nâm sǎam kɔ́k máa pǎn khɛ̀k
1 take cup water three cup come give guest
'I brought three cups of water for the guests.' (CONTAINER+CONTENTS)

Note that the container+contents reading is available when the substance noun combines with a numeral-classifier, as in (392). The verb *phùj* 'open' technically must apply to the container, rather than the contents. However, the noun here *nâmneŋ* 'tea' clearly denotes the substance. It must be possible to bridge the interpretation between the container and contents as the literal recipient of the opening event, perhaps due to the fact that the bottle is being opened to access the contents.

(392) kǎw phùj nâmneŋ sɔ̌ŋ taw jâw
1 open tea two bottle PFV
'I opened two bottle of tea.' (CONTAINER+CONTENTS)

Some examples from English can help elucidate the bridging involved in these constructions. If we think about the object of the verbs *opened* in (393) versus *drank* in (394), the interpretation of *two bottles of tea* is clearly somewhat different. What is being opened is the bottle and what is

being drunk is the tea, and in (393) *two* refers to the number of bottles and in (394) *two* refers to the volume of tea in bottles.

(393) I opened two bottles of tea. $\rightsquigarrow \exists x[\textsc{opened}(I, x) \wedge \textsc{bottle}(x) \wedge \mu_{\textsc{bottle}}(x) = 2 \wedge \exists y[\textsc{contain}(x, y) \wedge \textsc{tea}(y)]]$

(394) I drank two bottles of tea. $\rightsquigarrow \exists x[\textsc{drank}(I, x) \wedge \textsc{tea}(x) \wedge \exists y[\textsc{bottle}(y) \wedge \mu_{\textsc{bottle}}(y) = 2 \wedge \textsc{contain}(y, x)]]$

The standard measure term *liter* can also be part of the object of the verbs *opened* and *drank* equally well. For standard measure terms like *liter*, the sense is that they describe an amount—specifically a volume—rather than a container. There must again be some ability to shift between the container and contents to some degree.

(395) I opened two liters of tea. $\rightsquigarrow \exists x[\textsc{opened}(I, x) \wedge \textsc{container}(x) \wedge \exists y[\textsc{contain}(x, y) \wedge \textsc{tea}(y)] \wedge \mu_{\textsc{liter}}(x) = 2]$

(396) I drank two liters of tea. $\rightsquigarrow \exists x[\textsc{drank}(I, x) \wedge \textsc{tea}(x) \wedge \mu_{\textsc{liter}}(x) = 2]$

Things are different when it comes to verbs like *break*, which can only apply to containers. As (397) shows, the measure expression *two liters of tea* is not a good subject for the verb *break*. In contrast, the expression *two bottles of tea* in (398) is entirely acceptable. This is similar to example (389).

(397) #Two liters of tea broke when they/it hit the ground.

(398) Two bottles of tea broke when they hit the ground.

The contrast is milder with (399) and (400). When the verb applies to an amount and not to a container, the standard measure term is fine, as in (399), but the container term is odd, as in (400).

This is somewhat similar to the contrast in (390).[23]

(399) Add two liters of tea to the soup.

(400) #Add two bottles of tea to the soup.

(400) can be made better by adding *worth* to *bottles* and turning it into a measure. Note that the verbs that pertain to the container are now not very good (402)–(403).

(401) Add two bottles' worth of tea to the soup.

(402) #I opened two bottles' worth of tea.

(403) #Two bottles' worth of tea broke when they/it hit the ground.

3.5.2 Count versus measure

In Shan, there does not seem to be a distinction between whether the [CONTENTS NUM CLF] construction has the interpretation where the contents are measured in the same container or not. (404) can either be followed by (405) or (406). From this, we can see that the measure of wine described by the container-type classifier can be used to refer to the total measure of wine consumed whether or not the wine was in separate (405) or the same (406) containers. This seems to not have a CONCRETE PORTION distinction as described by Partee & Borschev (2012).

(404) ti pǎaŋkǐnlêŋ nân háw kǐn wáaj sǎam kɔ́k
at party that 1 drink wine three cup
'At the party I drank three glasses of wine.'

(405) háw ʔǎw wáaj sǎam kɔ́k ti nɤ̌ phɤ̌n ʔǎn pɤn tám wâj nân
1 take wine three cup at on table COMP others place stay that
'I took three glasses of wine from the table that someone had put (there).'

[23] I put a small # with (400) because it really is not that bad. There is a clear, salient interpretation where you are adding the tea and the bottle together to the soup, which leads to the oddness.

(406) jə̂npɣwaa kón hét kăan ti hâan nân máa thɛŋ păn wáaj háw tíktík
because person do work at restaurant that come add give wine 1 again_and_again
'Because the person working at the restaurant kept giving me more wine.'

(407) means that the speaker hit and broke three bottles containing wine and (408) means that the speaker hit and broke three empty bottles. It is not entirely clear why (407) can have an interpretation where the wine bottle broke. Perhaps this is a CONTAINER+CONTENTS reading. That (408) can only refer to the bottle is not surprising.

(407) háw phàa sàj wáaj săam taw tók tɛ̀k
1 collide into wine three bottle fall break
'I hit three bottles of wine, (which) fell and broke'

(408) háw phàa sàj taw wáaj săam taw tók tɛ̀k
1 collide into bottle wine three bottle fall break
'I hit three wine bottles, (which) fell and broke.'

Given that I showed in section 3.3.3 that adverbs cannot give us insight into the structure of the measure versus count construction, I will use selectional evidence to demonstrate that there is no apparent distinction between counting and measuring in Shan. Previously, I had argued that the fact that numeral-classifier expressions are selected by the verb *mí* rather than used predicatively suggests that they are nominal rather than predicative. An example of this is (269), repeated below in (409). (410) and (411) both include what must be expressions of measure. Still, they are selected by the predicate *mí* just as the count expression in (409) is.

(409) măa tsɣ́ nâj mí săam tŏ
dog PL this have three CLF.ANML
'These dogs are three.'

(410) mɣnâj háw kǐn nâm táaŋnăm mán mí jù pɛ̀t kɔ́k.
today 1 drink water amount 3 have stay eight cup
'Today, I drank water, in the amount of eight cups of water.'

(411) kón kô nɯŋ ʔăw màaktsɔ́k máa ti kéŋmàj táaŋnăm mán mí săam kìlò
person CLF.HUM one take orange come at Chiang_Mai amount 3 have three kilo
'A person brought oranges to Chiang Mai, the amount of them was three kilograms.'

I am assuming a numeral semantics as shown in (412), where S is a partition, written Π, of some entity x if it is a cover of x and its cells do not overlap. Since Shan mass and count nouns are not quantized, they require a classifier to quantize them in order for them to be able to combine with a numeral.

An alternative semantics for the numeral is given in (413).[24] The choice of numeral is not significant here beyond being able to generate a set of sums from a set of atoms.

(412) $[\![\text{TWO}]\!] = \lambda P \lambda x . \exists S [\Pi(S)(x) \wedge |S| = 2 \wedge \forall s \in S [P(s)]]$ (Ionin & Matushansky 2006)

(413) $[\![\text{TWO}]\!] = \lambda P \lambda x . x = \cup Y \wedge Y \subseteq P \wedge |Y| = 2$ (Bale et al. 2019)

The proposed classifier/measure semantics can be seen in (414). This is similar to Nomoto's (2013) analysis of classifiers in that it individuates; however, the unit of individuation is based on the classifier/measure term rather than the noun for all categories of classifiers that are not individual classifiers. This captures the data in section 3.4.1 that shows that the 'unit' of measure describing a noun can vary based on the classifier, group, or measure term. This allows for a unified analysis of the classifier/measure expression. (302) gave the semantics for the animal classifier *tŏ*. (414) gives the generic classifier semantics. (415) shows the noun denotation of 'cup', and (416) shows the corresponding classifier/measure term denotation.

(414) $[\![\text{CLF}]\!] = \lambda P \lambda x . P(x) \wedge \mu_{\text{CLF}}(x) = 1$

(415) $[\![\text{kɔ́k}_N]\!] = \lambda x .{}^{*}\text{CUP}(x)$

(416) $[\![\text{kɔ́k}_{\text{CLF}}]\!] = \lambda P \lambda x . P(x) \wedge \mu_{\text{CUP}}(x) = 1$

I have already discussed in section 3.4.1 that there is a need to define the relationship between a collective and the individuals that make up the collective using a semi-lattice homomorphism. It is necessary to define a distinct but related relationship between containers and their contents. On

[24] This semantics requires that the ∪ be able to generate a sum from a set of atoms that make up the sum.

the one hand, there seems the be a similar semantic bridging between the container and contents. On the other hand, the container and contents are not, in fact, made of the same *stuff*/*material*. Perhaps there is the same relationship between container nouns and container classifiers, where a relationship between the container and contents can be specified.

Semantics of different categories of Shan nouns/classifiers are as shown in (417)–(423). I am assuming that container nouns are relational nouns just like *mother*, etc. CONTAIN indicates that there is a containment relationship between the container and its contents. Presumably, there can be two possible denotations for container nouns: one where the noun denotes the container and one where it denotes a container and its contents. The classifier might have a similar ambiguity, shown in the contrast between (421) and (422). (421) looks like a regular classifier. (422) includes the relation between the container and contents. The distinction between the relational noun in (420) and the relational classifier (422) is that (i) the relational noun denotes the container rather than the contents and (ii) the classifier quantizes the noun denotation, while the noun is number neutral. (423) is the MEASURE term semantics which is identical to the version of the container classifier in (421).

(417) COUNT NOUNS: $\lambda x.^{*}\textsc{count}(x)$ (inherent atomic parts)

(418) MASS NOUNS: $\lambda x.^{*}\textsc{mass}(x)$ (no inherent atomic parts)

(419) CONTAINER N_1: $\lambda P \lambda x.^{*}\textsc{container}(x) \wedge \exists y[P(y) \wedge \textsc{contain}(x, y)]$

(420) CONTAINER N_2: $\lambda x.^{*}\textsc{container}(x)$

(421) CONTAINER CLF_1: $\lambda P \lambda x.P(x) \wedge \mu_{\textsc{container}}(x) = 1$

(422) CONTAINER CLF_2: $\lambda P \lambda x.P(x) \wedge \exists y[\textsc{contain}(y, x) \wedge \mu_{\textsc{container}}(y) = 1]$

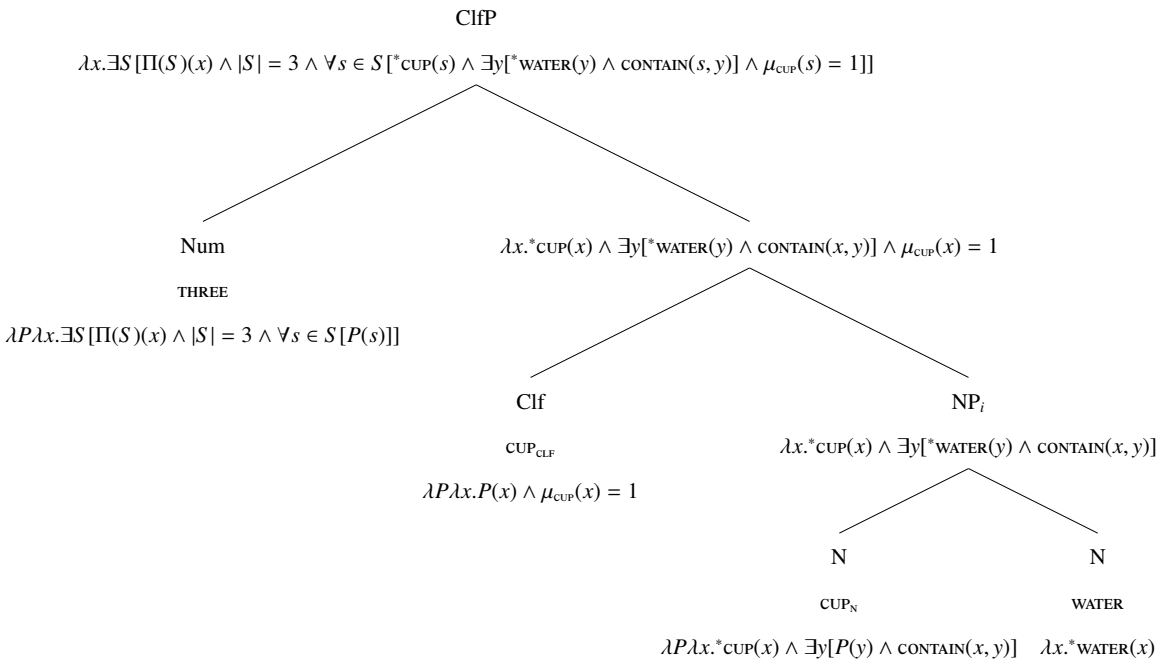
(423) MEASURE TERMS: $\lambda P \lambda x.P(x) \wedge \mu_{\textsc{measure}}(x) = 1$

Container classifiers like *kɔ́k* 'cup' are repeater classifiers, as discussed in section 3.3.1. One analysis of these constructions involved a lexical item in the classifier position that has a combined

classifier and noun semantics. This would not be very different from the semantics in (422).

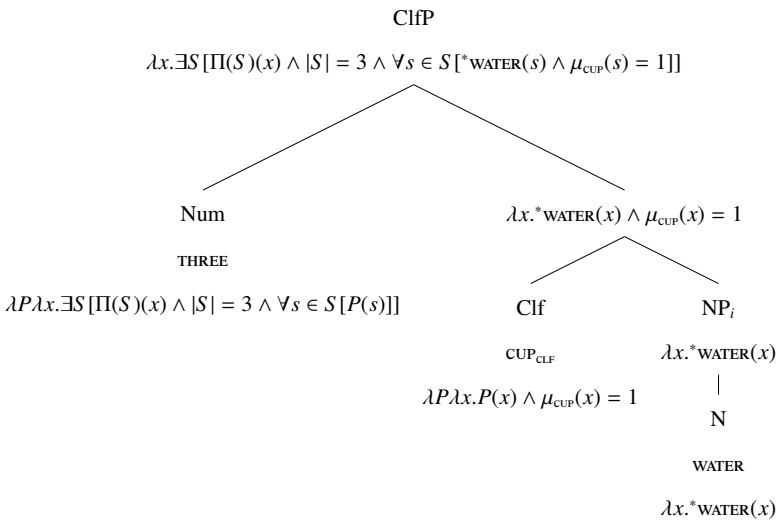
The interpretation of *kɔ́k nâm sǎam kɔ́k* in (389) can be seen in (424). (424) denotes sets of pluralities of three cups that contain water.

(424)

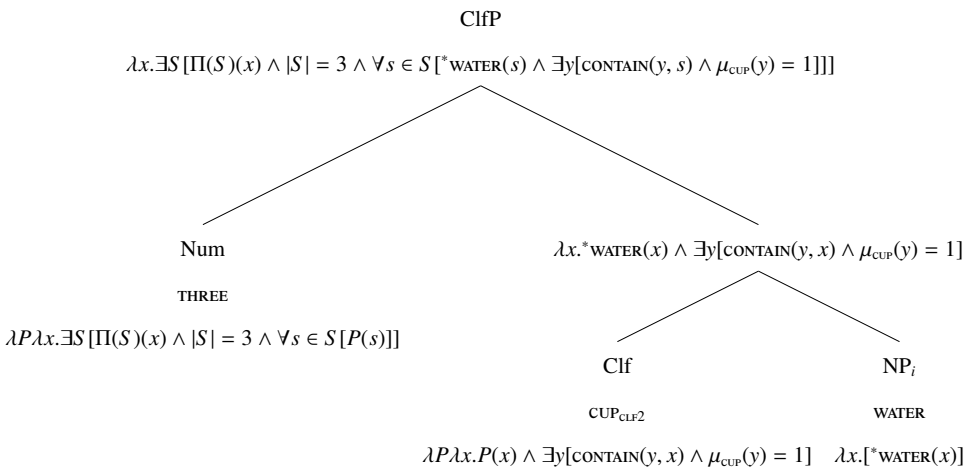


The interpretation of *nâm sǎam kɔ́k* in (390) could be either (425) or (426). (425) and (426) denote sets of pluralities of water that measure three cups in total. (425) uses the measure container classifier in (421) and (426) uses the relational container classifier in (422).

(425)



(426)



In the discussion of the collective terms, there was a homomorphism between the semi-lattice for a collective and the individuals that make up the collective. The relationship between the container and contents is different. Rather than being made up of an identical substance, the container-contents relationship in measure can be paraphrased as 'if the contents were poured into containers, the contents would fill *n* containers'. The container classifier *kɔ́k* 'cup', which is a non-standard measure, has to be able to quantize both cups themselves and any substance, such as water, that could fill the cup. The semantics of the classifier container in (422) requires that there be a physical container for the substance; however, we have seen that that is not always the case. This is why the MEASURE container form in (421) is an option.

3.5.3 Mass nouns and container classifiers

This discussion pertains to the question of whether to include the noun semantics as part of the classifiers that are not measure terms. For example, why do container classifiers allow mass nouns to combine with distributive predicates (322)–(323)? (322) cannot combine with the distributive predicate because *nâm* 'water' lacks clear atomic parts; therefore, the presupposition attached to the predicate would fail.

(322) #nâm nân jàj nàa
water that big very
intended: 'That water is big.'

(323) nâm kɔ́k nân jàj nàa
water cup that big very
'That cup of water is big.'

This can be accounted for if the container itself is what is being described as big when a container classifier combines with a mass noun. This motivates needing a version of the container classifier that includes more nominal semantics, such as (422). The semantic derivation of (323) proceeds as follows:

1. $[\![kɔ́k]\!]([\![nâm]\!]) = \lambda P \lambda x.P(x) \wedge \exists y[\textsc{contain}(y, x) \wedge \mu_{\textsc{container}}(y) = 1](\lambda x.{}^{*}\textsc{water}(x))$
2. $[\![kɔ́k]\!]([\![nâm]\!]) = \lambda x.{}^{*}\textsc{water}(x) \wedge \exists y[\textsc{contain}(y, x) \wedge \mu_{\textsc{container}}(y) = 1]$
3. $[\![nân]\!]([\![nâm\ kɔ́k]\!]) = \iota x[{}^{*}\textsc{water}(x) \wedge \exists y[\textsc{contain}(y, x) \wedge \mu_{\textsc{container}}(y) = 1]]$
4. $[\![jàj_1]\!] = \lambda x : \exists y \exists \mu[y < x \wedge \mu_C(y) = 1].\exists y[y < x \wedge \mu_C(y) = 1 \wedge \textsc{big}(y)]$
5. $[\![jàj_1]\!]([\![nâm\ kɔ́k\ nân]\!]) =$
$$\exists y[y \leq \iota x[{}^{*}\textsc{water}(x) \wedge \exists y[\textsc{contain}(y, x) \wedge \mu_{\textsc{container}}(y) = 1]] \wedge \mu_{C=\textsc{cup}}(y) = 1 \wedge \textsc{big}(y)]$$

When a measure term is the classifier, it is the measure term that is being described as big. This is what leads to the infelicity of (326) above. The measure itself cannot be described as big because it cannot refer to a physical object.

Previous accounts of count and measure expressions give a radically different semantics for the measure and count use of classifiers, such as Li & Rothstein 2012 for Mandarin. While a distinct syntax of count/measure could motivate a different semantic composition, a unification of the underlying measure function of classifiers could make those expressions more comparable within and across languages.

3.5.4 Nez Perce

This unified analysis of counting and measuring fits well with Deal's (2017) analysis of nouns in Nez Perce and Yudja (Lima 2014). In these languages, there is *prima facie* evidence for a lack of a mass-count distinction, since numerals can combine directly with both typical count and substance nouns. However, Deal (2017) gives evidence to support the idea that there is in fact a root distinction between count and substance nouns, and a syntactically null element intervenes to turn a mass denotation into a quantized denotation in some contexts.

The assumed count/measure semantics contrast was based on the fact that there are syntactic differences between count and measure constructions separate from count/mass distinctions. In a language where the only syntactic distinction is due to the count/mass distinction rather than the count/measure distinction, there is no reason to assume there is a difference between counting and measuring. The analysis proposed in the previous section can easily be extended to Nez Perce and Yudja. The primary difference would be in that the covert quantizing/classifier morpheme that Deal (2017) proposes would be get its semantic 'unit' from the context rather than from a classifier.

As (427) and (428) show, substance nouns in Nez Perce (NP) combine directly with distributive predicates or numerals. Despite that, Deal (2017) argues that substance nouns and count nouns are distinct in Nez Perce. When they combine with a quantifier, count nouns, but not substance nouns must first combine with a plural morpheme, as shown in the contrast between (429) and (430).[25]

(427) himeeq'is kuus
big water
'(the) big portion of water'

(428) mitaat lalx / kuus
three coffee / water
'three coffees/ waters'
(NP, Deal 2017: (58a),(32))

(429) 'oykal-oo ha-ayat/*aayat
all$_1$-HUM PL-woman/*woman.SG
'all the women'

(430) 'oykala ta'c hipt
all$_1$ good food
'all good food'
(NP, Deal 2017: (74),(88))

[25]For non-human nouns in Nez Perce, plurality is not marked on the noun but is visible on the adjective.

Deal (2017) proposes an analysis where there is a covert atomizing function α that quantizes the substance noun. For example, in (431), α atomizes the substance noun *kike't* 'blood' before it combines with a numeral, leading to the interpretation in (432). For (429)–(430) above, the quantifier meaning 'all' must combine with a cumulative predicate, so singular count nouns as in (429) are made cumulative by combining with the plural morpheme. Substance nouns like (430) do not combine with a plural morpheme before combining with such a quantifier. Deal (2017) proposes that the plural morpheme is essentially Link's (1983) * operator (closure under sum).

(431) Lepit kike't hi-sseew-n-e.
two blood 3SUBJ-drip-PFV-REM.PAST
'Two drops of blood fell.' (NP, Deal 2017: (65a))

lepit 'two' [Num:– [α_n $\sqrt{}$*kike't* 'blood']]

(NP, Deal 2017: (65b))

(432) $|\{x : AT_n(\text{blood})(x) \wedge \text{fell}(x)\}| \geq 2$ (Deal 2017: (66))

In a study where speakers of Nez Perce were asked to determine *who had more* of something, they were shown pictures of substances or objects, where on one side the substance/object had more volume but was less numerous and on the other side the substance/object had less volume but was more numerous. The objects were always more numerous, but for the substance nouns *more* meant more volume when there was no plural marking and meant more numerous when there was plural marking. Deal (2017) took this to indicate that when there was plural marking the substance had been covertly quantized and then combined with plural to be able to combine with a quantifier.

There is evidence in Nez Perce for a mass/count distinction without a count/measure distinction. This is compatible with the unified count/measure semantics proposed here. In Nez Perce, there is a null α that is the equivalent of a MEASURE classifier that quantizes a substance denoting noun.

3.5.5 Mandarin classifiers

Mandarin classifiers as described by Li & Rothstein (2012) are significantly different from Shan classifiers, regardless of the surface similarity of having obligatory numeral classifiers and number neutral nouns. For example, container nouns in count/measure constructions only appear once in Mandarin constructions whether the noun denotes the container and contents or just the contents (measure). This can be seen in (433). According to Li & Rothstein (2012), (433) can have two interpretations, one count and one measure. Under this analysis, *wan* 'bowl' still functions as a classifier in the measure reading, but it functions as a combined classifier and noun in the count reading. When the morpheme *de* intervenes between the classifier and the noun, the count interpretation is blocked, as shown in (434).

(433) yi wan tang
one CLF$_{bowl}$ soup
'one bowl of soup' (MC, Li & Rothstein 2012)

(434) #ta kai le san ping de jiu.
he open PFV three CLF-bottle DE wine
'He opened three bottles of wine.' (MC, Li & Rothstein 2012: (10))

In Shan, the container noun appears once in measure constructions and typically twice in count constructions, where the extra one is contributing the noun semantics of the container that controls the reference of the noun. In Mandarin it must be the case that the container noun in count constructions fulfills the double function of being a classifier and a container in a combined CONTAINER+CONTENTS reading.

(435) and (436) demonstrate another difference between Shan and Mandarin. In Shan, the classifier for counting bottles and the classifier for measuring bottles is the same, but in Mandarin they are different. Example (435) can have either a MEASURE or CONTAINER+CONTENTS reading where as (436) has only the COUNT CONTAINER reading. The verb 'break', for example, is not compatible with (435) but is compatible with (436).

(435) yi ping shui
one $\text{CLF}_{\text{bottle}}$ water
'a bottle of water'

(MC, Li 2011: (32a))

(436) yi *(ge) ping
one *(CLF) water
'a bottle'

(MC, Li 2011: (32a′))

In contrast, (437)–(438) both use the classifier *taw* 'bottle'. The expression in (438) depicts something that can be broken, but (437) typically does not.

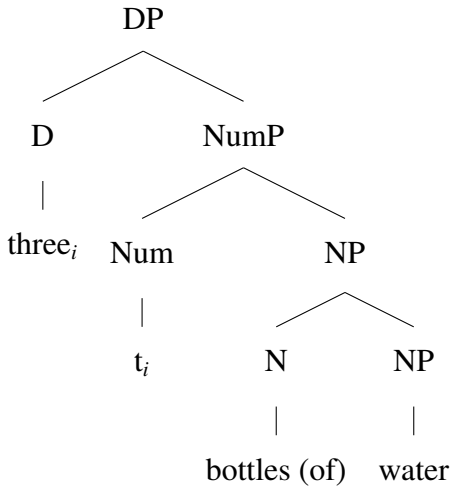
(437) nâm nɯŋ taw
water one bottle
'one bottle of water'

(438) taw nɯŋ taw
bottle one bottle
'one bottle'

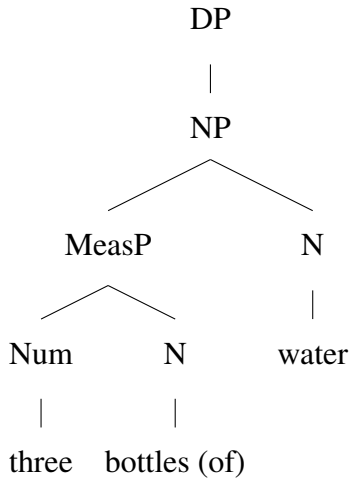
Li & Rothstein (2012) proposed that the counting and measuring readings available for (433) come about as a result of a separate syntax and semantics. Their proposed syntax for the counting reading is given in (439a) and the one for the measure reading is shown in (439b).

(439) Li & Rothstein 2012: (47a,b)

a. Counting reading



b. Measure reading



I will not reiterate all their arguments for this distinction, but it is, in part, based on the fact that certain morphemes, like *de* in (434) (repeated in (440)), can appear between the container noun and the contents noun, forcing a measure reading. In (434), the verb *kai* 'open' is only compatible

with a counting interpretation, making this example with *de* infelicitous. I will adopt their syntactic account of Mandarin classifier/measure phrases.

(440) #ta kai le san ping de jiu.
he open PFV three CLF-bottle DE wine
'He opened three bottles of wine.' (Li & Rothstein 2012: (10))

Below I will briefly go over a simplified version of the analysis of count and measure terms in Li & Rothstein 2012. I have been assuming that the basic denotation of nouns is a set of individuals and sums of individuals; whereas Li & Rothstein (2012) assumes that the basic denotation is a kind, so I have tried to represent Li & Rothstein's (2012) analysis without that aspect. The points discussed here do not depend on that choice.

(441) gives the numeral denotation, and (442) gives the noun denotation. (443) and (444) are the measure and classifier version of the classifier, respectively. (445) is the denotation of *de*. The measure version of the classifier (443) clearly requires a numeral and creates a set of entities that have the specified measure. The counting classifier version in (444) quantizes or atomizes the noun that it combines with at the specified unit.[26]

(441) $[\![liang]\!] = 2$

(442) $[\![jiu]\!] = \lambda x.\textsc{wine}(x)$

(443) $[\![ping]\!] = \lambda n \lambda x.\textsc{mea}(x) =< n, \textsc{bottle-ful} >$ (Li & Rothstein 2012: (61))

(444) $[\![ping_{clf}]\!] = \lambda P \lambda x.P(x) \wedge \textsc{bottle}_{unit}(x)$ (based on Li & Rothstein 2012: (64))[27]

(445) $[\![de]\!] = \lambda P \lambda Q \lambda x.P(x) \wedge Q(x)$ (based on Li & Rothstein 2012: (62))

[26]For Rothstein (2017) there is a difference between counting and measuring where counting associates an individual with a number and measuring simply measures the amount. I consider counting to be a kind of measuring, but I do not go into detail about the reasoning behind that here.

[27]This analysis assumes the basic denotation of a noun is a kind. I am not assuming that, but added type-shifting can incorporate that feature into an analysis if desired. I am leaving out the kind and 'k-indexed entities' for simplicity. See Li 2011 for why to use k-indexed entities.

A detour into Li & Rothstein (2012) semantics. As mentioned, the semantics presented here is not identical to the one presented by Li (2011) and Li & Rothstein (2012). The true classifier denotation they give is shown in (446).

(446) $\lambda\mathbf{k}\lambda x.\pi_1(x) \in^{\cup} \mathbf{k} \cap k \wedge \pi_2(x) = k \wedge P_{unit}(\pi_1(x))$

> A classifier denotes a function which applies to a kind denoting term and gives the set of k-indexed atomic instantiations of the kind, where all the atomic instances have the unit property expressed in P_{unit}. (Li & Rothstein 2012: (64))

k is a variable over kinds, k a variable over contexts, d is a variable over individuals, x is a variable of type $d \times k$, π_1 is a function that takes something of type $d \times k$ and returns the first projection of the ordered pair (the individual) and π_2 returns the second projection (the context). P_{unit} has the role of filtering out non-atomic entities. Then, the classifier denotation in (446) takes a kind and returns a set of $d \times k$ (k-indexed entities), where the entities are instantiations of a kind **k** in a context k which are atomic.

There is nothing wrong with using this formalization. Taking context into account is something that should be more overtly represented in the semantics, given how important it clearly it to interpretation. However, I do not think it is necessary to assume that the basic denotation of a noun in Mandarin Chinese (or Shan) is a kind as discussed in chapter 2, and I am ignoring the role of context for simplicity. While I think it is important, it is not pertinent to the argument here.

(447) gives the denotation of *ke shu* 'CLF tree' and (448) gives their proposed denotation for *wubai ke shu* '500 CLF tree'. There are a few problems with (448):

1. it is not possible to compositionally get from (447) to (448) with a numeral denotation;
2. it is not clear why the PLURAL function is necessary, since $^{\cup}\text{TREE} \cap k$ would likely already include plural entities;

3. the intended denotation of (448) is likely a set of $d \times k$ such that the entities consist of 500 atomic trees. In that case, the function P_{unit} cannot filter out non-atomic entities because that would filter out the entities that consist of 500 trees. The function $|\,|$ is clearly entity measuring, rather than set measuring. If P_{unit} simply ensured that the entities were made up of units with a particular property, then (447) would denote an atomic join semi-lattice, which seems to be not what the authors intended by their description of (447) as 'the set of k-indexed instantiations of the TREE kind which count as one tree in context k, i.e. k-atomic instantiations of TREE.' (Li & Rothstein 2012: p. 730).

(447) $[\![ke\ shu]\!] = \lambda x.\pi_1(x) \in^{\cup} \textsc{tree} \cap k \wedge \pi_2(x) = k \wedge \textsc{plant}_{unit}(\pi_1(x))$

(Li & Rothstein (2012): (65b))

(448) $[\![wubai\ ke\ shu]\!] = \lambda x.\pi_1(x) \in \textsc{plural}(^{\cup}\textsc{tree} \cap k) \wedge \pi_2(x) = k \wedge \textsc{plant}_{unit}(\pi_1(x)) \wedge |\pi_1(x)|_k = 500$

(Li & Rothstein (2012): (65c))

I see three ways of circumventing these problems:

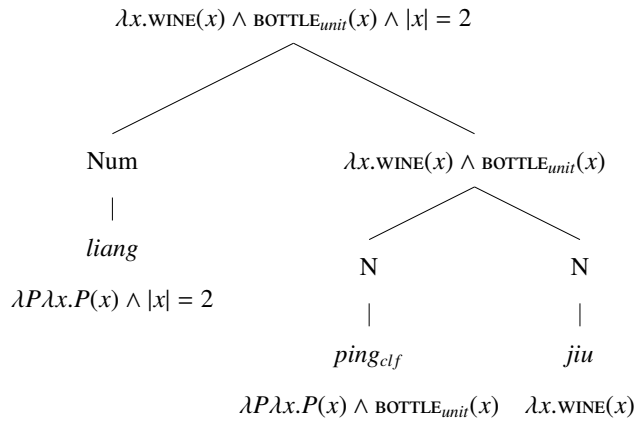
1. Build the numeral into the container-classifier denotation as it is in measure.
2. If $\textsc{bottle}_{unit}$ generates a set of atomic bottles, a partition numeral in the style of Ionin & Matushansky (2006) is needed.
3. If $\textsc{bottle}_{unit}$ simply says that the atomic parts of the predicate are bottle atoms (but that may be pluralities), a numeral that measures each entity is possible.

I am going to use the third option because I demonstrate the second option in my proposal. Note that this measure function would need to be able to count distinct atomic parts of the pluralities. This would involve a numeral semantics as proposed below:

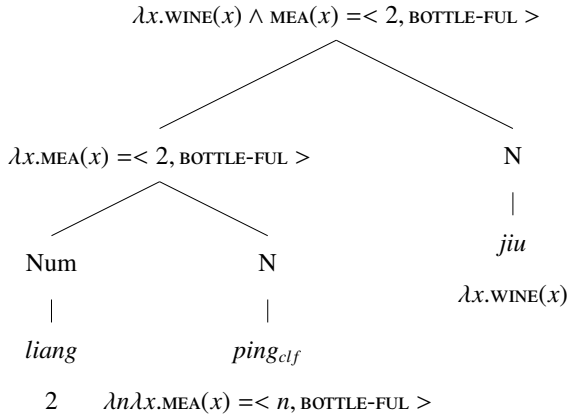
(449) $[\![liang]\!] = \lambda P \lambda x.P(x) \wedge |x| = 2$

The composition for the counting reading is given in (450). (451)–(452) give the composition for the measure reading, with (452) and without (451) the morpheme *de*. Both derivations give the same result.

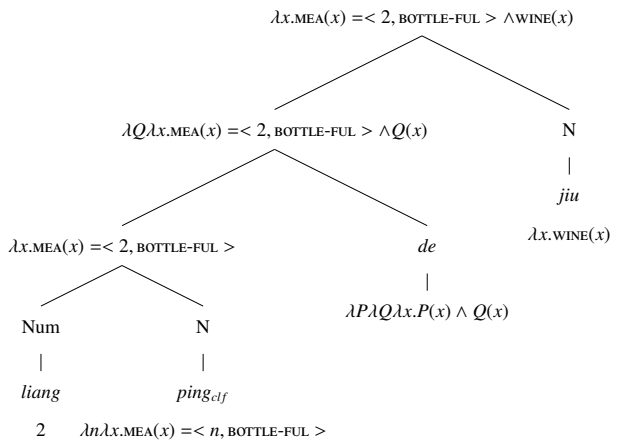
(450) Counting reading



(451) Measure reading



(452) Measure reading with *de*



The main difference between the semantics for counting and measuring as described by Li & Rothstein (2012) is in that the measure expression is a set of entities whereas the count expression is a set of context-indexed entities. In the way that I have altered their semantics here, that distinction is not apparent, and there is only a distinction in the count/measure semantics.

3.5.6 Analysis

I propose that it is possible to account for two possible interpretations of *liang ping jiu* 'two bottles of wine' using the simpler semantics that I proposed in section 3.5.1. I am still assuming the syntax proposed in Li & Rothstein 2012.

I will use the same semantics of numerals that creates a partition, shown in (453a). I will also assume that Chinese mass and count nouns have the same distinct semantics as are found in Shan—count nouns denotes atomic join semi-lattices and mass nouns denote a non-atomic join semi-lattice. I will propose three denotations for the container term *ping* 'bottle'.

If I assume that the Chinese classifier is roughly equivalent to the Shan classifier, the different syntactic structures would simply lead to a difference of scope of the noun being measured with respect to the partition. Given that mass nouns are presumably the same whether or not they have been divided into portions, this would yield an identical interpretation for the count (CONTAINER+CONTENTS) and MEASURE readings.

If, instead, we assume that the CONTAINER+CONTENTS comes about as a result of the container contributing more nominal semantic content, we can distinguish the two types of classifiers by syntactic position. Therefore, I will assume that in the measure term position, the container classifier has the denotation in (453c), whereas in the count classifier position, it has the MEASURE denotation in (453d) (which is equivalent to a relational count noun).

(453) PROPOSED DENOTATIONS

a. $[\![liang]\!] = \lambda P \lambda x.\exists S\,[\Pi(S)(x) \wedge |S| = 2 \wedge \forall s \in S\,[P(s)]]$ (Ionin & Matushansky 2006)

b. $[\![jiu]\!] = \lambda x.\textsc{wine}(x)$

c. $[\![ping_m]\!] = \lambda x.\mu_{\textsc{bottle}}(x) = 1$

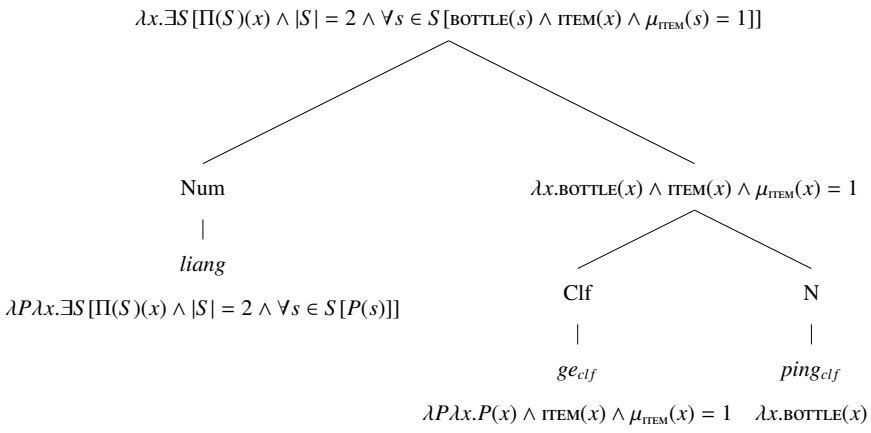
d. $[\![ping_{clf}]\!] = \lambda P \lambda x.\textsc{bottle}(x) \wedge \mu_{\textsc{bottle}}(x) = 1 \wedge \exists y[P(y) \wedge \textsc{contain}(x, y)]$

e. $[\![ping_n]\!] = \lambda x.{}^{*}\textsc{bottle}(x)$

Doubtless, there is more to be said about the relationship between the container classifier and its contents. I have included a simplified semantics where some relationship CONTAIN holds between the container and contents.

The fact that the container noun itself also requires a (morphologically distinct) classifier when it combines with a numeral supports the idea that it is the syntactic position of the classifier that is connected to the quantization of the container noun, not the lexical container noun itself. The count interpretation where the container is a head noun can be seen in (454).

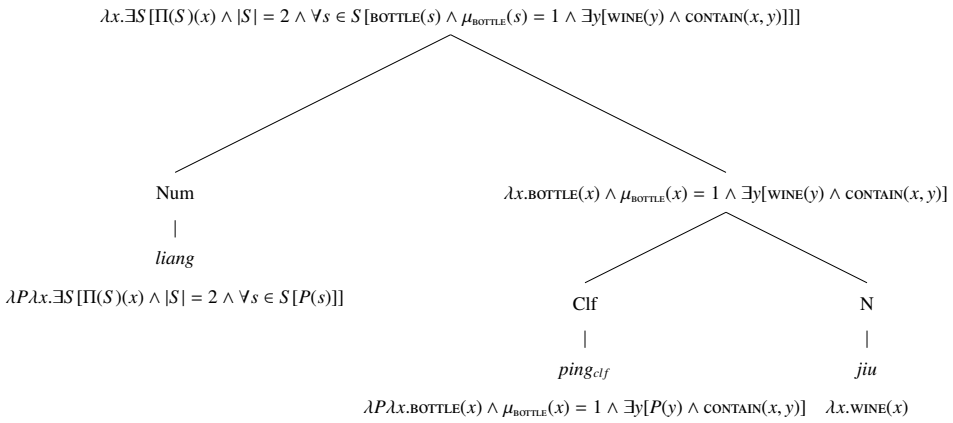
(454) Counting reading



The morpheme *de* that follows a numeral-classifier expression and enforces a measure reading tends to follows predicate, leading Li & Rothstein (2012) to propose that the numeral-classifier phrase in measure expressions is a predicate. This is still compatible with the analysis proposed here. The counting classifier does not denote a predicate and so would be incompatible with *de*.

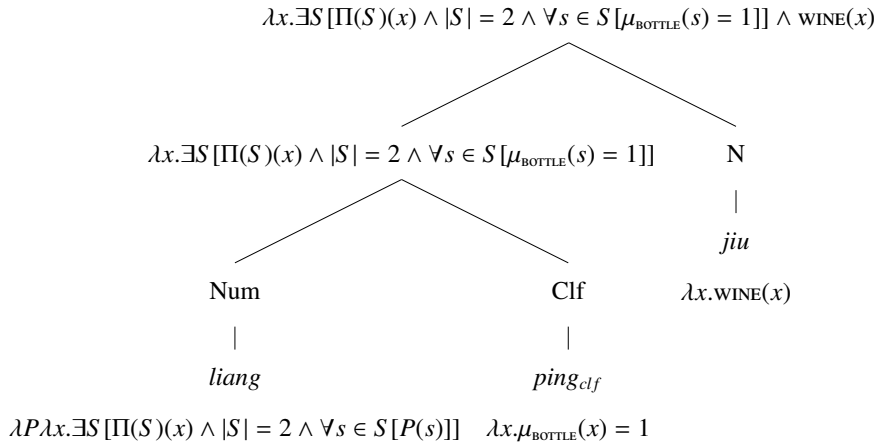
For the CONTAINER+CONTENTS derivation given in (455), the container noun and the contents noun form a constituent before combining with the numeral. The container noun has a measure component as part of the denotation. The numeral creates a partition, where each part is the size of a bottle. This would make Clf N constituents denote quantized predicates.

(455) Counting reading (CONTAINER+CONTENTS)



This contrasts with what occurs with the measure reading in (456). Here the numeral forms a constituent with the classifier, rather than the classifier forming a constituent with the noun. The type $\langle\langle e, t\rangle, \langle e, t\rangle\rangle$ modifying semantics of the 'count' classifier must be altered to be type $\langle e, t\rangle$ itself in order to combine with the numeral. With this adjustment it is simple to generate the distinct measure semantics where what is being denoted is the contents rather than the containers.

(456) Measure reading



Li & Rothstein 2012 that shows the numeral 'one' is optional for count interpretations, as in (457a) and obligatory for measure interpretations as in (457b). This is still compatible with the analysis of measure constructions proposed here. Under the count syntax, the classifier can

combine with a noun to form a quantized predicate, which can then be interpreted as a definite or indefinite singular, depending on other factors. The measure construction necessarily combines with the numeral before combining with the noun, so the numeral cannot simply be left out.

(457) a. wo mai le (yi) ben shu.
I buy PFV one $\text{CLF}_{\text{volume}}$ book
'I bought a book.' (Li & Rothstein 2012: (27a))

b. wo mai le *(yi) gongjin pingguo.
I buy PFV one CLF_{kilo} apple
'I bought a kilo of apples.' (Li & Rothstein 2012: (27b))

This section has shown that it is possible to explain the patterns of classifier data in Mandarin Chinese without needing to distinguish between what counting and measuring means. It is only necessary to propose differences in the syntax and semantics of measure terms in comparison to other classifiers, rather than requiring a distinct numeral denotation as well.

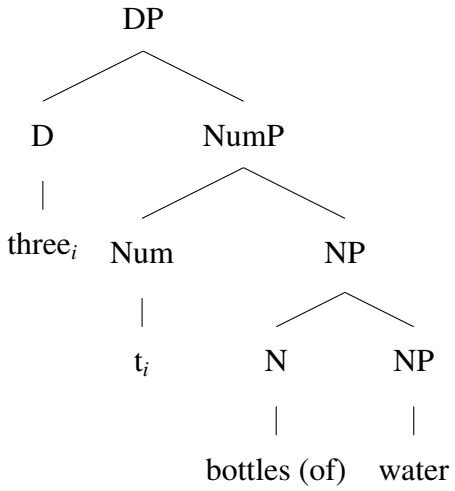
3.6 Chapter summary: Classifiers

I have proposed a unified underlying semantics for counting and measuring and claim that the distinction between measure and count interpretations comes from whether the entity denoted by the noun matches (COUNT) or not (MEASURE) with the unit of individuation. This captures the individuating function of the classifier and the Shan data presented here, wherein count/measure expressions are only grammatically distinguished by the head noun.

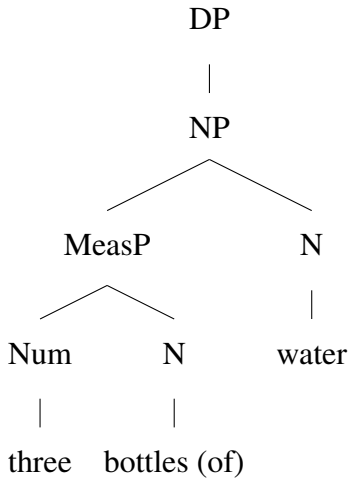
I argue that this unified count and measure semantics better captures the relationship between those two types of readings. It provides a simpler cross-linguistic analysis that is compatible with the variety of languages, including Nez Perce and Mandarin Chinese. Returning the structures proposed for count versus measure expressions in (439a) and(439b), repeated in (458a) and (458b), it is worth considering how numerals combine with cumulative nouns in English.

(458) Li & Rothstein 2012: (47a,b)

a. Counting reading



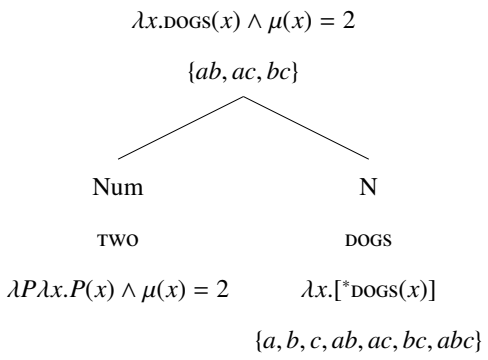
b. Measure reading



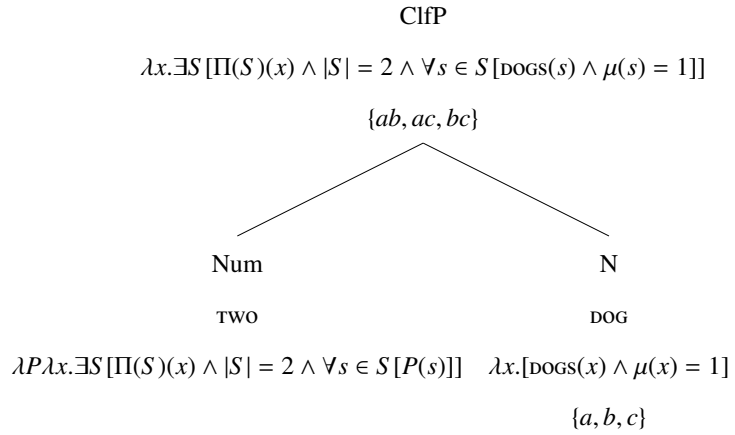
We could assume that the derivation of a numeral and count noun in English is the one shown in (459a). Alternatively, we could assume that the derivation is similar to the one in Shan, except that the plural morphology on *dogs* is only syntactically, not semantically relevant (Sauerland et al. (2005)).

(459) Num + Count Noun: English

a. Option 1: Plural noun (adapted from Little et al. to appear: (8))



b. Option 2: Singular noun

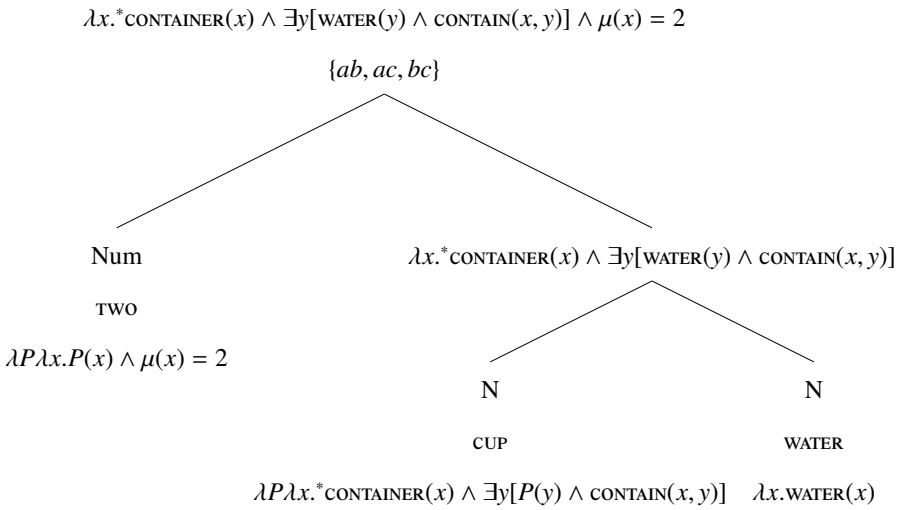


If we use the numeral semantics in Option 1 (459a) and the container noun semantics proposed in (419), repeated in (460), we can get the derivation in (461a) for the Container+Contents reading. It is more difficult to get the measure reading, even with a different syntax for the measure expression given in (461b).

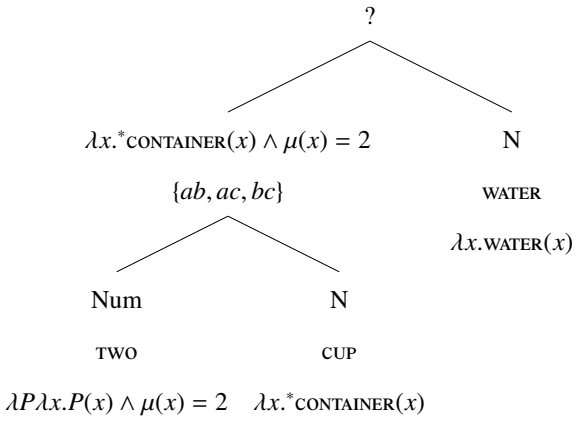
(460) Container N_1: $\lambda P\lambda x.{}^*\textsc{container}(x) \wedge \exists y[P(y) \wedge \textsc{contain}(x, y)]$

(461) Option 1

a. Container+Contents (adapted from Little et al. to appear: (8))



b. MEASURE



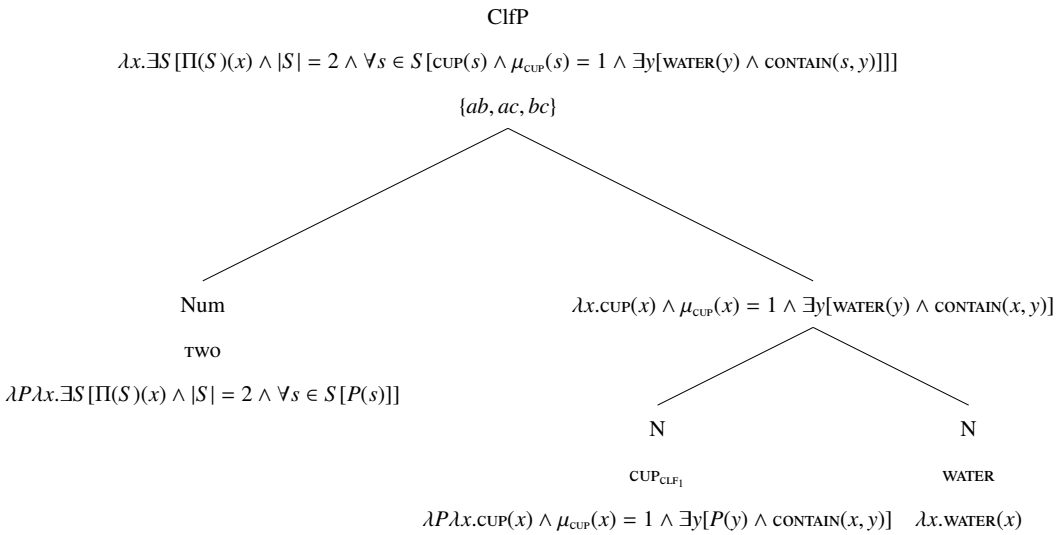
The classifier semantics from (453d) used for the CONTAINER+CONTENTS reading in Mandarin Chinese, shown in (462), can be used to get the same derivation in English, as demonstrated in (464a). If we use the container classifier semantics proposed for Shan in (422), repeated in (463), we can derive the MEASURE reading in (464b), using the same numeral semantics for both.

(462) $[\![\text{CUP}_{\text{CLF}_1}]\!] = \lambda P \lambda x.\text{CONTAINER}(x) \wedge \mu_{\text{CONTAINER}}(x) = 1 \wedge \exists y[P(y) \wedge \text{CONTAIN}(x, y)]$

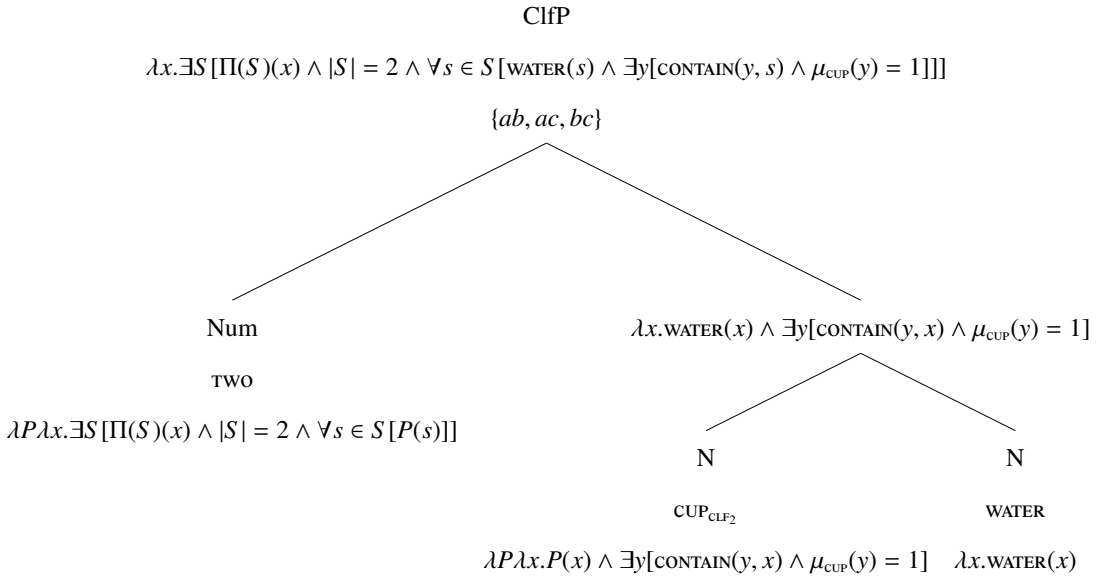
(463) $[\![\text{CUP}_{\text{CLF}_2}]\!] = \lambda P \lambda x.P(x) \wedge \exists y[\text{CONTAIN}(y, x) \wedge \mu_{\text{CONTAINER}}(y) = 1]$

(464) Option 2

a. CONTAINER+CONTENTS



b. MEASURE



I propose that the derivations for English CONTAINER+CONTENTS and MEASURE readings are best derived as in (464). Minimally, these examples show the connection between quantized nouns in English and container classifiers (and classifier more broadly) in Shan and Mandarin Chinese. This demonstrates the distinction between English and Shan nouns that I argued for in chapter 2 that mass nouns and *furniture* nouns in English pattern with Shan mass and count nouns, respectively. Cumulative nouns need something to quantize them before they can combine with a numeral. This might be another quantized noun in English or a classifier in Shan and Mandarin Chinese.

This chapter has investigated the syntax and semantics of a variety of classifier expressions in Shan and has argued that these expressions are largely syntactically the same. The semantics of the particular classifier expression is what generates most of the differences between classifier phrases.

CHAPTER 4

DEFINITENESS

DEFINITENESS AND QUANTIFICATION: EVIDENCE FROM SHAN

Mary Rose Moroney, Ph.D.

Cornell University 2021

This chapter demonstrates that bare nouns in Shan (Tai-Kadai) can express both unique and anaphoric definiteness, a distinction first noted by Schwarz (2009). This pattern of data as well as similar patterns found in Serbian and Kannada motivate adding a category to the typology of definiteness marking described by Jenks (2018) to include languages that allow bare nouns to express anaphoric definiteness. An extended type-shifting analysis offers an account of the availability of bare noun anaphora, as well as the other bare noun interpretations, such as the indefinite, generic, and kind interpretations. Variation in the use of bare nouns versus more marked anaphoric expressions is tied to pragmatic factors such as what anaphoric definite expressions are available in the language, contrast, information structure, and potential ambiguity of nominal expressions. This account proposes that a constrained semantic account of the interpretations of nominal expressions combined with a pragmatic account of their use can model much of the cross-linguistic data on definiteness. A version of this chapter was published in Moroney 2021.

4.1 Introduction

Bare nouns in languages without overt articles can be interpreted as indefinite, definite, generic, or as a kind (see, e.g., Mandarin; Yang 2001). Therefore, it is not surprising that bare nouns in Shan, a Southwestern Tai (Tai-Kadai) language of Myanmar which lacks overt articles, can have an indefinite, definite, generic, and kind interpretation, as discussed in chapter 2. The relevant data is repeated in (465). This chapter focuses on the expression of definiteness in Shan, though the proposed analysis will allow for all these interpretations.

(465) SHAN BARE NOUN INTERPRETATIONS

a. măa hàw jù.
dog bark IPFV
'Dogs are barking.' *indefinite*

'The dog(s) is/are barking.' *definite*

b. măa hàw.
dog bark
'Dogs bark.' *generic*

c. măa wɔt.wáaj.hăaj kwàa jâw.
dog disappear go finish
'Dogs are extinct.' *kind*

Based on data from German, Schwarz (2009) proposed splitting 'definiteness' into two types: uniqueness and familiarity. Uniqueness means uniqueness within a non-linguistic context, and familiarity refers to discourse anaphora. Building on this proposal, Jenks (2015), (2018) argued that bare nouns in Thai and Mandarin only express one kind of definiteness: unique definiteness. He offers the typology of definiteness in Table 4.1.

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

Table 4.1: Typology of definiteness marking (adapted from Jenks 2018)

	Both marked		One marked	
	same	different	unique	anaphoric
Unique (ι)	Def	Def_{weak}	Def_{weak}	$\emptyset$
Anaphoric (ι^x)	Def	$\text{Def}_{\text{strong}}$	$\emptyset$	$\text{Def}_{\text{strong}}$
Languages	Cantonese, English	German, Lakhota	(*unattested*)	Mandarin, Akan, Wu

(466) The Typology of Definiteness Marking

i. Both marked, same (Generally Marked): The primary strategy of definiteness marking is used in both unique and anaphoric contexts.

ii. Both marked, different (Bipartite): Unique and anaphoric definiteness are marked using different definiteness marking in at least some contexts.

iii. One marked, uniqueness (Marked Unique): Unique definiteness is obligatorily marked but anaphoric definiteness is not. (*unattested*)

iv. One marked, anaphora (Marked Anaphoric): Unique definiteness is not marked, but anaphoric definiteness marking is obligatory.

In this typology some languages, like German and Lakhota, differentially mark unique and anaphoric definiteness. Some languages, like Mandarin (Jenks 2018), Akan (Arkoh & Matthewson 2013), and Wu (Li & Bisang 2012; Simpson 2017), only mark anaphoric definiteness. English marks both types of definiteness using the same morpheme, meaning that *the* is compatible with both unique and anaphoric definiteness. Jenks (2018) claimed that there are no languages attested where only unique definiteness is morphologically marked.

This chapter explores the expressions of definiteness using bare nouns and makes the claim that Shan bare nouns can express both unique and anaphoric definiteness. This means that there should be one more category in Jenks's (2018) typology of definiteness, a category with 'unmarked' unique and anaphoric definiteness. It is expected that other languages would fall in this unmarked category and use bare nouns to express both unique and anaphoric definiteness. This

chapter provides evidence that Shan falls into this category and that other languages, such as Serbian and Kannada, do as well.

Section 4.2 discusses the unique and anaphoric types of definiteness identified by Schwarz (2009), primarily looking at data from Thai as discussed by Jenks (2015). Section 4.3 introduces data demonstrating that bare nouns in Shan, a language related to Thai, can express both unique and anaphoric definiteness. Section 4.4 discusses the typology of definiteness marking adding data from Serbian and Kannada. Section 4.5 presents a type-shifting analysis for bare nouns in Shan that uses two ι type-shifting operators and identifies a problem with using the Consistency test introduced by Dayal (2004) to decide whether a word counts as a determiner for the Blocking Principle. Additionally, this section proposes to use the economy principle *Don't Overdeterminate!* from Ahn (2019) to explain the variation in choosing bare nouns or other anaphoric expressions within a language. Section 4.6 discusses some remaining issues connected to the typology of definiteness marking, including the role of contrast and ambiguity, and Section 4.7 concludes.

4.2 Two kinds of definiteness: Unique and anaphoric

4.2.1 Background

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

the *weak* form, and the un-reduced form with the preposition + definite article combination, called the *strong* form, overtly represent unique and anaphoric definiteness, respectively.

The difference between these two forms can be seen in example (467). This example involves a unique definite context since it is common ground that there is only one mayor in this context. The uniqueness of the mayor in the situation triggers the use of the weak definite article form, *vom* ('by the'), and the strong form, *von dem*, is infelicitous. In contrast, the strong form is obligatorily used in familiar/anaphoric contexts.

(467) Weak versus strong articles in German

Der Empfang wurde vom / #von dem Bürgermeister eröffnet.
the reception was by-the$_{weak}$ / #by the$_{strong}$ mayor opened

'The reception was opened by the mayor.' (Schwarz 2009: (42))

Building on the categories of definiteness described by Hawkins (1978), Schwarz (2009) identified several contexts for definite expressions: immediate situation (local non-linguistic context), larger situation (global non-linguistic context), anaphoric/familiar, and bridging (associative anaphora). To these, Schwarz (2009) added a category for donkey anaphora. According to Schwarz (2009), German uses the weak form of the definite article in the contexts of situational uniqueness, which includes uniqueness in an immediate situation or larger situation, as well as in a type of bridging called part-whole bridging. The strong form of the definite article is used in anaphoric contexts, producer-product bridging, and donkey anaphora—which Schwarz (2009) argued all involve a kind of anaphora. Table 4.2 gives examples of these categories and the article form used for German. Most of these will be discussed more in section 4.2.2. This chapter will not include the evidence from bridging anaphora other than to note in the tables that it follows the predictions of Schwarz (2009).

Table 4.2: Types of definiteness described by Schwarz (2009), citing Hawkins (1978)

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

4.2.2 Uniqueness versus familiarity/anaphoricity

Schwarz (2009) argued that the weak definite article expresses *unique definiteness*. This means that the intended referent is unique in an immediate situation, as in (468), or in a larger or global context. In (468), there is only one glass cabinet in the immediate context, so it is expressed using the weak form of the definite article, *im*.

(468) German: Unique in immediate situation

Das Buch, das du suchst, steht im / #in dem Glasschrank.
the book that you look-for stands in-the$_{weak}$ / #in the$_{strong}$ glass-cabinet

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

Schwarz (2009) proposed that the strong definite article, in contrast, expresses *familiarity* or *anaphoricity*. This includes either being perceptually or generally familiar, or part of the preceding discourse. (469) provides an example with discourse anaphora. The second sentence must use the strong form of the definite article, *von dem Politiker* ('from the politician'), to refer back to the politician that was introduced in the first sentence.

(469) German: Anaphora

Hans hat einen Schriftsteller und einen Politiker interviewt. Er hat #vom / von
Hans has a writer and a politician interviewed He has #from-the$_{weak}$ / from

dem Politiker keine interessanten Antworten bekommen.
the$_{strong}$ politician no interesting answers gotten

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

This phenomenon is not limited to German. Schwarz (2013) found that the strong/weak contrast is apparent in many languages. Jenks (2015), (2018) showed that Mandarin and Thai use bare nouns in the same places where German would use the weak definite article and phrases with a noun, classifier, and a demonstrative (N Clf Dem) where German would use the strong definite article. I will be referring to expressions that include a noun and demonstrative, with or without a classifier, as 'demonstrative-noun phrases'.[1]

Examples (470) and (471) show the use of the bare noun in situations of unique definiteness in Mandarin and Thai, respectively. *Gou* 'dog' in (470) and *măa* 'dog' in (471) must refer to the unique dog or dogs in the context.

(470) Mandarin: Unique in immediate situation

Gou yao guo malu.
dog want cross road

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

(471) Thai: Unique in immediate situation

măa kamlaŋ hàw.
dog prog bark

'The dog is barking.' (Jenks 2015: (2))

[1]In chapter 2, I claimed that N Dem expressions can be interpreted as kinds, but I do not include any of those examples in this chapter.

(472) and (473) show how the demonstrative-noun phrase is used to express discourse anaphora in Mandarin and Thai. In the Mandarin example in (472a), a boy and a girl are first introduced into the discourse context. Then in (472b) and (472c), the demonstrative-noun phrase *na ge nansheng* ('the/that boy') refers back to the boy. According to Jenks (2018), bare nouns in Mandarin can express anaphoric definiteness in certain contexts, namely in the subject position. The classifier and demonstrative are optional in subject position, but not in object position, as shown in (472b) and (472c). Jenks (2018) claimed that this is because the Mandarin subject is a topic, and topic marking negates the effect of *Index!*, the requirement that an indexical expression be used whenever possible. This will be discussed further in section 4.5.2. The connection between topics and using bare nouns or weak definites to express anaphora has been noted before. For example, German can use a weak definite article to refer back to a referent that is a topic (Schwarz 2009: 47).

(472) Mandarin: Narrative sequence (Anaphoric)

a. Jiaoshi li zuo-zhe yi ge nansheng he yi ge nüsheng,
classroom inside sit-PROG one CLF boy and one CLF girl
'There are a boy and a girl sitting in the classroom...'

b. Wo zuotian yudao #(na ge) nansheng
I yesterday meet #(that CLF) boy
'I met the boy yesterday.'

c. (na ge) nansheng kanqilai you er-shi sui zuoyou.
(that CLF) boy look have two-ten year or-so
'The boy looks twenty-years-old or so.' (Jenks 2018: (15a,b,d))

Example (473) from Thai demonstrates anaphoric reference across two sentences. The first sentence introduces a student into the discourse context using *nákrian khon nɨŋ* 'one student'. In (473a), the demonstrative-noun phrase *nákrian khon nán* ('that boy') refers to back to that student. In order to have the anaphoric reading in Thai, Jenks (2015) says the demonstrative construction is required even in subject position. In this way, Mandarin and Thai differ in their choices of anaphoric expressions.

(473) Thai: Narrative sequence (Anaphoric)

mîawaan phŏm cəə kàp nákrian khon nɨŋ.
yesterday 1 meet with student clf indef

'Yesterday I met a student'

a. nákrian #(khon nán) chalàat mâak.
student #(clf that) clever very
'That student was very clever.' (Jenks 2015: (17))

4.2.3 Donkey anaphora

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

In (474), a referent is introduced by *khwaai tua nɨŋ* 'one buffalo' in the universally quantified relative clause, and a demonstrative-noun expression must be used to refer back to the buffalo that each farmer has. Using a bare noun to refer back to the buffalo gives the sentence a generic meaning: 'Every farmer that has a buffalo hits buffalo'.

(474) Thai: Donkey anaphora

chaawnaa thúk khon thîi mii khwaai tua nɨŋ tii khwaai tua nán
farmer every clf that have buffalo clf indef hit buffalo clf that

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

As demonstrated here, there is a contrast between unique and anaphoric definiteness in Thai and Mandarin, where unique definites are expressed with bare nouns and anaphoric definites are

expressed with demonstrative-noun phrases, with the exception of Mandarin subjects. In Table 4.3 is a summary of the required definite expressions in certain contexts in German, Thai, and Mandarin. Examples of all the contexts described by Schwarz (2009) for all three languages can be found in the cited sources.

Table 4.3: Expressions of definiteness in German, Thai, and Mandarin

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

4.3 Two kinds of definiteness with Shan bare nouns

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

4.3.1 Background on Shan

As noted in Chapter 2, Shan has SVO word order, as in (475) and is a number neutral languages as (476) demonstrates.

(475) háw hăn măa
1 see dog
'I see a dog/dogs.'

(476) tinaj mí măa
here have dog
'Here there is a dog/are dogs.'

When a numeral combines with a noun, a classifier must also appear with the phrase, as in (477)–(478). The same is true for mass nouns: when there is a numeral combining with the noun, there must also be a measure word, as in (479))–(480). Classifiers vary based on properties of the noun they combine with. The classifier for animals is *tŏ*, as in (477)–(478), and the classifier for humans is *kə̂*.

(477) măa nɯɯŋ *(tŏ)
dog one CLF.ANML
'one dog'

(478) măa săam *(tŏ)
dog three CLF.ANML
'three dogs'

(479) nâm nɯɯŋ *(kɔ́k)
water one cup
'one cup of water'

(480) nâm săam *(kɔ́k)
water three cup
'three cups of water'

Although some languages allow the classifier to appear alone with the noun (see Simpson et al. 2011), Shan does not allow this, as shown in (481). Classifiers in Shan seem to be derived from nouns. For example, as (482) shows, *tŏ*, the classifier for animals, is also the word for 'body'.

(481) N CLF

*măa tŏ
dog CLF.ANML

intended: 'the dog'

(482) CLF N

tŏ măa
CLF.ANML dog

'dog body', *not* 'the dog'

The demonstrative in Shan can either appear directly with the noun, as in (483), or with a classifier, as in (484). The difference between (483) and (484) in meaning is that (483) can refer to singular and plural definite objects, and (484) can only refer to a singular definite object.

(483) N DEM

măa nân
dog that

'that dog/those dogs'

(484) N CLF DEM

măa tŏ nân
dog CLF.ANML that

'that dog'

(485) gives the plural version of (484). The plural morpheme was discussed in section 3.4.4.

(485) N PL DEM

mǎa tsɤ́ nân
dog CLF.PL that

'those dogs'

(486) PL N

tsɤ́ mǎa
CLF.PL dog

'the group of dogs'

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

4.3.2 Uniqueness versus familiarity/anaphoricity in Shan

Shan patterns like Mandarin and Thai in that in unique situations, such as are shown in (487) and (488), a bare noun must be used. In (487), there is only one teacher in the context, so that teacher must be referred to using a bare noun. For (488), we know that there is only one sun in our global context. Consequently, a bare noun must be used to refer to the sun. The demonstrative is not allowed in either case. In these examples, the entity described by the bare noun (i.e., the teacher or sun) has not been mentioned before in the linguistic context, so these examples do not involve discourse anaphora.

(487) SHAN: UNIQUE IN IMMEDIATE SITUATION

(Context: classroom with just one teacher)

Náaŋ Lɤ̌n ʔàm tsaaŋ kwàa hǎa khúsɔ̌n (#kô nân).
Ms. Lun NEG able go find teacher (#CLF.HUM that)

'Nang Lun cannot find the teacher.' (Moroney 2019a: (11))

(488) SHAN: UNIQUE IN LARGER SITUATION

kăaŋwán (#hòj nân) lŏŋ hɤ sɔ̀ŋ.
sun (#CLF.RND that) very bright glitter

'The sun is very bright.' (Moroney 2019a: (12))

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

In contrast to Mandarin and Thai, Shan can use a bare noun in anaphoric contexts. The narrative sequence in (489) demonstrates this. In the first sentence, a man is introduced into the discourse context. Another individual, the store owner, is also introduced in the story in order to make *phu-tsáaj* 'man' a more natural anaphoric expression to use. Schwarz (2009) and Simpson et al. (2011) use a similar strategy when looking into anaphoric examples. This is because if a speaker introduces only one individual, she is much more likely to refer back to that individual using a pronoun, rather than a nominal expression. This will be discussed more in section 4.5.4. In (489), since at least two discourse referents are available, anaphoric reference to the man is made either with a bare noun, *phu-tsáaj* 'man', or with a demonstrative-noun phrase (N Clf Dem), *phu-tsáaj kɔ̂ nân* 'that man'.

(489) SHAN: NARRATIVE SEQUENCE (ANAPHORA)

phu-tsáaj kɔ̂ nɯŋ kwàa ti hâan khăaj măa tàa sûɯ măa ʔɔ̀n tŏ nɯŋ păn
person-man CLF.HUM one go at store sell dog for buy dog small CLF.ANML one give

lukjíŋ mán-tsáaj... phu-tsáaj (kɔ̂ nân) khɯ́n tɔ̀p waa,
daughter 3-man person-man (CLF.HUM that) back respond that

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

(Moroney 2019a: (13))

An example with inanimate referents can be seen in (490). In the first clause, three inanimate things are introduced: *kɔ́k kɔ̀fì* 'a cup of coffee', *phɤ̌n* 'a table', and *pâplik* 'a book'. In the second clause, the cup of coffee and the book can be referred back to using a bare noun or a demonstrative-noun phrase.[2]

[2]While it is possible to use a demonstrative-noun phrase for either noun being used anaphorically, it is awkward to

(490) Shan: Cross-clausal Anaphora

jôn kɔ́k kɔ̀fî mí nɤ̌ phɤ̌n mí hímtsăm pâplik lɛ, háw laj ʔăw kɔ́k kɔ̀fî (nân)
because cup coffee have on desk have near book and 1 get take cup coffee (that)

he sàj pâplik (nân)
spill in book (that)

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

Previous work has proposed that contrast can license bare noun anaphora (Jiang 2012; Jenks 2018). In particular, Jenks (2018) discusses that contrastive topics (Büring 2003) license bare noun anaphora in Mandarin. These contrastive contexts occur when there is 'a topical set of alternatives relevant to a particular QUD' (Jenks 2018: 526). This is different from Ahn's (2019) discussion of contrast in connection with the use of bare nouns in languages where a morphologically simplex pronoun otherwise blocks bare noun anaphora. In such languages, the presence of multiple salient entities that would be referenced using the same pronoun, creates a situation where a bare noun can be used since a pronoun would not unambiguously identify the intended referent. Some might think that (490) only allows bare noun anaphora because there are two contrastive salient entities. The notebook and cup of coffee are not contrastive topics in the sense of Büring (2003) because Shan uses specific morphemes to indicate contrastive topics.

It is possible to see bare noun anaphora in contexts where there is only one salient, animate entity, as in (491). This example further involves an anaphoric expression in object position. In (491), the classifier and demonstrative are optional. In this example, the squirrel is the only significant individual introduced into the narrative, even though other objects, such as the tree and the storm, are discussed.[3] The anaphoric reference to the squirrel in (491) does not occur in a highly contrastive context, given that it is the only animate, third-person entity in the narrative. There

use the demonstrative with both at the same time. The effect of using the demonstrative here seems to be to emphasize the noun it combines with, so it may seem unnatural to emphasize both.

[3]However, it is likely relevant that there are other entities discussed in the example because otherwise, a pronoun or null pronoun might be more natural to use. Animals and non-animate entities are referred to using the same 3rd person pronoun, *mán*, so the pronoun might not uniquely identify the intended referent, the squirrel. This is related to Ahn's (2019) notion of contrast. The relevance of this will be discussed further in section 4.5.4.

also certainly is no contextually salient set of alternatives that are relevant to a QUD. This supports the idea that in Shan the bare noun anaphora is not licensed only by contrast in the sense of Jenks (2018).

(491) SHAN: NARRATIVE SEQUENCE (ANAPHORA)

tsɔn tǒ nɯŋ máa mí nɤ̌ tonmâj ʔǎn mí hímtsăm hɤ́n háw nân sě
squirrel CLF.ANML one come have on tree COMP have near house 1 that and

mɤnâj phǒn lóm haaŋ hâaj nàa lɛ hét haj kìŋ-mâj jàat tók njăa tě
today rain wind appearance bad very and do cause branch-tree break fall almost IRR

phât njáa tsɔn (tǒ nân) páa jâw
hit meet squirrel (CLF.ANML that) with finish

'A squirrel was on the tree near my house. Then one day, a bad storm caused a branch to break and fall almost hitting the squirrel.' (Moroney 2019a: (15))

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

4.3.3 Donkey anaphora

German, Thai, and Mandarin all use a strong/demonstrative form to refer back to a nominal in cases of donkey anaphora. However, Shan does not obligatorily use a demonstrative or strong definite article to express donkey anaphora. (492) gives an example of donkey anaphora in the second sentence. This example uses the conditional construction. The antecedent clause in (492)

introduces *mέw* 'cat' and *nŭ* 'mouse' using bare nouns and in the consequent the same bare nouns refer back to them.[4]

(492) SHAN: CONDITIONAL

mέw lɛ nŭ ʔàm mɛn kăn. pɔ́ mέw hăn nŭ năj-tsŭŋ, mέw lɯp lám tî
cat and rat NEG right together if cat see rat then cat follow chase grab

njɔ́p nŭ tàasè
snatch rat always

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

A demonstrative-classifier-noun phrase can be used more naturally as the anaphoric component of a donkey anaphora sentence in an example like (493), but the demonstrative and classifier are not necessary. In (493), *tŏ lăj* ('which one') quantifies over individual cats, and the demonstrative-classifier-noun phrase can refer back to each cat.

(493) SHAN: DONKEY ANAPHORA

măa nâj hăn mέw tŏ lăj kɔ tĕ lɯp mέw (tŏ nân) tàasè.
dog this see cat CLF.ANML which PRT will follow cat (CLF.ANML that) always

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

(Moroney 2019a: (20))

In donkey anaphora constructions, my consultants did not typically use *mέw tŏ* nɯŋ 'one cat' in the antecedent clause. Shan does not have an indefinite article but uses the numeral 'one' in some cases where English uses an indefinite article. If 'only' is used, it is possible to use *mέw tŏ* nɯŋ in the antecedent, and in that case it is necessary to use a demonstrative to refer back to it, as in (494). When 'one cat' is used in a donkey anaphora construction, it sounds as if the number of cats is relevant to whether the dog chases the cat, in the same way it does in the English translation of (494).

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

(494) SHAN: DONKEY ANAPHORA

mă̌a ku tǒ nâj pɔ́ hă̌n mɛ́w tǒ nɯɯŋ kój kɔ tě lɯp lám
dog every CLF.ANML this if/when see cat CLF.ANML one only PRT will follow chase

mɛ́w *(tǒ nân) tàasè.
cat *(CLF.ANML that) always

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

(Moroney 2019a: (21))

Table 4.4: Expressions of definiteness in German, Thai, Mandarin, and Shan

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

Table 4.4 summarizes the range of definite expressions found in German, Thai, Mandarin, and Shan. This section has investigated the expression of definiteness in Shan in different kinds of definite contexts. In Shan, bare noun can be used in all of the contexts described by Schwarz (2009). Demonstrative-noun phrases are compatible with the contexts that require the strong definite form in German, Thai, and Mandarin, but they are not obligatory. The bridging anaphora data is not included here, but the data from Shan is consistent with the pattern discussed so far. Bare nouns can appear in both part-whole and producer-product bridging. Even contexts like cross-sentential and donkey anaphora allow for bare nouns.

4.4 A revised typology of definiteness

This section argues for a revised typology of definiteness, adding a category where anaphoric definiteness is unmarked to the typology given above in Table 4.1. As shown in Section 4.3, Shan bare nouns can express both unique and anaphoric definiteness. In addition to Shan, there are several languages that could potentially fit into this unmarked category, including Serbian (§4.4.1) and Kannada (§4.4.2). Section 4.4.3 revisits some previous work on the typology of definiteness marking to discuss where other languages might fit in the expanded typology. Section 4.4.4 gives the proposed revised typology.

4.4.1 The case of Serbian

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

Example (495) demonstrates that a bare noun in object position can refer anaphorically in Serbian. (495) gives an anaphoric bare noun example using an inanimate noun. After the noun, *novanik* 'wallet', is introduced, it is referred back to several times in the story using a bare noun. The first time it is referred to anaphorically, the bare noun is the object of the verb *uzela* 'took'. This kind of anaphora using a bare noun is possible with animate nouns as well, such as *policajca* 'policeman'.

(495) Serbian: Cross-sentential Anaphora

Isidora Stojanović, . . . pronašla je novčanik pun para gde se nalazilo 4000
Isidora Stojanovic found is wallet full money where refl. located 4000

evra i 16.000 dinara a Isidora se nije dvoumila šta treba da učini. Uzela
euros and 16.000 dinars and Isidora self did-not think_twice what needs that do Took

je novčanik i otišla do MUP-a i predala novac...
is wallet and left to police_station and turned_in money

'Isidora Stojanović, ... found a wallet full with money, which contained 4000 euros and 16.000 dinars. Isidora did not think twice what she needs to do. She took the wallet and went to the police station and turned in the money... "[5]

Bare nouns can also be used in donkey anaphora examples in Serbian, as shown in (496).[6] In these examples, it is better if there are two possible antecedents for the donkey anaphor, since a pronoun is often preferred to using a bare noun when expressing anaphora.

(496) Serbian: Donkey Anaphora

Moj jednogodišnji sin ponekad radi neočekivane stavi. Recimo, svaki put kada
My one-year-old son sometimes does unexpected things. For instance, every time when

vidi kofícu i loptu, on pokuša da stavi kofícu na loptu.
sees bucket and ball he tries to put bucket on ball.

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

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

[5]Source: http://www.orsbap.com/test/rukometasica-sloge-isidora-stojanovic-vratila-novcanik-pun-para/

[6](496) is a sentence provided by one linguist native speaker of Serbian and judged as correct by another native speaker.

4.4.2 The case of Kannada

In a recent presentation, Srinivas & Rawlins (2020) demonstrated that Kannada bare nouns express both unique and anaphoric definiteness in much the same way that Shan bare nouns do. For example, (497) demonstrates that the anaphoric noun *pustaka-(d)alli* 'in the book' does not have to be a subject. It is also not simply a unique definite, as there are many other books in the library.

(497) Kannada: Cross-sentential Anaphora

Bengaloor-(i)na doDDa granthalaya-(d)alli halsinahaNN-(i)na bagge ondu pustaka ide.
Bangalore-GEN big library-LOC jackfruit-GEN about one book COP

Monne naanu alli-ge hogiddaaga pustaka-(d)alli halsinahaNN-anna he:ge kariyadu
Recently I-NOM there went book-LOC jackfuit-ACC how fry

anta noDide
that saw

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

(Srinivas & Rawlins 2020: (9))[-30pt]

Even donkey anaphora sentences, such as (498)–(499) use a bare noun to refer back to the donkey introduced in the first part of the sentence.

(498) Kannada: Donkey Anaphora

katte-annu uLLuva pratiyobba raitanuu katte-ge uuTa haakuttaane
donkey-ACC having every farmer-EMPH donkey-DAT food puts

'Every farmer who has a donkey feeds the donkey.'

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

(499) Kannada: Donkey Anaphora

Raita-na hattira katte iddare avanu katte-ge ooTa haakuttaane
Farmer-GEN near donkey have.if he donkey-DAT food puts

'If a farmer has a donkey, he feeds the donkey.'

(Srinivas & Rawlins 2020: (8), added emphasis)

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

4.4.3 Filling in the typology

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

Ahn (2019) discusses several languages that could fall into the unmarked definiteness category. For example, Korean and American Sign Language (ASL) appear to allow bare nouns to express anaphoric definiteness fairly robustly. An example of Korean is shown in (500), here *namca-ka* 'man' is introduced as the subject in the first sentence and referred back to with a bare noun *namca-lul* as the object in the second sentence. Case marking accounts for their different forms. In ASL, simple inter-sentential anaphora might suggest that bare nouns cannot be anaphoric as shown in (501). However, (502) shows that a bare noun is acceptable to use in some cases of anaphoric reference. Korean and ASL would then be languages that do not obligatorily overtly mark anaphoric definiteness.

(500) namca-ka tulewa-ss-ta. na-nun namca-lul chyetapwa-ss-ta.
man-NOM enter-PAST-DECL I-TOP man-ACC stare-PAST-DECL
'A man entered. I stared at the man.' (Korean, Ahn 2019: (24), added emphasis)

(501) JOHN BUY IX$_A$ BOOK, IX$_B$ MAGAZINE. #(IX$_A$) BOOK EXPENSIVE.
'John bought a book and a magazine. The book was expensive.'
(ASL, Ahn 2019: (333), citing Irani 2016; see Irani 2019: (27), added emphasis)

(502) BOY$_i$ ENTER CLUB. MUSIC ON. BOY$_i$ DANCE.

'A boy$_i$ entered a club. Music was on. The boy$_i$ danced.'

(ASL, Ahn 2019: (330), Condition A for N, format and emphasis mine)

One Marked or Unmarked?

Jenks (2018) argued that Mandarin is a 'marked anaphoric' language, where the demonstrative is the primary marker of anaphora. While bare nouns can be used in anaphoric contexts in subject position, Jenks (2018) argued that they are only able to do so because they are functioning as topics.

The previous sections have discussed specific languages from different language families that have data available to demonstrate that bare nouns can be used to express anaphoric definiteness. There are several other languages, such as those mentioned by Despić (2019)—Japanese, Hindi, and Turkish—that have a bare noun/marked anaphoric definiteness contrast. All of these languages have a way of overtly marking anaphoric definiteness. What is not clear is whether this morpheme is obligatory in cases of anaphoric definiteness. This obligatoriness is what distinguishes languages with an anaphoric definite determiner from ones that do not have any overt definite determiners. This distinction has implications for the range of possible definite interpretations of bare nouns and overt lexical instantiations of unique and/or anaphoric definiteness. The semantic account of bare noun definiteness that will be proposed in section 4.5 is constrained by the lexical determiners found in the language.

Schwarz (2013) identifies two languages that use bare nouns to express unique definiteness and have a determiner that is only used to express anaphoric definiteness. These languages are Akan and Mauritian Creole. Schwarz (2013) notes that the anaphoric definite article is found largely in the same places as the strong definite article in Germanic languages. In addition to Thai (Jenks 2015) and Mandarin (Jenks 2018)—which will be discussed further in section 4.5.4, Jenks (2018)

also classifies Akan as a language with an anaphoric definite article. Jenks (2018) further identifies Wu (Tibeto-Burman) as such a language.

The data available for Mauritian Creole from Wespel 2008 show that the determiner *la* is used in anaphoric and not unique definite contexts. There is no data that shows that bare nouns can express anaphoric definiteness. The data in (503) demonstrates that the determiner is used even in an apparent case of contrastive focus, which supports treating *la* as an anaphoric definite determiner in Mauritian Creole.

(503) Enn garson ek enn tifi ti pe lager. Garson la ti paret an koler, tifi la ti res
one boy and one girl PST PROG argue boy DEF PST appear in rage girl DEF PST stay
kalm.
calm

'A boy and a girl were arguing. The boy seemed furious, the girl stayed calm.'

(Wespel 2008: 143)

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

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

and Jinyun Wu support categorizing these languages as marking anaphoric definiteness. If bare nouns do express anaphoric definiteness in these languages, they should instead be characterized as being unmarked for definiteness.

Classifiers as definiteness markers

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

Among the languages Simpson et al. (2011) describe, Hmong seems to pattern clearly as a language that marks both unique and anaphoric definiteness using a classifier-noun phrase. Li & Bisang (2012) note that the Hmong classifier seems to be a highly grammaticalized marker of definiteness, supporting the language's inclusion in the typological category that marks both kinds of definiteness. The classifier-noun phrase, the overt marker of definiteness, was preferred in all contexts. The data from Bangla suggest that Bangla is a marked anaphoric language like Akan. The classifier-noun phrase was rated as more grammatical in anaphoric contexts and the bare noun was rated as more grammatical in unique contexts. A more detailed investigation of Bangla largely seems to confirm that (Simpson & Biswas 2016). As for Cantonese, which Jenks

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

(2018) calls a 'generally marked' language, the Hong Kong Cantonese speakers are consistent with what one would expect from a language that uses the same definiteness marking for both kinds of definiteness, but the Malaysian Cantonese speakers patterned closer to the Bangla speakers.

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

Both Marked

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

English falls into the category where both unique and anaphoric definiteness are marked using the same determiner, *the*. As discussed in section 4.1, Schwarz (2009) identified that some Germanic languages mark both unique and anaphoric definiteness using separate definite articles. Other languages that express unique and anaphoric definiteness using two distinct overt morphemes include Lakhota and Hausa (Schwarz 2013).

Other patterns of definiteness

Some languages are more difficult to fit into the typology. For example, Schwarz (2013) notes that Haitian Creole shows a type of definite contrast different from the unique/anaphoric one. Haitian

Creole typically uses the same definite article to express both unique and anaphoric definiteness, but there are certain definite contexts where a determiner is not used.

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

4.4.4 Summary: A revised typology

A revised typology of definiteness marking is given in Table 4.5. This table adds the category 'Unmarked' definiteness to the typology and puts Shan, Serbian, and Kannada in that category.

Table 4.5: Revised typology of definiteness marking

	Both marked		One marked		Unmarked
	same	different	unique	anaphoric	
Unique	Def	Def_{weak}	Def_{weak}	∅	∅
Anaphoric	Def	$\text{Def}_{\text{strong}}$	∅	$\text{Def}_{\text{strong}}$	∅
Languages	Cantonese, English	German, Lakhota	(*unatt.*)	Akan, Wu, Mauritian Creole	Shan, Kannada, Serbian

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

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

nouns in the typology of definiteness marking.

4.5 Analysis

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

Section 4.5.1 gives the background for the type-shifting analysis used to derive the interpretations of bare nouns. Section 4.5.2 extends the analysis proposed by Jenks (2018) that distinguishes between unique and anaphoric definiteness. This allows us to explain why some languages obligatorily mark anaphoric definiteness and some do not. Section 4.5.3 describes predictions of definiteness marking for the typology based on the type-shifting analysis. Section 4.5.4 discusses how an economy constraint connected to the definite expressions available in a language can account for variation of definiteness marking within languages. Section 4.5.5 summarizes this section.

4.5.1 Type-shifting

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

(Deal & Nee 2018).

Languages like Shan that lack overt definite determiners frequently use bare noun arguments. Languages like English that have overt determiners use bare noun arguments in a more restricted way. Only English plural and mass nouns can be bare arguments, as shown in (504)–(505).

(504) Dinosaurs are extinct.

(505) Gold is valuable.

Chierchia (1998) proposed a Neo-Carlsonian approach following Carlson (1977), claiming that bare plurals in English are type $\langle e, t\rangle$, and they can type-shift to function as arguments. The type-shifting operators defined by Chierchia (1998) and updated by Dayal (2004), are given in (506). Chapter 2 discussed my approach to bare nouns. It is compatible with these type-shifters.

(506) Type-shifting operators (Dayal 2004):

a. $^{\cap}$: $\lambda P\ \lambda s\ \iota x[P_s(x)]$

b. ι: $\lambda P\ \iota x[P_s(x)]$

c. $\exists$: $\lambda P\ \lambda Q\ \exists x[P_s(x) \wedge Q_s(x)]$

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

Using Dayal's (2004) update of Chierchia's (1998) analysis, Deal & Nee (2018) discuss the predicted available interpretations for bare nouns in Teotitlán Del Valle Zapotec. These interpreta-

tions are all available for Shan bare nouns,[8] so this kind of analysis can account for the bare noun interpretations introduced in example (465), repeated below.

(465) Shan Bare Noun Interpretations

a. mă̌a hàw jù.
dog bark ipfv
'Dogs are barking.' *indefinite*

'The dog(s) is/are barking.' *definite*

b. mă̌a hàw.
dog bark
'Dogs bark.' *generic*

c. mă̌a wɔ̀t.wáaj.hă̌aj kwàa jâw.
dog disappear go finish
'Dogs are extinct.' *kind*

This analysis relates to the typology of definiteness marking because it offers some constraints as to which type-shifting operators are expected in which languages. The Blocking Principle (Chierchia 1998; Dayal 2004), defined in (507), constrains type-shifting by prohibiting the use of covert type-shifting operators that are duplicated by overt determiners in a language. Essentially, if a language has an overt determiner form of a type-shifter, then covert type-shifting using that operator is unavailable. For example, in English *the* is said to correspond to ι, which explains why bare nouns in English cannot type-shift using ι.

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

A previous account using type-shifting to capture the interpretations of bare nouns is from Despić (2019). This account specifically discusses anaphoric definiteness, claiming that whether

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

a language uses number marking is relevant for what interpretations are available for bare nouns. Table 4.6 summarizes the distribution. In a language without definite articles, ι can be used to generate a definite interpretation, as expected. However, only languages that mark number can use bare nouns to express anaphoric reference to a kind. This is because the singular count noun can be construed as ranging over the taxonomic domain as proposed by Dayal (2004).

Table 4.6: Languages without definite articles: Bare Nouns (Despić 2019: 28)

	+Number						**–Number**			
	Kind-level			**Object-level**			**Kind-level**		**Object-level**	
	Mass	Count		Mass	Count		Mass	Count	Mass	Count
		SG	PL		SG	PL				
Anaphoric	*	✓	*	✓	✓	✓	*	*	✓	✓
Type-Shift	$\cap$	ι	$\cap$	ι	ι	ι	$\cap$	$\cap$	ι	ι

Shan is a language without definite articles or overt plural marking. Therefore, Shan should pattern with the –Number category in Table 4.6. This is consistent with what has been discussed in Section 2.2.3 and Section 4.3. However, in order to discuss the full typology of definiteness marking, I will extend the table to include a distinction between unique and anaphoric definiteness. This will be discussed in Section 4.5.3.

4.5.2 Two types of definiteness

Jenks (2018) follows Schwarz (2009) in supporting the existence of two types of definiteness. To account for the obligatory use of the demonstrative in anaphoric definite environments in Mandarin, Jenks (2018) provides the denotations of unique and anaphoric definiteness shown in (508), where (508a) is the type-shifting operation ι and (508b) is the denotation of the Mandarin demonstrative. Since the English determiner *the* is used in both unique and anaphoric definite contexts, Jenks (2015) proposes that *the* is ambiguous between the unique and anaphoric definite meaning.

(508) a. UNIQUE DEFINITE ARTICLE: (Jenks 2018: (22))

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

b. ANAPHORIC DEFINITE ARTICLE: ι^x

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

According to Jenks (2018), a bare noun cannot express anaphoric definiteness due to a principle called *Index!*, defined in (509), which says that when an indexical expression is available, it must be used, meaning that a bare noun cannot be used to express anaphoric definiteness. The exception to this comes from Mandarin subject position, where both bare nouns and demonstratives can be used. Jenks claims that Mandarin subjects are topics and this topic marking negates the effect of *Index!*.

(509) *Index!* (Jenks 2018: (50))

Represent and bind all possible indices.

It is clear from the data in §4.3 that the Shan demonstrative is not an anaphoric definite determiner since it is not obligatory in all anaphoric contexts, as has been described for Thai. One option is to say that the pragmatic principle, *Index!*, introduced by Jenks (2018), can be overruled in Shan and other 'unmarked' languages in considerably more contexts than Jenks proposed. However, this seems unsatisfactory if we want the analysis to make strong predictions about what interpretations are available for bare nouns in a given language. Further, Dayal & Jiang (to appear) provide evidence demonstrating that Mandarin does allow bare nouns to express anaphoric definiteness. Instead, I propose that Shan has a null anaphoric type-shifter ι^x in addition to the ι type-shifter. These denotations would be the same as those in (508). This ambiguous type-shifting analysis would predict that the anaphoric interpretation of bare nouns should be available in contexts where anaphora is possible since ι^x has more presuppositions. *Maximize presupposition* (Heim 1991)[9] would predict that when ι and ι^x are in competition, ι^x should win out whenever

[9] *Maximize Presupposition* says to 'presuppose as much as possible'. Meaning, when choosing between competing

there is an indexical property available.

We might at this point ask what the difference is between the Thai and Shan demonstratives such that the Shan demonstrative would not count as a definite determiner but the Thai one would. Using the Consistency test (Dayal 2004, based on Löbner 1985) to distinguish between demonstratives and true definites does not offer an explanation.[10] For determiners like *the*, conjoining two clauses where a determiner-noun phrase is the argument of contradictory predicates, a contradiction results. Shan and Thai examples of the Consistency test with demonstratives are shown in (510) and (511), respectively.

(510) SHAN: CONSISTENCY TEST

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

kɔ́k hòj nâj pěn sǐ khǎaw. kɔ́k hòj nâj pěn sǐ lăm.
cup CLF.RND this be color white cup CLF.RND this be color black

'This cup is white. This cup is black.'

(511) THAI: CONSISTENCY TEST

dèk khon nán nɔɔn yùu tɛ̀ɛ dèk khon nán mâi.dâi nɔɔn yùu.
child CLF that sleep IPFV but child CLF that NEG sleep IPFV

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

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

Here, both demonstratives pattern like demonstratives in causing no contradiction when using deixis. However, both sound contradictory when uttered out of the blue, when the demonstrative-noun expression in each clause is interpreted anaphorically as referring to the same individual. Anaphoric uses of the demonstrative in English sound similarly contradictory. The Consistency test does not seem to resolve the question of the difference between Thai and Shan. However, the discussion in 4.5.4 will actually support the idea that Thai and Shan definiteness are not so different.

expressions where one has more presuppositions than the other, choose the one that has the most true presuppositions.

[10] See more discussion of issues related to the Consistency test in Moroney 2019b.

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

4.5.3 The full typology

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

- +/- Number represents whether a language in the typological category morphologically marks singular and plural.
- Object-level/Kind-level distinguishes between nominal expressions that refer to individual entities or to kinds/sub-kinds.
- Mass/Count distinguishes between mass and count nouns.
- For count nouns in the +Number category, singular (SG) and plural (PL) nouns are distinguished.
- Unique/Anaphoric (Anaph.) categorizes the type of definiteness.
- Type-Shift/Overt indicates whether that interpretation can come about through a type-shifting operator or an overt determiner.

The assumptions for filling in the table, based on Despić 2019 and Dayal 2004, are the following:

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

Both Marked

Table 4.7 gives the predictions for languages that mark both unique and anaphoric definiteness, like English. In these languages, the determiner is required to express both unique and anaphoric definiteness. Bare nouns, both mass and plural, can get a kind level interpretation, but bare nouns cannot be used to refer to kinds anaphorically. This predicts, for example, that plurals can have a unique, kind interpretation either with a bare plural or with a determiner.[11] The determiner relies on the taxonomic reading of the noun and the bare plural can type-shift with $\cap$.

Table 4.7: Definite bare noun interpretations, both marked, adapted from Despić 2019

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

In a language where both unique and anaphoric definiteness are marked differently, like German, the predictions would be identical to those found in Table 4.7 except that the determiner form would vary depending on whether unique or anaphoric definiteness was being expressed. For example, with weak definite articles, both singular and plural nouns can have a kind reference. This

[11] See Section 2.4 for discussion of $\cap$ blocking the overt use of the determiner for the kind interpretation in English.

is the *Unique-Overt* row in the *Kind-Count* columns in the above table. Schwarz (2009) confirms that this is the case for singular nouns in German (65: (66)) as well as in Fering, a Germanic language with a weak/strong definite determiner contrast that appears with both singular and plural nouns (66: (67)). We would also expect that anaphoric reference to a kind would be possible with the strong definite article and not with bare nouns.

One Marked

Table 4.8 shows the predictions for a language where only anaphoric definiteness is marked. In these languages, the determiner is required to express anaphoric definiteness. Bare nouns can get a kind level interpretation, but they cannot be used to refer to kinds or objects anaphorically.

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

As discussed in section 4.4.3, a classifier can be used as a marker of definiteness. Simpson & Biswas (2016) present data on Bangla showing that the classifier is obligatory with a noun in anaphoric reference, bridging contexts, and reference to salient visible entities.[12,13] Dayal (2014) notes that bare nouns can refer to kinds and that the plural classifier combined with a noun cannot be used to refer to a kind in Bangla except anaphorically. This is predicted with this typology if we treat the classifier as 'det.' in Table 4.8.

[12]There is an interfering sociolinguistic factor that prohibits a classifier when referring to a respected referent. See Simpson & Biswas 2016 for more details.

[13]According to Simpson & Biswas (2016), the classifier-noun form is used when the referent is 'identifiable'.

Table 4.8: Definite bare noun interpretations, one marked, adapted from Despić 2019

		+Number						**–Number**			
		Kind-level			**Object-level**			**Kind-level**		**Object-level**	
		Mass	Count		Mass	Count		Mass	Count	Mass	Count
			SG	PL		SG	PL				
Unique	*Type-Shift*	∩	ι	∩	ι	ι	ι	∩	∩	ι	ι
	Overt	-	-	-	-	-	-	-	-	-	-
Anaph.	*Type-Shift*	*∩	*ι	*∩	*ι	*ι	*ι	*∩	*∩	*ι	*ι
	Overt	det.	det.	det.	det.	det.	det.	det.	det.	det.	det.

Unmarked

Finally, in Table 4.9 are the predictions for languages like Shan and Serbian that do not obligatorily mark either unique or anaphoric definiteness. In these languages, the determiner is not required to express either type of definiteness. Bare nouns can get a kind level interpretation, and in number marking languages, singular bare nouns can be used to refer to kinds anaphorically.

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

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

Table 4.9: Definite bare noun interpretations, unmarked adapted from Despić 2019

		+Number						**–Number**			
		Kind-level			**Object-level**			**Kind-level**		**Object-level**	
		Mass	Count		Mass	Count		Mass	Count	Mass	Count
			SG	PL		SG	PL				
Unique	*Type-Shift*	∩	ι	∩	ι	ι	ι	∩	∩	ι	ι
	Overt	-	-	-	-	-	-	-	-	-	-
Anaph.	*Type-Shift*	*∩	ι	*∩	ι	ι	ι	*∩	*∩	ι	ι
	Overt	-	-	-	-	-	-	-	-	-	-

4.5.4 A pragmatic proposal

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

Ahn (2019) proposes that Korean and Thai have the following options for marking anaphoric definiteness, shown in (512) and (513), respectively.[14] Each language has a ranked scale of anaphoric expressions, where the more semantically complex expression will only be used in case the less complex expression cannot uniquely identify the referent. Since Korean does not lexicalize a simplex pronoun, bare nouns are the simplest way to express anaphora. In Thai, there is a simplex lexical pronoun which is preferred to the bare noun in contexts where there is only one available antecedent.

[14] Ahn (2019) departs from Schwarz (2009) in separating the anaphoric index from the lexical expression. I am not adopting this assumption, but this choice is not essential to the analysis.

(512) Korean: ⟨N, *ku* N⟩ (Ahn 2019: (83))

a. $[\![\text{N}]\!] = \iota x : \text{entity}(x) \wedge \phi(x) \wedge P(x)$

b. $[\![\textit{ku}\ \text{N}]\!] = \iota x : \text{entity}(x) \wedge \phi(x) \wedge P(x) \wedge R(x)$

(513) Thai: ⟨pronoun, N, N *nán*⟩[15] (Ahn 2019: (84))

a. $[\![\text{pronoun}]\!] = \iota x : \text{entity}(x) \wedge \phi(x)$

b. $[\![\text{N}]\!] = \iota x : \text{entity}(x) \wedge \phi(x) \wedge P(x)$

c. $[\![\text{N}\ \textit{nán}]\!] = \iota x : \text{entity}(x) \wedge \phi(x) \wedge P(x) \wedge R(x)$

According to Ahn (2019), the effect of non-uniqueness inferred for demonstrative-noun phrases comes as a result of this ranking of anaphoric expressions.[16] If the bare noun is not used, that indicates that the bare noun phrase cannot uniquely identify the intended referent, so non-uniqueness can be inferred.

This competition-based account predicts that if the type-shifted (anaphoric) bare noun denotation is a subset of the denotation of a demonstrative-noun phrase, we would expect to find bare nouns preferred to the demonstrative-noun phrase in some contexts. When there are enough available referents in the discourse context that the pronoun cannot uniquely identify the intended referent, we would expect the bare noun to be preferred to the demonstrative-noun phrase. I am assuming that bare nouns are both less morphologically and semantically complex than more marked definite expressions.[17]

This type of analysis seems to work for Thai. Here are two examples with donkey anaphora which mirror other examples used in this chapter. (514) is the analogue of (496), and (515) has a very similar meaning to (492).[18] The donkey anaphors in these examples are bare nouns 'bucket'

[15] In Thai, the word order for demonstrative-noun phrases is N *nán*, so I have used that word order here.

[16] An alternative approach to demonstratives by Dayal & Jiang (to appear) includes a presupposition of non-uniqueness within a widened domain.

[17] The relative importance of semantic and morphological complexity in this competition-based account will not be discussed here.

[18] These examples were translated into Thai from English by one native Thai speaker, and then judged as grammatical by a second native speaker.

and ‘ball’ in (514) and ‘rat’ in (515). With these two available referents, bare noun anaphora is possible in Thai.

(514) Thai: Donkey Anaphora

lûuk-chaai wai nùŋ khùap khɔ̌ŋ chăn baaŋ kràŋ kɔ̂ tham nai sìŋ thî mâi khâatkhít.
child-male aged one year poss 1 some time prt do in thing that neg expect

nai thúk thúk khráŋ thî khăw hěn thăŋ-nám lɛ lûukbɔn khăw mák ca waaŋ
in every every time that 3 see bucket-water and ball 3 often irr put

thăŋ-nám bon lûukbɔn.
bucket-water on ball

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

(515) Thai: Donkey Anaphora

mɛw lɛ nǔ mâi khɔ̂ɔi ca thùuk kan. mûa-rài thî mɛw hěn nǔ, man mák ca
cat and rat neg quite irr correct together when that cat see rat, 3 often irr

lâi-càp nǔ
chase-catch rat

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

This account has more difficulty explaining the pattern of anaphora in Mandarin. This can be seen in the following two examples, repeated from (472a)–(472b), above. Here, we might expect to be unable to use a pronoun because *nansheng* ‘boy’ and *nüsheng* ‘girl’ would be referred back to with the same pronoun.[19] In this case, we might expect the bare noun to refer back to the boy. However, it seems like the demonstrative-noun phrase must be used.

(472a) Jiaoshi li zuo-zhe yi ge nansheng he yi ge nüsheng,
classroom inside sit-prog one clf boy and one clf girl
‘There are a boy and a girl sitting in the classroom...’

[19]This is true in spoken language rather than written. See Ahn 2019 for more discussion.

(472b) Wo zuotian yudao #(na ge) nansheng
I yesterday meet #(that CLF) boy
'I met the boy yesterday.' (Mandarin, Jenks 2018: (15a,b))

If the demonstrative in a language has been grammaticalized to function as an anaphoric definite determiner, we might expect the Blocking Principle to disallow anaphoric definite type-shifting in the language. Therefore, a bare noun cannot be used in this sort of anaphoric context. Another possible explanation is that word order has an effect on the available bare noun interpretations with the result that the bare noun is not a potential anaphoric competitor in this syntactic context. It is well-known that there is an effect of sentence structure, where nominal expressions in the post-verbal position are typically indefinite (Li & Bisang 2012). Additionally, more recent work has suggested that the subject-object asymmetry in anaphoric bare nouns in Mandarin is not so robust. Ahn (2019) notes some variability in judgments from five Mandarin speakers looking at (472a)–(472b). Dayal & Jiang (to appear) give detailed evidence supporting the idea that non-subject bare nouns can be anaphoric in Mandarin.[20]

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

This section has shown that a competition based approach to definite expressions can explain some of the variation found both within and across languages. This analysis uses *Don't Overdeterminate!* from Ahn (2019) to choose which definite expression is appropriate in which context. In contexts where a pronoun fails to uniquely identify the referent, a bare noun can express anaphoric definiteness, as the examples of donkey anaphora in Thai showed. This supports the idea that Thai is a language that does not obligatorily mark anaphoric definiteness with an overt morpheme, but

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

instead has a bare noun as an option for anaphoric definiteness. In a similar way, this account can explain why certain anaphoric expressions might be preferred in specific contexts in Shan.

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

4.5.5 Summary

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

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

4.6 Discussion

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

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

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

(516) i. The role of contrast

ii. The role of relative sentential prominence

iii. The role of disambiguation

This section offers a brief discussion of contrast (§4.6.1) and sentential prominence (§4.6.2) and some discussion of the role of ambiguity in definiteness marking (§4.6.3).

4.6.1 Contrast

Discussions of the role of contrast can be difficult since 'contrast' can refer to different things. As has been discussed in §4.3.2, for Ahn (2019) contrast is connected to the number of salient entities. Ahn (2019) claims that for ASL, for example, having multiple animate salient entities in a given contexts allows the bare noun to be used. For this reason, bare nouns are more common for languages like Shan, Serbian, and Thai—where a pronoun might be competing with a bare noun to express anaphora—in contexts where there is more than one individual in the discourse context.

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

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

4.6.2 Sentential prominence

Simpson et al. (2011) noted that the classifier-noun phrase can indicate relative prominence of a referent within a sentence. Discourse old referents might be more likely to be bare nouns rather than classifier-noun phrases in languages like Vietnamese. Jenks (2015) noted a similar pattern for

Thai, where in a long narrative, bare nouns can be used anaphorically, saying this happens 'after an individual has been established and it is clear that they are the only individual of the relevant type' (113, fn. 7).

Bare noun use in both of these cases seems compatible with what Ahn (2019) said about anaphoric bare nouns: they indicate that the property of being that noun is sufficient to uniquely identify the referent. However, sentential prominence is also connected to information structural properties such as Givenness and Focus as discussed by Krifka & Musan (2012), among others. It would be worthwhile to investigate the connection between information structure and bare noun definiteness.

4.6.3 Ambiguity

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

Perhaps what *Index!* from Jenks (2018) represents is the preference for a language to unambiguously specify an intended meaning. Since bare nouns can have a variety of interpretations, using a demonstrative-noun phrase can demonstrate clearly that a definite interpretation is intended. We would expect this to come into play whenever an expression is potentially ambiguous. This ambiguity might more often arise with bare nouns or with *pro*, both of which can have several interpretations.[21] Ambiguity could even arise with expressions like pronouns or other definite expressions.

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

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

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

In a discussion of types of ambiguous expressions, Sennet (2016) included both flexible types (i.e., from the possibility of type-shifting) and the generic versus episodic readings of sentences, which is connected to the contrast between kinds and individuals. Thus, bare noun expressions have already been identified as being potentially ambiguous. An example of this is shown in (517). Under Chierchia's (1998) analysis of the bare plural *dinosaurs*, the base denotation of the plural is a kind and the possible episodic and generic interpretations come from existential closure (as a result of Chierchia's (1998) Derived Kind Predication) and from a generic operator quantifying over instances of the kind, respectively.

(517) Dinosaurs ate kelp. (Sennet 2016: (31))

Returning to Shan, the competition between the generic and definite readings are apparent from the contrast between (518a) and (518b) as possible follow up sentences to (518).[22] (518) comes

[22]The complementizer *waa* here is distinct from the one that appears with relative clauses. See Jenks (2011) for a discussion of the cognate forms found in Thai.

as part of a longer story where two friends are discussing possible pets that Jai Kham might have bought at the pet store. The addressee of this sentence knows that Jai Kham was interested in buying a particular black cat and white dog. Both (518a) and (518b) are judged as good follow ups to sentence (518). In (518a), the bare noun *méw* 'cat' is interpreted as the aforementioned black cat, but in (518b), the bare noun *mǎa* 'dog' is most easily interpreted as referring to dogs in general.[23] Both the definite and generic readings are possible for both sentences, but one reading is more salient than the other. Given that there is already one salient dog and cat in the context, the bare noun should be the most appropriate definite expression given the economy principle. However, some speakers do prefer a demonstrative-noun phrase in these sentences. I argue that this stems from the possible ambiguity between a definite and generic interpretation.

(518) SHAN: AMBIGUITY AND ANAPHORA

Tsáaj Khám lɤk sûː méw lăm tŏ nân, kójkaa ʔàm laj lɤk
Jai Kham choose buy cat black CLF.ANML that but NEG get choose

mǎa khǎaw tŏ nân.
dog white CLF.ANML that

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

a. jɔ̂n waa méw láklɛ̌m nàa.
because COMP cat clever very
'because the cat is very smart.'

b. jɔ̂n waa mǎa ʔàm láklɛ̌m.
because COMP dog NEG clever
'because dogs aren't smart.'

Table 4.10 shows potential competing expressions (on the horizontal axis) and interpretations (on the vertical axis) for a subset of nominal expressions in Shan. For example, the [N dem.] expression indicates anaphoric definiteness, but it can be anaphoric to a kind, a singular entity, or a plural entity.

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

[23] Bare nouns can be definite in sentences with negation in Shan.

Table 4.10: Anaphoric and kind interpretations, potential ambiguity in Shan

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

4.7 Chapter summary: Definiteness marking

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

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

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

This chapter also discussed other factors that influence the choice of definiteness expression and proposed that ambiguity is a significant factor. If a nominal expression can have different interpretations in different contexts, use of that nominal expression could lead to ambiguity and failure to uniquely identify an intended referent. I proposed, therefore, that in potentially ambiguous contexts, a semantically more complex definite expression might be chosen. Taking the role of ambiguity into account could explain some of the patterns of data that the economy principle cannot account for.

CHAPTER 5

LEFT-LOCATED CONSTRUCTIONS

DEFINITENESS AND QUANTIFICATION: EVIDENCE FROM SHAN

Mary Rose Moroney, Ph.D.

Cornell University 2021

This chapter investigates left-located constructions in Shan. These constructions are either definite or quantificational, and serve to restrict the context in which the matrix clause is evaluated. Using several syntactic diagnostics, I demonstrate that the definite left-located constructions are topicalized to that position in some cases and left-dislocated in others. There is a construction found in definite left-located contexts that looks like an internally headed relative clause (IHRC). These IHRCs obligatorily appear in left-located positions unlike other relative clauses in Shan. Finally, this chapter describes the function of a particle that marks universal quantification over situations and analyzes donkey anaphora in Shan.

5.1 Introduction

The previous chapters have built up a syntax and semantics of Shan nominal expressions. This chapter takes structures and interpretations that have been developed over the previous chapters and applies them to examples of left-located constructions that are found in Shan.

There is a construction in Shan that looks like an internally headed relative clause (IHRC).[1] This construction is always in the left-most position of a clause. It looks identical to an externally headed relative clause, except for the position of the head. These constructions tend to be judged as more grammatical when the head of the relative is not the subject of the relative clause. I argue that these are, in fact, internally headed relative clauses based on the fact that there is a clear internal head that is co-referenced with an argument of the matrix clause in a way that indicates that the clause refers to the head.

Left-located (LL) constructions in Shan are either topicalized, leaving a gap in the matrix clause, or left-dislocated, leaving a pronoun. There are both topicalized and left-dislocated IHRC constructions. Shan IHRCs and other LL constructions function as topics to limit the situation described by the sentence, and I provide a semantics of both the IHRC construction and LL constructions to capture this. A left-located constituent that appears lower in the syntax than topicalizations and left-dislocations is associated with a sentence particle *kɔ*. This syntactic position is connected to universal quantification over situations described by the sentence.

The structure of this chapter is as follows: To finish this introduction §5.1.1 discusses other types of left-located (LL) constructions. Then, §5.1.2 introduces the apparent IHRC construction. §5.1.3 briefly looks at left-located constructions in Thai. Then, section 5.2 examines several syntactic features of left-located constructions in Shan in order to better characterize them. It looks as though the left-located construction is topicalized to a left-position when there is a gap in the ma-

[1] Another variety of Shan (Yŭn Shan) seems to have IHRCs more productively. See Moroney (2019c) for more details. The data from this dissertation comes from Southern Shan, spoken in the southern part of Shan State, Myanmar.

trix clause since an overt pronoun is required when there is a syntactic island. Thus, overt pronouns might be evidence of left-dislocations. Section 5.3 offers a semantic analysis of the LL constructions and IHRC constructions. Section 5.4 discusses the syntax and semantics of the particle *kɔ* and analyzes examples of donkey anaphora. Section 5.5 summarizes this chapter.

5.1.1 Left-located (LL) constituents

Several kinds of nominal and prepositional phrases can be found in a left-located (LL) position of a clause. Left-located constituents in Shan can be names (519), anaphoric expressions (520), prepositional phrases (521), specificational constructions (522), noun complement clauses (523), cleft constructions (530), and apparent IHRCs (527)–(534). These constructions can be identified by the demonstrative *nâj* 'this' or *nân* 'that', which appears after the left-located constituent.[2]

In both (519) and (520), the LL constituent refers to a familiar individual or group of individuals, and a pronoun in the main clause co-references the LL constituent.[3]

(519) [jíŋ lăawŋúːn nâj]$_i^{DP}$ háw jŭːnjăn laj waa mán$_i$ tě ʔàm máa tinaj
Ying Lao_Nguen this 1 confirm ACH that$_v$ 3 IRR NEG come here
'Ying Lao Nguen$_i$, I confirm that she$_i$ will not come here.'

(NAME)

(520) [khúsɔ̆n săam kɔ̂ nân]$_i^{DP}$ tsɤ́-khăw$_i$ ʔɔ̆nkăn thátsaaŋ kăanwaan …
teacher three CLF.HUM that CLF.PL-3 together examine homework
'The three teachers examined homework …' (ANAPHORIC DP)

[2]One of these demonstratives appears with nearly every example of a left-located construction, even left-located prepositional phrases. When there is no demonstrative separating the left-located constituent from the rest of the sentence, there might be a slightly longer pause after the LL constituent. An example without the demonstrative is in (1).

(1) tsáajkham, náaŋʔɔ̀n, lɛ náaŋsĕŋ, tsɤ́-khăw jù ti hâan-khăaj-tŏsát
Jai_Kham Nang_Orn and Nang_Saeng CLF.PL-3.PL stay at story-sell-animal
'Jai Kham, Nang Orn, and Nang Saeng, they are at the pet store.'

[3]It is clear that *khúsɔ̆n săam kɔ̂ nân* 'those three teachers' is left-located and a separate constituent from *tsɤ́-khăw* 'they' in (520) because the pronoun would appear to the left of the quantificational material, if it formed a constituent with it.

(521) includes a left-located prepositional phrase, where a member of the group described by the PP is referred to in the main clause—here with either a name or a numeral+classifier.[4,5]

(521) [náj phujíŋ sǎam kô nân]$_i^{PP}$ tsáaj lǎawkhám tsaaŋ kwàa hǎa [náaŋ lǎawŋúun]$_{k<i}$/
in women three CLF.HUM that Jai Lao_Kham can go find Nang Lao_Nguen/

[kô nɯɯŋ]$_{k<i}$
CLF.HUM one

'Among [those three women]$_i$, Jai Lao Kham could find Nang Lao Nguen$_{k<i}$/[one of them]$_{k<i}$.' (PP)

(522) and (523) are examples of LL specificational constructions. (522) has the form of a question (indicated by superscript *Q*), followed by the particle *nǎj* that is also used at the end of if-clauses and the demonstrative *nâj*. The clause following specifies the answer to the question.

(522) [[khǎw len tsŷŋhŭ̌ŭ]^Q nǎj nâj] khǎw ʔàm tsâɯ tǐn kójkaa khǎw tsâɯ mú́ hě len
3.PL play how PRT.NAJ this 3.PL NEG use foot but 3.PL use hand and play

ʔɔ̌nkǎn ʔɤ
together ASSERT

'The way they play is they don't use their feet but they use their hands and play together.'

(QUESTION: SPECIFICATIONAL)

(523) has the form of a noun complement clause for the noun *lɔŋtáaŋ* 'reason', and the following adjunct clause specifies the reason. LL constituents followed by *jốn waa* like this can have the copula verb *pěn* between the LL constituent and *jốn waa*, but the only examples of these include IHRC LL constructions. I have not specifically tested alternating *pěn* in these constructions.

(523) [lɔŋtáaŋ ʔǎn tsáaj lǎawkhám ʔàm lajtsǎj mɛ́w nân]^{NCC} jốn waa kŏ mɛ́w
reason COMP Jai Lao_Kham NEG like cat that because COMP$_v$ fear cat
'The reason that Jai Lao Kham doesn't like cats is because (he) fears cats.'

(NCC: SPECIFICATIONAL)

[4]The index notation $k < i$ just indicates that the nominal with index k is a sub-part of the nominal with index i.
[5]See section 3.3.2 for a discussion of NP-ellipsis associated with this example.

There is also a different sort of construction which appears to involve movement to a leftward position. This construction is associated with quantification or modality in some way. These are examples like (524) and (525), which involves a modal *laj* 'can'.[6] There are also examples like (526) where the left-located construction contains a question word (*kaahʉ̌ʉ* 'how much' in this example).

(524) [[náj tɯɯk nâj]$_{PP}$ háw mɛ́p]$_{VP^{?}}$ ʔàm laj
in box this 1 hide NEG can
'We can't hide inside this box.' (MODAL)

(525) [lɤ̀k táŋ sì ʔǎn] kɔ laj
choose all four CLF.GN PRT can
'You can choose all four (pictures).' (MODAL)

(526) [lò pàj thɛŋ hʉ̌ʉŋ kaahʉ̌ʉ] kɔ ʔàm hû
must wait additionally long.time how.much PRT NEG know
How much more I need to wait, I don't know. (EMBEDDED Q)

I am going to categorize this type of construction as distinct from the nominal left-located constructions, since the left-located constituent does not seem to be a nominal expression. Note that in (524), the left-located prepositional phrase *náj tɯɯk nâj* 'in the box' appears to the left of the fronted T/VP *háw mɛ́p* 'we hide'. I will come back to these constructions later.

In (527), is an example of the constructoin that looks like an internally headed relative clause. The left-located constituent is *ʔǎn phujíŋ ʔǎw táaŋkǐn lêŋ mɛ́w nân* 'the food that the woman fed to the cats'. Here, the first word *ʔǎn* is the complementizer found with externally headed relative clauses. The last word, *nân* is a demonstrative.[7] This left-located constituent is the subject of a question that asks what it is. In this case, the question is asking about what *táaŋkǐn* 'food' is being fed to the cats, so if this is a relative clause, the head is internal to the clause.

[6]Note that (525) has an example of the generic classifier use of *ʔǎn*, which is separate from the relative clause complementizer use.

[7]The proximal demonstrative *nâj* can be used here as well if the clause head is 'close' to the speaker.

(527) [ʔăn phujíŋ ʔăw táaŋkǐn$_i$ lêŋ méw nân]$^{IHRC}_i$ pěn săŋ
COMP woman take food feed cat that COP what
'The food that the women fed to the cats is what?' (IHRC?)

The externally headed version of (527) can be seen in (528).

(528) [táaŋkǐn$_i$ ʔăn phujíŋ ʔăw păn méw nân] pěn săŋ
food COMP woman take give cat that COP what
'The food that the women gave to the cats is what?' (EHRC)

These examples all look like cleft constructions like (529) in Japanese.

(529) [Taroo-ga e$_i$ mita-no] -wa nan(i)$_i$-na -no
Taro-NOM saw-NM -TOP what-COP -Q
'What is it that Taro saw?' (Japanese, Kizu 2005: 41, (60c))

Adjunct questions often take the form where what would be the main clause in English is a left-located, nominalized clause that is followed a predicated adjunct question. These are shown in (530) and (531). These examples also look like the IHRC construction in (527). It seems likely that the examples in (527)–(531) are cleft constructions.

(530) [ʔăn phujíŋ ʔăw táaŋkǐn lêŋ méw nân]^{event} pěn tsŷŋhŭ
COMP woman take food feed cat that COP how
'How is the women feeding food to the cats?'

(ADJUNCT Q)

(531) [ʔăn phujíŋ ʔăw táaŋkǐn lêŋ méw nân]^{event} pěn ti-lăj
COMP woman take food feed cat that COP place-which
'Where is it that the women fed food to the cats?'

(ADJUNCT Q)

(532)–(533) demonstrate the form of an embedded question and the form of a free relative, respectively. (532) is an answer to the question 'Do you know what Jai Lao Kham bought at the store yesterday?' The embedded question has the form of a regular question, with in situ question

word *sǎŋ* 'what', embedded under the verbal complementizer *waa*. The question word stays in-situ. (533) is a follow up to (532), elaborating on the information that Ying Lao Nguen knows. The free relative clause (FRC), *ʔǎn tsáaj lǎawkhám sûː nân* '(the thing) that Jai Lao Kham bought', is identical to an externally headed relative or the apparent internally headed relative, except that there is no overt head.

(532) jíŋ lǎawŋúːn tɔ̀ːp waa háw hû waa tsáaj lǎawkhám sûː sǎŋ kójkaa
Ying Lao_Nguen reply COMP$_v$ 1 know COMP$_v$ Jai Lao_Kham buy what but

háw ʔàm hû waa tsáaj lǎawkhám khǎaj sǎŋ
1 NEG know COMP$_v$ Jai Lao_Kham sell what

'Ying Lao Nguen says, 'I know what Jai Lao Kham bought but I don't know what Jai Lao Kham sold.'' (EMBEDDED Q)

(533) [ʔǎn tsáaj lǎawkhám sûː nân] tê pěn sɤ̌
COMP Jai Lao_Kham buy that CONTR COP shirt
'What Mr. Lao Kham bought was a shirt.' (FRC)

5.1.2 Shan IHRCs

As was introduced in this previous section, there is a construction in Shan (Southwestern Tai) that looks like an internally-headed relative clause (IHRC)—that is, a relative clause where the noun that is being described by the relative is inside the clause itself. Examples (534)–(535) are nominalized clauses where an argument of the nominalized clause is also functioning as an argument of the matrix clause.

In (534), the subject of the clause is inside of the [ ] brackets, and the apparent internal head is in blue text. Here, this constituent is the subject of a predicate meaning 'received a prize'. The head here is *pɔŋkwáam* 'article'. The clause internal head is actually the external head of a relative clause *pɔŋkwáam ʔǎn mí tsɯ waa mɤ́ŋ táj* 'the article that is called 'Shan State'.

(534) [ʔǎn jíŋ lǎawŋúun tɛm [pɔŋkwáam$_i$ ʔǎn mí tsɯ waa mɤ́ŋ táj] nân]$^{IHRC}_i$
COMP Ying Lao_Nguen write article COMP have name COMP$_V$ state Shan that

laj hâp sú-láŋwán
get receive prize

'The article that Ying Lao Nguen wrote that is called "Shan State" won a prize.' (IHRC?)

In (535), the left-located construction is the subject of a question about the *mɛ́w* 'cats' in the left-located clause. This example is different from the previous ones because the head of the left-located clause is co-referenced with an overt pronoun *tsɤ́-khǎw* 'they' in the matrix clause. In (527) and (534), there is no overt pronoun in the matrix clause cross-referencing the head of the nominalized clause. In examples like (527) and (534), it would be useful to know whether these constructions are in the subject position, whether they have moved and left a gap, or whether there is a null pronoun left in the subject position. This will be addressed in section 5.2.

(535) [ʔǎn jíng lǎawŋúun ʔǎw táaŋkǐn lêŋ mɛ́w tsɤ́ nân$_i$]^{IHRC} tsɤ́-khǎw$_i$ mí
COMP Ying Lao_Nguen take food feed cat CLF.PL that CLF.PL-3 have

lǎaj tŏ
how.many CLF.ANML

'The cats that Ying Lao Nguen feeds, how many are there?'

These IHRC constructions can only appear in a left-located position. (536) gives an intended IHRC in the object position of the matrix clause. An externally headed relative clause (EHRC) is fine in that position, as (537) shows.

(536) *náaŋ ʔɔ̀n kǐn [ʔǎn tsáaj khám pɔ̀k màakmoŋ]
Nang Orn eat COP Jai Kham peel mango
intended: 'Nang Orn ate the mangoes that Jai Kham peeled.' (IHRC)

(537) náaŋ ʔɔ̀n sɔ́j [màakmoŋ sì hòj ʔǎn tsáaj khám pɔ̀k mɤwáa nân]
Nang Orn cut mango four CLF.RND COMP Jai Kham peel yesterday that
'Nang Orn cut the four mangoes that Jai Kham peeled yesterday.' (EHRC)

(538) and (539) give examples of headless relatives. These look like externally headed and internally headed relatives except that there is no overt head. In (538), there is no overt head in the relative clause *ʔăn tsáaj lăawkhám pɔ̀k wâj nân* 'what Jao Lao Kham peeled'. The head is understood to be the fruit (*màakmâj*) that Jai Lao Kham peeled. Like externally headed relatives and unlike internally headed relatives, the headless relative can appear in the argument position, as in (539). Here, the headless relative is the object of *kǐn* 'eat' and it appears after that verb.

(538) tsáaj lăawkhám pɔ̀k màakmâj jù. [ʔăn tsáaj lăawkhám pɔ̀k wâj nân]
Jai Lao_Kham peel fruit IPFV COMP Jai Lao_Kham peel stay that

jíŋ lăawŋúːn ʔăw sɔ́j
Ying Lao_Nguen take cut

'Jai Lao Kham is peeling fruit. What Jai Lao Kham peeled, Ying Lao Nguen cut (pieces of fruit).'

(539) nŭ nâj méw tîɲɔ́p mán jù hím hɤ́n jíŋ lăawŋúːn.
mouse this cat catch 3 stay near house Ying Lao_Nguen

méw lajtsăj kǐn [ʔăn ɲɔ́p máa laj nân]
cat like eat COMP catch come ACH that

'Mice, cats catch them near Ying Lao Nguen's house. Cats like to eat the ones they catch.'

When IHRCs are in a left-located position, they are acceptable. It is even possible for the co-referential argument in the matrix clause to have a partitive interpretation of the left-located clause's internal head. This is shown in (540). Here, the internal head seems to be *màakmoŋ sì hòj* 'four mangoes' and the matrix clause verb applies to *sɔ̌ŋ hòj* 'two' of those mangoes.

(540) [ʔăn tsáaj khám pɔ̀k wâj màakmoŋ sì hòj mɤwáa nân] náaŋ ʔɔn ʔăw sɔ́j.
COP Jai Kham peel keep mango four CLF.RND yesterday that Nang Orn take cut

sɔ̌ŋ hòj jâw mɤnâj
two CLF.RND PFV today

'The four mangoes Jai Kham peeled yesterday, Nang Orn cut two today.'

These constructions look similar to internally-headed relative clauses (IHRCs) in other languages. (541) gives an example from Navajo, and (542) gives an example from Japanese.[8]

(541) [Mary 'aghaa' 'aɬníí'dóó yizdiz]-ę́ę́ Alice yiyíiɬchíí'.
Mary wool half 3OBJ.3SBJ.spin.PFV-IGII Alice 3OBJ.3SBJ.dye.red.PFV
'Mary spun half of the wool. Alice dyed it.'

(Navajo; Bogal-Allbritten & Moulton 2017: (3))

(542) John-wa [[Mary-ga san-ko-no ringo-o muitekureta] -no]-o tabeta.
John-TOP Mary-SUBJ three-CLF-GEN apple-ACC peeled NO-ACC ate
'Mary peeled three apples and John ate them all.'

(Shimoyama 1999, citing Hoshi 1995)

They also appear somewhat similar to Hindi correlatives; compare (543) to (544). In (543), the subject of the matrix clause is *kónkhăajnɤ̂ kɔ̂ nân* 'that meat-seller', which seems to be the head of the left-located clause. In (544), the subject of the matrix clause is *laRke=ko* 'the boy', the head of the correlative.

CONTEXT FOR (543): Four men love Nang Lao Nguen. She loves two of them, a meat-seller and a teacher. The other two are a meat-seller and a doctor. Someone says, 'She should marry the doctor because he is rich.' Another person responds with:

(543) [ʔăn náaŋ lăawŋʉ́n hâk kónkhăajnɤ̂ kɔ̂ nân] kɔ pěn kónmí mɤ̌nkăn
COMP Nang Lao_Nguen love meat-seller CLF.HUM that ALSO COP wealthy_person same
'The meat-seller that Nang Lao Nguen loves is also a rich man.'

(544) [$_{IP}$ [$_{CP}$ jis laRke=ko Sita pyaar karti hai] voh bahut ameer hai]
REL.OBL.SG boy=ACC Sita love do.HAB.F be.PRS DEM very rich be.PRS
'The boy who Sita loves is very rich.' (Bhatt 1997: p. 56 (7a))

A issue with analyzing these constructions in Shan as IHRCs is that there seems to be a subject/non-subject asymmetry when it comes to the acceptability of the IHRC construction. If

[8]The Navajo IHRCs all appear with OSV word order instead of the more common SOV order in Bogal-Allbritten & Moulton 2017. They suggest that this is due to the length of the IHRC nominal.

the head of the relative clause is the subject of the relative clause, it is judged less grammatical than if it is not the subject.[9] This can be seen in the contrast between the EHRC in (545) and the IHRC in (546). In (545), the EHRC head is the subject of the relative clause. The IHRC equivalent in (546) is not very grammatical, but the free relative (FRC) version is, as shown in (547). In comparison, the internal heads in (534)–(535) are the objects of the relative clauses. The Hindi correlative equivalent of (546) in (548) is acceptable.

(545) phujíŋ [ʔăn tsúk jù nân]$_i$, tsɤ́-khăw$_i$ sŭŋ nàa.
woman COMP stand stay that PL-3.PL tall very
'The women that are standing, they are tall.'

(546) ???[ʔăn phujíŋ tsúk jù nân]$_i$, tsɤ́-khăw$_i$ sŭŋ nàa.
COMP woman stand stay that PL-3.PL tall very
'The women that are standing, they are tall.'

Consultant comment: sounds better with *phujíŋ ʔăn* (EHRC version)

(547) [ʔăn tsúk jù nân]$_i$, tsɤ́-khăw$_i$ sŭŋ nàa.
COMP stand stay that PL-3.PL tall very
'The ones that are standing, they are tall.'

(548) [jo laRkiyaaN khaRii haiN] ve lambii haiN
which girls standing be-PR they tall be-PR
'The girls who are standing are tall.' (Hindi; Dayal 1996: p.12, (20a))

(549)–(550) give examples of headless relatives which can easily relativize the subject of the relative clause. In both cases, the classifier for humans (*kɔ̂*) appears at the beginning of the clause. As discussed in chapter 3, a classifier can precede a relative clause. In (549) there is a pronoun cross-referencing the head of the free relative, and in (550) there is not.

(549) [kɔ̂ ʔăn tsúk jù nân]$_i$ mán$_i$ sŭŋ nàa
CLF.HUM COMP stand stay that 3 tall very
'The one that is standing, she is tall.'

[9]It is possible for the head of the IHRC to be the subject, object, or have some other role in the matrix clause.

(550) [kɔ̂ ʔǎn tsúk jù nân] sŭŋ nàa
CLF.HUM COMP stand stay that tall very
'The one that is standing is tall.'

This subject/object asymmetry could stem from the reverse subject/object asymmetry for externally headed relatives, wherein the object relatives are more difficult to process than subject relatives, as predicted by the Accessibility Hierarchy (Keenan & Comrie 1977). This has been confirmed in studies of second language acquisition, such as Doughty 1991. Since the externally headed relative clauses involving subjects are very good, these constructions are preferred to the internally headed versions. Externally headed relatives involving objects are not as good as subject relatives, so by comparison the internally headed versions are that much more acceptable.

Another issue is that I have not found much evidence of these in naturalistic data. The IHRC construction is only accepted when it is in a left-located position with respect to the matrix clause. The reminder of this chapter focuses on left-located constructions, which includes these IHRC constructions.

5.1.3 Thai

It is somewhat surprising that these IHRC constructions are acceptable at all in Shan, when they are not possible in related languages like Thai. Shan is an SVO word order language (as is Thai), and these languages do not frequently have IHRCs. At one point, it had been observed that only OV languages have IHRCs (e.g., Kuroda 1974; Gorbet 1974; Langdon 1977; Downing 1978; Cole 1987). There are, however, some SVO languages with IHRCs, such as, Gur languages (Hiraiwa 2005; Tellier 1989), Riau Indonesian (Gil 2000), Tagalog (Aldridge 2017), and ASL (Wilbur 2017). Additionally, IHRCs have been reported in at least some dialects of Karen (Tibeto-Burman), another SVO language spoken in Karen State, Myanmar (Shee 2008).[10]

[10] Although my brief look at S'gaw Karen did not find any IHRCs.

(551)–(552) give examples of the closest constructions that Thai has to Shan left-located IHRCs. These examples use the cognate form of the Shan complementizer ʔ*ăn*, which is ʔ*an*. In (551), the left-located constituent ʔ*an kʰwa:mkaruna:pra:ni:* is nominal, but not clausal like it can be in Shan. The same is true of ʔ*an pʰi:* ʔ*uən kʰɔŋ raw nan* in (552). There is no verbal element in these left-located constituents. In (552), the left-located element does begin with complementizer ʔ*an* and end with demonstrative *nan* 'that', so there are some similarities with the Shan construction. The examples of ʔ*an* here seem to be the 'topic' use identified by Kitsombat (1981), where the morpheme appears before a nominal expression at the beginning of a sentence.

(551) ʔan kʰwa:mkaruna:pra:ni:, ca mi: kʰraj baŋkʰap kɔ ha: maj
ʔan mercy IRR have someone force PRT find NEG
'For the mercy, it is not something that is coerced.'

(552) ʔan pʰi: ʔuən kʰɔŋ raw nan pen supʰa:pburut tʰi: nɔ:ŋ-nɔ:ŋ sa:w-sa:w
ʔan older.sib fat POSS 1 that COP gentleman COMP younger.sib-RED woman-RED
so:t-so:t naj opfit kɔ ma:jpɔ:ŋ kan tʰaŋnan
single-RED in office PRT target together all
'For our fat guy, he is someone that young and single ladies in the office all have their eyes on.'

The externally headed clause in (553) does not have a grammatical internally headed counterpart in Thai, as (554) shows.[11]

(553) [Naŋsŭŭ thîi raw hěn níi] man dɛɛŋ.
book COMP 1 see this 3.SG red
'This book that I see is red.' (Thai)

(554) *[ʔan naŋsŭŭ raw hěn níi] man dɛɛŋ.
ʔ*an* book 1 see this 3.SG red
intended: 'This book that I see is red.' (Thai)

Note also that in both (553) and (554), the relative clause is in a left-located position. It is possible to tell this because there is a pronoun *man* 'it' cross-referencing the head of the relative clause.

[11]These data were checked with one native Thai speaker.

Iwasaki & Ingkaphirom (2005) notes that in current usage, ʔ*an* as a complementizer is limited to being used in literary style. The complementizers *thîi* and *s*ə̂ŋ are more common. The more frequent use of the morpheme ʔ*an* is as a general classifier. Shan also has this classifier use of ʔ*ăn*.

This raises the question of what leads to the possibility of having IHRC constituents in Shan but not in Thai. In looking for an answer to this question, this chapter examines the features of left-located constructions in Shan, but a comprehensive answer requires further investigation. The following section focuses on the syntax of the definite left-located constructions.

5.2 Syntax of left-located constructions

As a first step towards investigating the IHRC constructions in Shan, it is useful to consider whether left-located nominals should be characterized as left-dislocated or topicalized constructions. Before that, I want to be clear about the terminology I will be using, defined in (555)–(557).

(555) **Left Located (LL):** a constituent in a position higher than the subject

(556) **Left Dislocated (LD):** a left-located constituent that is cross-referenced by a (covert or overt) pronoun in a lower position of the clause

(557) **Topicalized (Top):** a left-located constituent that has been moved from a position lower in the clause, leaving a gap

According to Ross (1967), left-dislocation is not sensitive to the Complex NP Constraint (CNPC), the Coordinate Structure Constraint (CSC), the Sentential Subject Constraint (SSC), or the Left Branch Condition on pied piping (LBC). Topicalization, in contrast, is subject to all of these in English. I will follow this claim and assume that co-reference between an LL constituent and an individual within one of these syntactic islands is because the LL constituent is left-dislocated rather than topicalized.

There is variation in Shan with respect to whether an overt pronoun in the matrix clause cross-references the left-located constituent. We have already seen in (534) and (535), repeated below in (558) and (559), that IHRC constructions can either appear to be in subject position, as in (558), or have an overt pronoun in the matrix argument position, as in (559). It is also possible to have a clear gap in the object position, as in (560). Examples like (558) leave open the question of whether IHRCs need to be left-located at all, or whether they can be in a subject position. It is possible that there is either a gap or null pronoun in subject position.

(558) [ʔǎn jíŋ lǎawŋúːn tɛm [pɔŋkwáam$_i$ ʔǎn mí tsɯ waa mɤ́ŋ táj] nân]$_i^{IHRC}$
COMP Ying Lao_Nguen write article COMP have name COMP$_V$ state Shan that

laj hâp sú-láŋwán
get receive prize

'The article that Ying Lao Nguen wrote that is called "Shan State" won a prize.' (IHRC?)

(559) [ʔǎn jíng lǎawŋúːn ʔǎw táaŋkǐn lêŋ mɛ́w tsɤ́ nân$_i$]^{IHRC} tsɤ́-khǎw$_i$ mí lǎaj
COMP Ying Lao_Nguen take food feed cat CLF.PL that CLF.PL-3 have how.many

tǒ
CLF.ANML

'The cats that Ying Lao Nguen feeds, how many are there?'

(560) [ʔǎn tsáaj lǎawkhám pɔ̀k wâj màakmoŋ$_i$ nân] jíng lǎawŋúːn ʔàm ʔǎw sɔ́j t$_i$
COP Jai Lao_Kham peel keep mango that Ying Lao_Nguen NEG take cut
'The mangoes that Jai Lao Kham peeled, Ying Lao Nguen did not cut.'

In the absence of a syntactic island, left-located nominals can either be co-referenced with an overt pronoun or not. In examples like (561) and (562), where there is nothing blocking movement, the cross-referencing pronoun is optional. This is true whether it is in a left-located position within an embedded clause (562) or at the far left position in the sentence (561).[12] The embedding verb *wôn* 'think' in these examples might be a bridge verb (Chomsky (1977)), meaning it would not be uncommon for this sort of verb to allow extraction out of the embedded clause.

[12]Shan allows left-located nominals within embedded clauses, as (562) demonstrated. This suggests that embedded clauses with complementizer *waa* have enough structure to support what might be an LL position.

(561) [tsáaj lăawkhám nâj]$_i$ háw wôn waa náaŋ lăawsɛ̌ŋ hûtsák (mán$_i$) lǐlǐ
Jai Lao_Kham this 1 think that Nang Lao_Saeng know (3) well
Jai Lao Kham, I think that Nang Lao Saeng knows (him) well.

(562) háw wôn waa [tsáaj lăawkhám nâj]$_i$ náaŋ lăawsɛ̌ŋ hûtsák (mán$_i$) lǐlǐ
1 think that Jai Lao_Kham this Nang Lao_Saeng know (3) well
I think that, Jai Lao Kham, Nang Lao Saeng knows (him) well.

Other, non-nominal constituents that are left-located cannot be co-referenced with a pronoun or by any other means. Most left-located elements are either nominal or prepositional phrases. The nominals often can have a co-referential constituent, but the PPs cannot.[13]

The following sections examine how Shan left-located constructions pattern with respect to the diagnostics for movement described by Fernández-Sánchez & Ott (2020). This data demonstrates that (i) some LL constructions must be base-generated since the co-referential pronoun appears within a syntactic island, (ii) for constructions where the LL constituent does not need an overt co-referential matrix clause pronoun, there might be movement, and (iii) there is some evidence that LL constituents can reconstruct.

5.2.1 Position of LL constituents

While there are some embedded clauses that support left-located constituents, such as those embedded under bridge verbs as in (562), for conditionals, as in (563), or embedded questions, as in (564), the left-located constituent cannot be moved to a left position within the subordinate clauses. Assuming that the if-clause in (563) is an event conditional as described by Haegeman (2003), it is not unexpected that it would not have a position for topicalization within the subor-

[13]The only other constituent that can be in a leftward position seems to be TP/VP in the presence of a modal, such as in (524), repeated in (1), but this section will not discuss this construction.

(1) [[náj tɯ̀k nâj]$_{PP}$ háw mɛ́p]$_{VP}$ ʔàm laj
in box this 1 hide NEG can
'We can't hide inside this box.'

dinate clause. This demonstrates that this is not an example of scrambling. In (564), there is a left-located embedded question subordinated to the verb *hû* 'know'. It is similarly ungrammatical to have the expression *pă̆a nâj* left-located to a position on the left within the embedded question. It is possible, however, to have a constituent left-located from the embedded question, as in (565).

(563) *pɔ́ [pă̆a nâj]$_i$ tsáaj lă̆awkhám tom t$_i$ jâw háw tě ǩǐn t$_i$
if fish this Jai Lao_Kham cook PRFV 1 IRR eat
intended: 'If this fish, Jai Lao Kham cooks it, I will eat it.'

(564) *[tsɔ̀ŋ [pă̆a nâj]$_i$ tsáaj lă̆awkhám tě tsû fáj sě tom (mán$_i$) nân] háw ʔàm hû
POLAR.Q fish this Jai Lao_Kham IRR assemble fire and cook 3 that 1 NEG know
'Whether, the fish, Jai Lao Kham will start the fire and cook it, I don't know.'

(565) [pă̆a nâj]$_i$ [tsɔ̀ŋ tsáaj lă̆awkhám tě tsû fáj sě tom (mán$_i$) nân] háw ʔàm hû
fish this POLAR.Q Jai Lao_Kham IRR assemble fire and cook 3 that 1 NEG know
'The fish, whether Jai Lao Kha will start the fire and cook it, I don't know.'

Typically, the subject of a question with the polar question marker *ts*ɔ̀ŋ will appear either before *ts*ɔ̀ŋ, as with *pʉ́n mán* 'his/her history' in (566) or after *ts*ɔ̀ŋ, as with *mán* 'he/she' in (567). If there is an overt pronoun cross-referenced with the LL constituent, the pronoun appears to the right of *ts*ɔ̀ŋ. I would argue that the expressions that appear to the right of *ts*ɔ̀ŋ are in subject position, and those that appear to the left of *ts*ɔ̀ŋ are left-located constructions. If we assume that *ts*ɔ̀ŋ appears in a left-periphery position due to the fact that it marks polar questions, the left-located construction would be in a higher left-periphery position.

(566) pʉ́n mán tsɔ̀ŋ lǐ hâa
history 3 POLAR.Q good Q.PRT
'Is his/her history good?'

(567) tsɔ̀ŋ mán jaam mí pʉ́n haaŋ-hâaj
POLAR.Q 3 ever have history apperance-bad
'Does he/she have a bad history?'

As mentioned in the previous section, when an IHRC is the subject of the matrix clause, it is difficult to tell whether the IHRC is in the subject position or a higher position. The question

marker *ts*ɔ̀ŋ can be used to address this question since it appears between the LL and the subject. The IHRC can only appear to the left of *ts*ɔ̀ŋ, as (568)–(569) show. The externally headed version of these examples can appear to the right of the question marker *ts*ɔ̀ŋ, as in (570). This supports the idea that Shan IHRC constructions can only appear in an LL position, whereas EHRCs can appear in argument positions.

(568) *tsɔ̀ŋ [ʔǎn jíŋ lǎawŋúːn tɛm [pɔŋkwáam$_i$ ʔǎn mí tsɯ waa
POLAR.Q COMP Ying Lao_Nguen write article COMP have name COMP$_V$

mɤ́ŋ táj] nân]$_i^{IHRC}$ laj hâp sú-láŋwán jù hâa
state Shan that get receive prize IPFV Q.PRT

intended: 'Did the article that Ying Lao Nguen wrote that is called "Shan State" win a prize?' (IHRC)

(569) [ʔǎn jíŋ lǎawŋúːn tɛm [pɔŋkwáam$_i$ ʔǎn mí tsɯ waa mɤ́ŋ táj] nân]$_i^{IHRC}$
COMP Ying Lao_Nguen write article COMP have name COMP$_V$ state Shan that

tsɔ̀ŋ laj hâp sú-láŋwán jù hâa
POLAR.Q get receive prize IPFV Q.PRT

'The article that Ying Lao Nguen wrote that is called "Shan State", did (it) win a prize?' (IHRC)

(570) tsɔ̀ŋ [pɔŋkwáam$_i$ ʔǎn jíŋ lǎawŋúːn tɛm waa mɤ́ŋ táj nân]$_i^{IHRC}$
POLAR.Q article COMP Ying Lao_Nguen write COMP$_V$ state Shan that

laj hâp sú-láŋwán jù hâa
get receive prize IPFV Q.PRT

'Did the article that Ying Lao Nguen wrote that is called "Shan State" win a prize?' (EHRC)

Shan allows more than one left-located constituent assuming PPs can count as LL constituents, as in (571) and (572). From this, it seems that LL constituents do not necessarily block other LL constituents. However, it is notable that in (571)–(572), there is an overt co-referent matrix clause pronoun, *mán* 'she/he/it', corresponding to the left-located nominal. This is likely because the

pronoun is a possessor in the matrix clause.[14,15]

(571) [ti náj rôtfáj nân], háw wôn waa [jíŋ lăawŋúŋ nâj]$_i$, pɤn tě ʔăw thŏŋ-ŋúŋ mán$_i$
at in train that 1 think that Ying Lao_Nguen this another IRR take bag-money 3
'On the train, I think that Ying Lao Nguen, someone will steal her wallet.'

(572) [ti náj rôtfáj nân], [jíŋ lăawŋúŋ nâj]$_i$, háw thàaŋ pɤn tě ʔăw thŏŋ-ŋúŋ mán$_i$
at in train that Ying Lao_Nguen this 1 think another IRR take bag-money 3
'On the train, Ying Lao Nguen, I believe that someone will steal her purse.'

Looking at (573) and (574), the order of the left-located prepositional phrase and the left-located name does not matter. Additionally, the pronoun cross-referencing the left-located 'Aung San Suu Kyi' is optional, so this could be an LD or a topicalized construction.

(573) [mɤ pǐ 1991 nân] [ʔɔ̀ŋsáantsûkjì nâj]$_i$ tsúmpɔ̆ŋkăan sú-ŋámjěn nòpɛ́(l) ʔăw
when year 1991 that Aung San Suu Kyi this committee award-peace Nobel take

sú-ŋámjĕn ʔap păn (mán-náaŋ$_i$)
award-peace deliver give (3-FEM)

'In 1991, Aung San Suu Kyi, the committee gave her the Nobel peace prize.'

(574) [ʔɔ̀ŋsáantsûkjì nâj]$_i$ [mɤ pǐ 1991 nân] tsúmpɔ̆ŋkăan sú-ŋámjěn nòpɛ́(l) ʔăw
Aung San Suu Kyi this when year 1991 that committee award-peace Nobel take

sú-ŋámjĕn ʔap păn (mán-náaŋ$_i$)
award-peace deliver give (3-FEM)

'Aung San Suu Kyi, in 1991, the committee gave her the Nobel peace prize.'

This is consistent with the idea that argument and adjunct topics in Chinese, Japanese, and Korean can be freely ordered as described by Paul & Whitman (2017). Note that in (575a)–(575b),

[14]There does seem to be a (gradient) contrast between alienable and inalienable possession in Shan. Body parts relations seem to be the most likely to have no overt possessor when the possessor is in a left-located position. An example of this can be seen with *hŏ* 'head' in (1).

(1) măa tŏ nâj hŏ món nàa
dog CLF.ANML this head round very
'This dog's head is round.'

[15]See Badan & Del Gobbo 2011 for a discussion of how PPs fit into the categorization of 'hanging topics' versus left-dislocations.

Liú Xiáobō 'Liu Xiaobo' is co-indexed with a pronoun in the matrix clause, similar to (573) and (574).

(575) (Chinese: Argument and Adjunct topics; Paul & Whitman 2017: (39))

a. Liú Xiáobō$_i$, 2010 nián, wěiyuánhuì shòyǔ-le tā$_i$ Nuòbèi'ěr hépíng jiǎng.
Liu Xiaobo 2010 year committee award-PRF 3SG Nobel peace prize
'Liu Xiaobo, in 2010, the committee awarded him the Nobel peace prize.'

b. 2010 nián, Liú Xiáobō$_i$, wěiyuánhuì shòyǔ-le tā$_i$ Nuòbèi'ěr hépíng jiǎng.
2010 year Liu Xiaobo committee award-PRF 3SG Nobel peace prize
'In 2010, Liu Xiaobo, the committee awarded him the Nobel peace prize.'

From the data in this section, it is clear that the left-located constituent is in a left-periphery position, and a left-located prepositional phrase and nominal construction can appear in the same clause. In this chapter, I will primarily focus on evidence connected to determining whether the left-located constructions are left-dislocated or topicalized. Thus far, I have not explained how to distinguish the two, so the next section will cover that.

5.2.2 Islands

This section examines the interaction between left-located constructions and syntactic islands for movement, such as adjunct islands and complex noun phrase islands. Sensitivity to island constraints is a significant difference between topicalized and left-dislocated constructions (Ross 1967).

According to Dayal (1996), left-adjoined constituents in Hindi cannot be extracted out of complex noun phrase islands. In (576), the left-adjoined constituent *ravi* is co-indexed with a trace inside of a complex noun phrase, and in (577), the correlative is co-indexed with a pronoun *vo* inside a complex noun phrase. This suggest that the trace in (576) and the pronoun in (577) that are co-referential with a left-adjoined constituent, have the status of a variable rather than a pronoun.

(576) *ravi$_i$ maiN [yeh baat ki t_i nahiiN aayegaa] jaantii thii
Ravi I this matter that not come-F know-P
'Ravi I knew the fact that he will not come."

(Ravi, I knew the fact that he will not come)

(Hindi, Dayal 1996: p.183, (44a))

(577) *jo vahaaN rahtaa hai maiN [yeh baat ki vo nahiiN aayeega] jaantii thii
who there life-PR I this matter that he not come-F know-P
"Who lives there, I knew the fact that he will not come."

(Hindi, Dayal 1996: p.183, (44b))

The Hindi data in (576)–(577) are interesting because they show that an overt link between the correlative and an expression within a syntactic island is unacceptable even when there is an overt co-referential pronoun inside of the island. This is different from the situation in Japanese where it seems that islands can be ameliorated by an overt external head as in doubly-headed relative clauses (assuming that DHRCs are derived from IHRCs) (Grosu & Hoshi 2019).

In Shan, LL constituents are not necessarily sensitive to complex noun phrase islands in the same way that Hindi correlatives are. For example, (578) demonstrates that a left-located name is possible, and (579) shows that a left-located externally headed relative clause is possible. For these, there is an obligatory overt pronoun cross-referencing the LL constituent.

(578) [jíŋ lăawŋúŋ nâj]$_i$ háw hû [lɔŋtáaŋ ʔăn *(mán$_i$) tě ʔàm máa tinaj]
Ying Lao_Nguen this 1 know reason COMP *(3) IRR NEG come here
'Ying Lao Nguen, I know the reason that she will not come here.'

(579) [phujíŋ ʔăn ʔăw táangk̆ǐn păn mɛ́w háw nân]$_i$ háw hû [lɔŋtáaŋ ʔăn *(mán$_i$) tě ʔàm
woman COMP take food give cat 1 that 1 know reason COMP *(3) IRR NEG

máa tinaj]
come here

'The woman who feeds my cat, I know why she will not come here.'

As (580) shows, when an LL constituent co-refers with the object of the matrix clause, an overt pronoun is similarly required. For the data in the remainder of the section, I do not have

examples demonstrating that it is ungrammatical to have a covert pronoun inside the island that cross-references the LL constituent. However, given that the pronoun is obligatory in (578)–(580), it is likely that the same holds true for the following examples.

(580) [phujíŋ ʔăn ʔăw táangkǐn păn mɛ́w háw nân]$_i$ háw hû [lɔŋtáaŋ ʔăn
woman COMP take food give cat 1 that 1 know reason COMP

tsáaj lăawkhám ʔàm lajtsăj *(mán-náaŋ$_i$)]
Jai Lao_Kham NEG like *(3-FEM)

'The woman$_i$ who feeds my cat, I know why Jai Lao Kham does not like her$_i$.'

(581) gives an example with a something that looks like an internally-headed relative clause in a left-located position.

(581) [ʔăn phujíŋ ʔăw táangkǐn păn mɛ́w nân]$_i$ háw hû [lɔŋtáaŋ mán waa jə̂nsăŋlɛ mán$_i$
COMP woman take food give cat that 1 know reason 3 COMP why 3

lajtsăɯ mɛ́w].
like cat

'The woman that fed cats, I know why she likes cats.'

(581) has the form of a subject-head internally headed relative clause, which—as discussed in the introduction to this chapter—tends to be judged less grammatical than its externally-headed counterpart. Such examples are not always ungrammatical.

Shan left-located constituents can also be co-indexed with a word inside of a relative clause, further supporting the idea that the left-located constituent is not necessarily moved. This can be seen in example (582). Here *jíŋ lăawŋúm* 'Ying Lao Nguen' is left-located and anaphoric reference to Ying Lao Nguen is inside the relative clause subject.

(582) [jíŋ lăawŋúŋ nâj]$_i$ [méw tŏ ʔăn tĵɲɔ́p nŭ jù hím hɤ́n mán$_i$ nân]
Ying Lao_Nguen this cat CLF.ANML COMP catch mouse stay near house 3 that

lajtsăɯ kĭn nŭ ʔăn ɲɔ́p máa laj.
like eat mouse COMP catch come ACH

'Ying Lao Nguen, the cats that catch mice near her house likes to eat the mice they catch.'

(583), similarly, shows that a nominal can be left-located in front of a IHRC clause. The relative clause is functioning as the subject of the main clause.

(583) [jíŋ lăawŋúŋ nâj]$_i$ [ʔăn méw tĵɲɔ́p nŭ jù hím hɤ́n mán$_i$ nân] tî laj
Ying Lao_Nguen this COMP cat catch mouse stay near house 3 that catch ACH

lăaj tŏ?
how_many CLF.ANML

'Ying Lao Nguen, the cats that catch mice near her house manage to catch how many mice?'

In (584), the LL constituent is cross-referenced by a pronoun both within another LL constituent, *ʔăn pɤn laat lɔŋ mán-náaŋ nân* 'what others said about her', and by the subject of the matrix clause.

(584) [náaŋʔɔ̀n nâj]$_i$ [ʔăn pɤn laat lɔŋ mán-náaŋ$_i$ nân]$_j$ mán$_i$ khɯ́n laj jín t$_j$
Nang_Orn this COMP other say about 3-FEM that 3 back ACH hear
'Nang Orn, what others said about her, she heard about herself.'

Since topicalization is impossible out of syntactic islands, those cases with obligatory overt pronouns within syntactic islands can be categorized as left-dislocations. It is not surprising that left-located nominals are insensitive to some islands if they are left-dislocations. Prince (1998) noted that fixing an island violation is one of the functions of left-dislocation. However, the fact that an overt pronoun appears in the preceding data does not rule out the possibility that movement is occurring when the co-referential element in the matrix clause is not overt.

5.2.3 Topic drop

The status of the overt or covert argument cross-referencing a left-located constituent is important for determining whether the LL constituent is a left-dislocated (LD) or topicalized expression. In the previous section, I argued that when a left-located constituent is co-referential with an overt pronoun within a syntactic island, that LL constituent is left-dislocated. This section discusses the role of null arguments in the matrix clause and whether they are used in conjunction with topicalized or left-dislocated constructions.

According to Huang (1984), Chinese null topics as in (585) and overt topics as in (586) bind variables in object position. This is on par with 'pronoun zap' in German (Ross 1982).

(585) $[_{TopP}$ [e$_i$] Zhāngsān$_j$ shuō [Lǐsì$_k$ bù rènshi t$_{i/*j/*k}$]].
Zhangsan say Lisi NEG know
'[Ø$_i$] Zhangsan$_j$ said Lisi$_k$ didnt know [him$_{i/*j/*k}$].'

(586) $[_{TopP}$ $[_{DP}$ Nèi ge rén]$_i$, [Zhāngsān$_j$ shuō [Lǐsì$_k$ bù rènshi t$_{i/*j/*k}$]]].
that CLF person Zhangsan say Lisi NEG know
'That man$_i$, Zhangsan$_j$ said Lisi$_k$ didnt know [him$_{i/*j/*k}$].'

(Chinese; Paul & Whitman 2017: (55a,b), citing (Huang 1984: 542, (31), (34)))

In (585), there is some previously mentioned individual indexed by *i* that serves as the null topic of this sentence. Two other individuals, *Zhāngsān* and *Lǐsì* are also overtly mentioned. The null argument of the verb meaning 'know' can only refer to the null topic, not to the two overtly mentioned individuals. Similarly, in (586), where the topic *Nèi ge rén* 'that man' is overtly mentioned, the null object of 'know' can only refer to 'that man'.

There is evidence of null topics in Shan. In (587), a particular owl has been mentioned in the previous sentence. The null object of this sentence is co-referential with the previously mentioned owl.

(587) méw háw ʔàm lajtsǎj _
cat 1 NEG like
'My cat doesn't like (that owl).'

The examples in (588)–(589) replicate the data in (585)–(586) for Shan. In (588), speaker A introduces two individuals, *tsáaj lǎawkhám* 'Jai Lao Kham' (index *j*) and another person (index *i*). Speaker B could respond with (589a) or (589b). In (589a), there is an overtly mentioned topic *phutsáaj kɔ̂ nâj* 'that man' in an LL position, and in (589b), there is no overtly mentioned topic. In both cases, the null object of *hûtsák* 'know' must refer to the topic (index *i*) rather than one of the other two individuals mentioned in the sentence.

(588) A: nâj$_i$ pěn t'ajkɔ̂ tsáaj lǎawkhám$_j$ hâa?
this COP friend Jai Lao_Kham Q.PRT
A: 'Is this$_i$ Jai Lao Kham$_j$'s friend.'

(589) B:

a. phutsáaj kɔ̂ nâj$_i$, tsáaj nùmsɛ̌ŋ$_k$ laat waa tsáaj lǎawkhám$_j$
man CLF.HUM this Jai Noom_Saeng say COMP$_v$ Jai Lao_Kham
ʔàm hûtsák t$_{i/*j/*k}$
NEG know
'That man$_i$, Jai Noom Saeng$_k$ says that Jai Lao Kham$_j$ doesn't know (him$_{i/*j/*k}$).'

b. B: _$_i$ tsáaj nùmsɛ̌ŋ$_k$ laat waa tsáaj lǎawkhám$_j$ ʔàm
Jai Noom_Saeng say COMP$_v$ Jai Lao_Kham NEG
hûtsák t$_{i/*j/*k}$
know
B: 'That man$_i$, Jai Noom Saeng$_k$ says that Jai Lao Kham$_j$ doesn't know (him$_{i/*j/*k}$).'

Under Huang's (1984) analysis, the null object is not a zero object pronoun, but instead a variable bound by the topic. The topic moves from the object position to the topic position and then can be deleted. This is meant to explain the asymmetry of zero subjects versus zero objects. Null subjects are also found in Chinese, but Huang (1984) argues that these are pronouns. If this process occurs in Shan, then null objects would be evidence of topicalization.

In externally headed relative clauses where we expect that the relative head is moving to its surface position, we see island effects when trying to extract a head from within another relative

clause, as in (590). So it seems that in this case, a null pronominal cannot circumvent the island. As expected, the same is true of (591). However, if there is movement involved in creating the relative clause headed by *phujíŋ* 'woman', we would still expect this to be ungrammatical, since movement is blocked.

(590) *[phujíŋ$_i$ ʔăn [măa$_j$ ʔăn t$_i$ lup t$_j$ mɤwáa nân] tăaj kwàa nân] pĕn mɔ̆
woman COMP dog COMP pet yesterday that die go that COP doctor
intended: 'The woman such that the dog she pet yesterday died, is a doctor.'

(591) *[phujíŋ$_i$ ʔăn [măa$_j$ ʔăn t$_j$ khóp t$_i$ mɤwáa nân] tăaj kwàa nân] pĕn mɔ̆
woman COMP dog COMP bite yesterday that die go that COP doctor
intended: 'The woman that the dog that bit (her) yesterday died, is a doctor.'

It does seem to be the case that overt resumptive pronouns can ameliorate relative clause island violations. This can be seen in how (592) is fixed by adding a pronoun in (593). Here, *luk*ʔɔ̆*n* 'child' is the head of the most embedded clause and there is a gap in that clause. The head of the less-embedded clause is *măa* 'dog' which is cross-referenced with a pronoun *mán* inside the relative.

(592) *mɤnâj háw hăn [măa$_i$ ʔăn tsáaj lăawkhám hûtsák [lukʔɔ̆n$_j$ ʔăn t$_j$ lɯplám t$_i$]]
today 1 see dog COMP Jai Lao_Kham know child COMP chase
Intended: 'Today I saw the dog$_i$ that Mr. Lao Kham knows the child that chased (it$_i$).

(593) mɤnâj háw hăn [măa$_i$ ʔăn tsáaj lăawkhám hûtsák [lukʔɔ̆n$_j$ ʔăn t$_j$ lɯplám mán$_i$ nân]]
today 1 see dog COMP Jai Lao_Kham know child COMP chase 3 that
'Today I saw the dog$_i$ that Mr. Lao Kham knows the child that chased it$_i$.

With externally headed relative clauses, it is not necessary to have the co-referential argument of the matrix clause be overt whether the argument is an object (594) or subject (595). However, it is necessary to have an overt pronoun when cross-referencing another referent of the relative clause, as (596) shows. It seems that (594) has an instance of topicalization, where *j* moved from the trace position to a left-located position leaving a bound variable behind. If null subjects in Shan can be pronouns, rather than bound variables, we might expect (596) to be grammatical. It

is identical to (594) except that the subject of the matrix clause is null. We might expect that the null subject pronoun could refer back to *tsáaj lăawkhám*, but that does not seem to be possible. We would expect (596) to be ungrammatical under the reading where the object is a null-pronoun cross-referencing *tsáaj lăawkhám* because it would require that a null-topic bind that pronoun. This null-topic would generate a Condition C violation since it would c-command *tsáaj lăawkhám*.

(594) [măa ʔăn khóp tsáaj lăawkhám$_i$ nân]$_j$ mántsáaj$_i$ ʔàm lajtsăj t$_j$
dog COMP bit Jai Lao_Kham that 3-MASC NEG like
'The dog$_j$ that bit Jai Lao Kham$_i$, he$_i$ doesn't like (it$_j$).'

(595) [măa ʔăn khóp tsáaj lăawkhám$_i$ nân]$_j$ t$_j$ ʔàm lajtsăj mántsáaj$_i$
dog COMP bit Jai Lao_Kham that NEG like 3-MASC
'The dog$_j$ that bit Jai Lao Kham$_i$, (it$_j$) doesn't like him$_i$.'

(596) *[măa ʔăn khóp tsáaj lăawkhám$_i$ nân]$_j$ t$_i$ ʔàm lajtsăj t$_j$
dog COMP bit Jai Lao_Kham that NEG like
intended: 'The dog$_j$ that bit Jai Lao Kham$_i$, (he$_i$/it) doesn't like (it$_j$/him$_i$).'

(597) [măa ʔăn khóp tsáaj lăawkhám$_i$ nân]$_j$ t$_{i/j}$ ʔàm lajtsăj mán$_{i/j}$
dog COMP bit Jai Lao_Kham that NEG like
'The dog$_j$ that bit Jai Lao Kham$_i$, (he$_i$/it$_j$) doesn't like (it$_j$/him$_i$).'

Note that the IHRC version of (594)is ungrammatical in (598). Here the internal head must be the dog since there is no overt pronoun for it in the matrix clause. This is an example of the subject internal heads, which tend not to be grammatical.

(598) *[ʔăn măa$_i$ khóp tsáaj lăawkhám$_i$ nân]$_j$ mántsáaj$_i$ ʔàm lajtsăj t$_j$
COMP dog bit Jai Lao_Kham that 3-MASC NEG like
intended: 'The dog$_i$ that bit Jai Lao Kham$_j$, (it$_i$) doesn't like him$_j$.'

It does seem that null objects might be bound variables. However, there is more work to be done to determine this and to demonstrate clearly how null subjects should be represented. Null pronouns certainly appear to have a different status than overt pronouns. One thing that is clear is that having multiple null arguments cross-referencing arguments in a left-located nominal clause is more difficult, if not completely ruled out.

According to Hoonchamlong (1992), having a left-located constituent co-referential with a null object is fine in Thai even when there is a syntactic island, as in (599).[16] Here, the left-located constituent is *su1daa1* 'Suda' and is co-referential with the object of *paa1* 'take' in the noun complement clause. In comparison, the Shan equivalent of (599) in (600) requires an overt pronoun. An expression like *tsáaj lăawkhám nâj* 'Jai Lao Kham' in (600) circumvents the island but only if it is cross-referenced with an overt pronoun within the syntactic island.

(599) [su1daa1 na3]$_i$ chan4 da3jin1 [khaaw2 waa3 c@@n1 phON3 phaa1 khaw4$_i$/EC$_i$ pay1
Suda TM I hear news COMP John just take s/he DIR

rooN1pha1yaa1baan1 mUa3 chaaw4 nii4]
hospital morning this

'Suda, I heard the news that John just took her/EC to the hospital this morning. (Thai, Hoonchamlong 1992: 93 (27))

(600) [tsáaj lăawkhám nâj]$_i$, khaa laj ŋín [khàaw waa naaŋ lăawŋɯ́m ʔăw *(mán$_i$)
Jai Lao_Kham this 1.HUM ACH hear news COMP$_v$ Nang Lao_Nguen take *(3)

kwàa tsú hóŋyăa mɤwáa-nân]
go to hospital yesterday-that

'Jai Lao Kham, I heard the news that Nang Lao Nguen brought him to the hospital today.'

From this we can see that Shan left-located constructions like (600) are sensitive to syntactic islands, they require an overt co-referential pronoun within the island. It is premature to conclude from this piece of data that they are distinct from Thai in this way, but it suggests that further comparison between Thai and Shan on this topic is warranted.

5.2.4 Binding

This section looks at some of the relationships between the binding conditions and LL constructions.

[16]Example (599) uses the orthography given in Hoonchamlong 1992.

The sentence in (601) sets up *náaŋ lăawŋúːn* 'Nang Lao Nguen' as the one salient referent for *mán-náaŋ* 'her' in possible follow ups (601a)–(601c). The two books in the context sentence, one about Nang Lao Nguen and one about Jai Lao Kham, makes it more likely to use a pronoun or name to distinguish the books in the follow ups. The other names introduced, *tsáaaj nùmsĕŋ* 'Jai Num Saeng' and *tsáaj lăawkhám* 'Jai Lao Kham', are names for men and thus not compatible with the pronoun *mán-náaŋ* 'her'.

It looks as though the left-located expression in (601a) is not reconstructing to the position below the pronoun, since it is possible for *náaŋ lăawŋúːn* 'Nang Lao Nguen' to be co-referential with the pronoun subject *mán-náaŋ* 'her'. It is also possible to have the pronoun as part of the left-located constituent be co-referential with a name in subject position, as (601b) shows. However, it is not possible to have the subject pronoun in (601c) be co-referential with the name that it c-commands. It is, of course, possible for the pronoun to refer to some other woman, who would be named *náaŋ lăawŋúːn* 'Nang Lao Nguen' if the pronoun is part of the NP refering to the book.

(601) mɤwáa khú păn náaŋ lăawŋúːn$_i$ lɛ tsáaaj nùmsĕŋ pâplik lɔŋ
yesterday teacher give Nang Lao_Nguen and Jai Num_Saeng book about

náaŋ lăawŋúːn$_{i/j}$ lɛ pâplik lɔŋ tsáaj lăawkhám sɔ̌ŋ pâp.
Nang Lao_Nguen and book about Jai Lao_Kham two CLF.BOOK

'Yesterday, the teacher gave Nang Lao Nguen$_i$ and Jai Num Saeng$_k$ two books, a book about Nang Lao Nguen$_i$ and a book about Jai Lao Kham$_l$.'

a. [pâplik lɔŋ náaŋ lăawŋúːn$_i$ nân]$_k$, mán-náaŋ$_{i/j}$ ʔàan t$_k$ jâw.
book about Nang Lao_Nguen that 3-FEM read PFV
'The book about Nang Lao Nguen$_i$, she$_{i/j}$ already read.'

b. [pâplik lɔŋ mán-náaŋ$_{i/j}$ nân]$_k$, náaŋ lăawŋúːn$_i$ ʔàan t$_k$ jâw.
book about 3-FEM that Nang Lao_Nguen read PFV
'The book about her$_{i/j}$, Nang Lao Nguen$_i$ already read.'

c. mán-náaŋ$_{j/*i}$ ʔàan pâplik lɔŋ náaŋ lăawŋúːn$_i$ jâw.
3-FEM read book about Nang Lao_Nguen PFV
'She$_{j/*i}$ already read the book about Nang Lao Nguen$_i$.'

As (602) shows, anaphors in Shan need to be bound locally. In (602), there is no left-located

constituent, but the *náaŋ lăawŋúⁿ* in the matrix clause cannot bind the anaphor *tŏtsawkàw* in the relative clause.

(602) #náaŋ lăawŋúⁿ$_i$ ʔàan pâplik ʔăn tsáaj lăawkhám$_j$ ʔăw păn [tŏtsawkàw mán-náaŋ]$_{*i/j}$
Nang Lao_Nguen read book COMP Jai Lao_Kham take give self 3-FEM

nân jâw
that PRFV

intended: 'Nang Lao Nguen$_i$ read the book that Jai Lao Kham$_j$ gave herself$_i$.'

?'Nang Lao Nguen$_i$ read the book that Jai Lao Kham$_j$ gave himself$_j$.'

(603) shows that the anaphor in the LL constituent cannot be bound by *náaŋ lăawŋúⁿ* in the matrix clause. It must be bound by *tsáaj lăawkhám*, which is also in the LL clause.

(603) #[ʔăn tsáaj lăawkhám$_i$ ʔăw pâplik păn tŏtsawkàw$_{i/*j}$ nân]$_k$ náaŋ lăawŋúⁿ$_j$ ʔàan
COMP Jai Lao_Kham take book give self that Nang Lao_Nguen read

t$_k$ jâw
PRFV

Intended: 'The book that Jai Lao Kham gave herself, Nang Lao Nguen read.'

Consultant comment: Sounds like the took the book and give it to himself.

(604)–(606) are relevant to the pronoun zap described for Chinese. In (604) the reflexive *tŏ-tsawkàw* is potentially ambiguous between Nang Lao Saeng and Nang Lao Nguen. In order to disambiguate to specify that Nang Lao Saeng is the antecedent, (605) is used.[17] Here, the antecedent is in an LL position. In order to make Nang Lao Nguen the antecedent, Nang Lao Nguen is placed in an LL position inside the embedded clause. In (606), *náaŋ lăawŋúⁿ* 'Nang Lao Nguen' is the topic of the embedded clause and is co-indexed with the pronoun and anaphor in that clause.

(604) #náaŋ lăawsĕŋ$_i$ wôn waa náaŋ lăawŋúⁿ$_j$ ʔàm lajtsăj tŏ-tsawkàw$_{i/j}$
Nang Lao_Saeng think that$_v$ Nang Lao_Nguen NEG like body-self
'Nang Lao Saeng$_i$ thinks that Nang Lao Nguen$_j$ doesn't like herself$_{i/j}$.'

[17]Note that a different reflexive *tŏ-mán-náaŋ* is used.

(605) [náaŋ lăawsɛ̌ŋ nâj]$_i$ mán$_i$ wôn waa náaŋ lăawŋúun$_j$ ʔàm lajtsăj tŏ-mán-náaŋ$_{i/*j}$
Nang Lao_Saeng this 3 think that$_v$ Nang Lao_Nguen NEG like body-3-FEM
'Nang Lao Saeng$_i$, she thinks that Nang Lao Nguen$_j$ doesn't like her$_{i/*j}$.'

(606) náaŋ lăawsɛ̌ŋ$_i$ wôn waa [náaŋ lăawŋúun nâj]$_j$ mán$_j$ ʔàm lajtsăj khúun
Nang Lao_Saeng think that$_v$ Nang Lao_Nguen this 3 NEG like back

tŏ-mán-náaŋ$_{*i/j}$
body-3-FEM

'Nang Lao Saeng$_i$ thinks that Nang Lao Nguen$_j$, she$_j$ doesn't like herself$_j$.'

Looking at (607), the possessive construction *pâplik lɔŋ tŏ-mán-náa*ŋ 'books about herself' can be in a left-located position in a contrastive context. This suggests that this expression is moved from a position below the subject to the left-located position, and it can reconstruct to that position.

(607) náaŋ lăawsɛ̌ŋ$_j$ ʔàm lajtsăj ʔàan pâplik kójkaa [pâplik lɔŋ tŏ-mán-náaŋ$_j$]$_i$ tɛ̂ mán$_j$
FEM Lao_Saeng NEG like read book but book about BODY-3-FEM CONTR 3

lajtsăj ʔàan t$_i$ tɛ̂tɛ̂
like read really

'Nang Lao Saeng doesn't like to read books, but books about herself, she really likes to read.'

It seems that in the cases of topicalization, the topicalized constituent can reconstruct to bind an anaphora, but does not have to reconstruct, to avoid having a pronoun c-command a name.

5.2.5 Summary

This section has discussed features of several different kinds of left-located structures. The findings of this section are summarized in table 5.1. The column headings 'null argument' and 'overt pronoun' indicate the form of the argument that cross-references the LL constituent or head of the the LL constituent. 'LL in embedded' indicates that a left-located construction is possible embedded under *waa* 'say' or 'that'. 'Multiple LL' means that it is possible to have more than one

left-located construction in a sentence. 'IHRC non-head x-ref matrix object' is whether the non-head argument of an IHRC construction can be cross-referenced by the object of the matrix clause. 'Pronoun x-ref name in LL' indicates whether a pronoun in the matrix clause can cross-reference a name in an LL constituent. 'LL anaphor x-ref matrix subject' indicates whether an anaphor within an LL constituent can cross-reference the subject of the matrix clause.

Table 5.1: Summary of syntactic features of LL constructions in Shan

	null argument	overt pronoun
LL in embedded	✓	✓
multiple LL	✓	✓
LL from island	✗	✓
IHRC non-head x-ref matrix object	✗	✓
pronoun x-ref name in LL	✓	✓
LL anaphor x-ref matrix subject	✓	-

For the properties that are fine with an overt cross-referencing pronoun in the matrix clause but not with a covert argument, the properties are all related to movement. This suggests that when there is co-reference between certain arguments of the matrix clause and arguments of the LL constituent without an overt pronoun linking them, topicalization movement is taking place. When there is an overt pronoun, it is a case of left-dislocation.

If we assume that the LL constituent is a kind of topic, it is useful to consider the features that characterize topics in 'topic prominent' languages as described by Paul & Whitman (2017).

Shan seems to pattern closely with Chinese with respect to these features. This is summarized

Table 5.2: Features related to topic prominence Paul & Whitman 2017 Table 1

		Chinese	Japanese	German	English
(i)	Obligatory topics	no	yes	yes	no
(ii)	Null topic	yes	yes	yes	no
(iii)	Overt spellout of Top°	optional	yes	yes	no
(iv)	*Wh*-in-situ	yes	yes	no	no
(v)	Generics obligatorily topicalized	no	yes	no	no
(vi)	Multiple topics	yes	yes	yes(?)	yes(?)
(vii)	Adjunct/argument topic order free	yes	yes	no?	no?
(viii)	Top°filled by Internal Merge	no	no	yes	yes

in Table 5.3. When I do not have the data demonstrating a feature, I put '?' in the table for Shan.

There are examples of null topics in Shan, such as (587), repeated in (608).

(608) mɛ́w háw ʔàm lajtsǎj ∅
cat 1 NEG like
'My cat doesn't like (that owl).'

The demonstrative *nâj*/*nân* 'this'/'that' often follows a left-located constituent even if it would not typically appear with a demonstrative, like a name. It does seem to be an obligatory marker with topics in some cases, as shown in (609). However, in examples like (1), repeated in (610), the demonstrative is not required.

(609) [jíŋ lǎawŋúːn *(nâj)]$_i$ háw hû waa mán$_i$ tě ʔàm máa tinaj
Ying Lao_Nguen this 1 know COMP 3 IRR NEG come here
'Ying Lao Nguen, I know she will not come here.'

(610) [tsáajkham, náaŋʔɔ̀n, lɛ náaŋsɛ̌ŋ]$_i$, tsɤ́-khǎw$_i$ jù ti hâan-khǎaj-tɔ̌sát
Jai_Kham Nang_Orn and Nang_Saeng CLF.PL-3.PL stay at story-sell-animal
'Jai Kham, Nang Orn, and Nang Saeng, they are at the pet store.'

(611) and (612) give examples of *wh*-in-situ in embedded (611) and un-embedded (612) clauses.

(611) ... mánnáaŋ ʔàm hû waa pěn mɛ́w phǎj
3-FEM NEG know COMP COP cat who
'. . . , She does not know whose cat it is.'

(612) mí khaw sǐ sǎŋ phɔŋ
have rice color what PRT
'What colors of rice do you have?'

Generics do not obligatorily need to be left-located, as (613) shows. Here, even though this means that the cat does not like owls in general, *kâw* 'owl' can appear in the typical object position. Compare with the Chinese example in (614), which similarly has a generically interpreted *yú* 'fish' in-situ.

Table 5.3: Features related to topic prominence for Shan Paul & Whitman 2017 Table 1

		Chinese	Shan	Shan example
(i)	Obligatory topics	no	?	-
(ii)	Null topic	yes	yes	(587)
(iii)	Overt spellout of Top°	optional	obligatory(?)	(609)
(iv)	*Wh*-in-situ	yes	yes	(611)–(612)
(v)	Generics obligatorily topicalized	no	no	(613)
(vi)	Multiple topics	yes	yes	(616)
(vii)	Adjunct/argument topic order free	yes	yes	(573)–(574)
(viii)	Top°filled by Internal Merge	no	?	-

(613) mɛ́w háw ʔàm lajtsǎj kâw
cat 1 NEG like owl
'My cat doesn't like owls.'

(614) Māo xǐhuān chī yú
cat like eat fish
'Cats like to eat fish.'(Chinese; Paul & Whitman 2017: (64a))

However, generic sentences often do have the generic nominal left-located with the topic marker *nâj*, such as in (122), repeated in (615).

(615) mǎa nâj kâj lɯplám hǎaŋ tǒ-tsawkàw mán khɯ́n
dog this often chase tail body-self 3 back
'Dogs often chase their own tails.'

(616) gives an example with multiple left-located constructions, both the name *náaŋ lǎawŋɯ́n nâj* 'Nang Lao Nguen' and the IHRC expression appear before the matrix clause subject.

(616) [náaŋ lǎawŋɯ́n nâj]$_j$ [ʔǎn tsáaj lǎawkhám ʔǎw pâplik$_i$ pǎn mán$_j$ nân]
Nang Lao_Nguen this COMP Jai Lao_Kham take book give 3 that

man-náaŋ$_j$ ʔàan t$_i$ jâw
3-FEM read PRFV

'Nang Lao Nguen, the book that Jai Lao Kham gave her, she already finished reading.'

Based on the data that we see so far, Shan patterns with Chinese with respect to features of topic prominence. Therefore, I propose the following structure for left periphery in Shan:

(617) Shan left periphery (based on Paul & Whitman 2017: (42))

Comp [$_{\text{topic field}}$ Top*] *kɔ*-Focus T v VP

If we take overt pronouns co-referencing the left-located topic to indicate the presence of a syntactic island, then we see evidence for both left-dislocations and topicalizations in the left-located position. I will propose that this is a topic position, given that it appears before elements that are likely in the left periphery.

5.3 Semantic analysis

The left-located position in Shan is associated with a certain type of information: given information. It is not possible to have a focused element in a position for left-dislocations or topicalizations. The position for focus is at the end of the sentence. This is consistent with an existential closure account such as was proposed by Diesing (1992) since it would be expected that nominals in the VP could undergo existential closure, which is associated with introducing new information. The far left position in a clause is outside of the scope of existential closure, so it would need to be either definite or generic.

I propose that the topic left-located construction is associated with domain restriction connected to the situation of the matrix clause, perhaps the requirement that the left-located constructions be in that position stems from their being in the restrictor of such quantification. The left-located construction sets the frame of the sentence.

For left-dislocated expression, the assumption is that there is a pronoun in the matrix clause that is co-referential with the left-dislocated constituent. In this case, there need to be a slightly different semantics. (618) gives the proposed semantics of left-dislocation (Top$_{LD}$) that takes a proposition and a left-dislocated noun. If the topicalization construction semantically frames the sentence in a similar way to the left-dislocation, that can be accomplished with a slightly modified

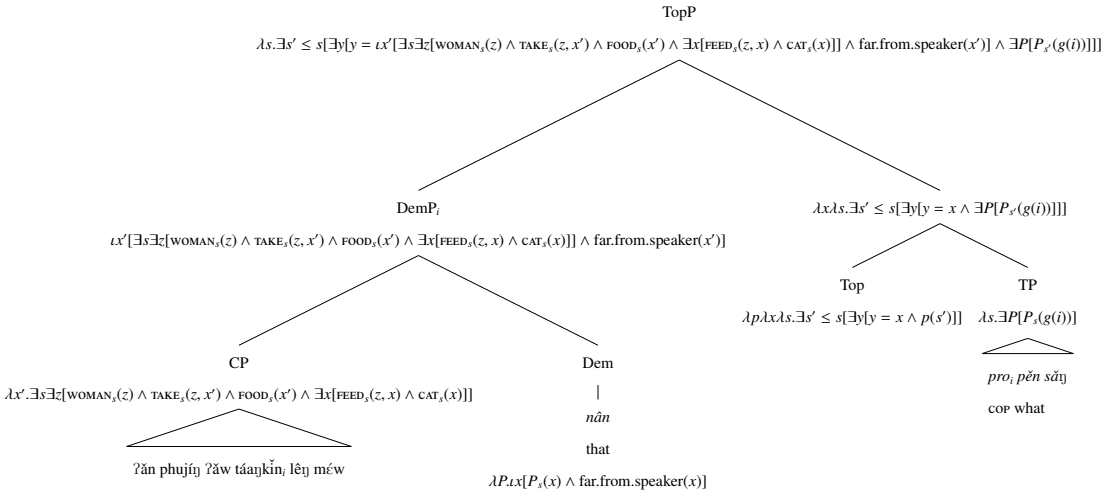
semantics, as in (619). A generic version is given in (620).[18] This just take into account that clause that the topicalized expression combines with includes a variable that is lambda-abstracted over.

(618) $[\![\text{Top}_{LD}]\!] = \lambda p \lambda x \lambda s . \exists s' \leq s [\exists y [y = x \wedge p(s')]]$

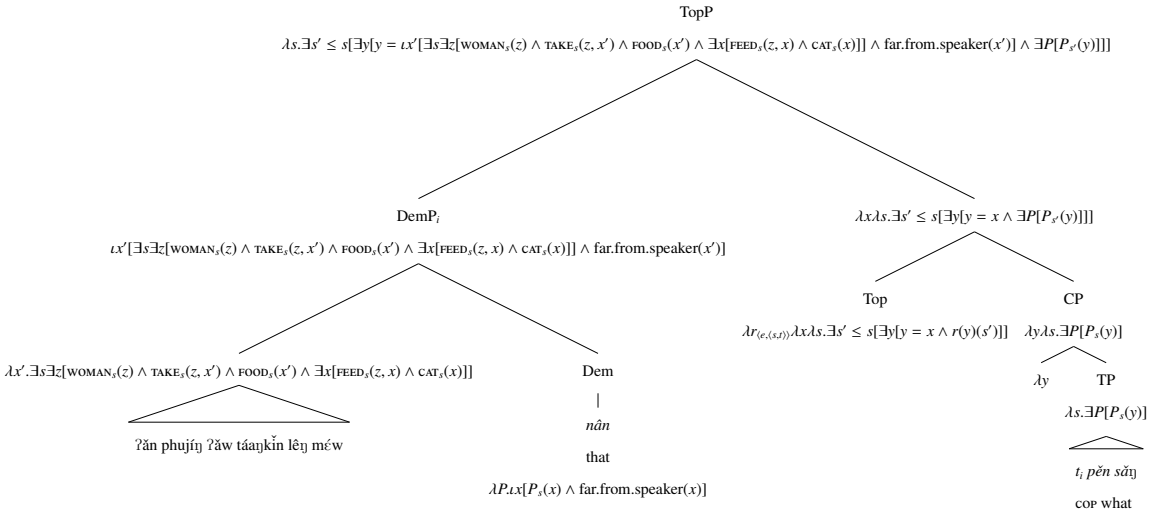
(619) $[\![\text{Top}_{Top}]\!] = \lambda r_{\langle e,\langle s,t\rangle\rangle} \lambda x \lambda s . \exists s' \leq s [\exists y [y = x \wedge r(y)(s')]]$

(620) $[\![\text{Top}_{Gen}]\!] = \lambda r_{\langle e,\langle s,t\rangle\rangle} \lambda x \lambda s . Gn\ y, s' \leq s [y = x \wedge r(y)(s')]$

(621)



(622)



[18] Compare the semantics of topicalization to the generic (97), repeated in (1). Both involve restriction of the situation being described by the clause. Perhaps these could be unified if generic sentences just involve (possibly covert) topicalization.

(1) $[\![\text{Gen}_{entity}]\!] = \lambda y_e \lambda h_{\langle e,\langle s,t\rangle\rangle} \lambda s' . Gn\ x, s \leq s' [x = y \rightarrow h(x)(s)]$

5.3.1 IHRC construction

If the role of the IHRC construction is to use the head and event of the relative clause to frame the sentence, it would make sense that that clause needs to be in the left-located position of the sentence.[19] This is somewhat similar to Shimoyama's (1999) e-type proposal, where the IHRC clause is adjoined to the matrix IP. In this proposal for Japanese, the IHRC clause must move covertly to this higher clause position. However, in Shan it is necessary for the IHRC to either be externally merged to that position or moved there overtly.

[19]The construction in (533), repeated in (1), is very similar to Japanese clefts as described by Kizu (2005).

(1) [ʔǎn tsáaj lǎawkhám sûɯ nân] tê pěn sɤ
COMP Jai Lao_Kham buy that CONTR COP shirt
'What Jai Lao Kham bought was a shirt.'

As shown in (2), Japanese clefts consist of a free relative *Hanako-ga atta no* 'the one that Hanako met', a focus phrase *sono hito-ni* 'that person' and a copula verb *da*. In comparison, the Shan example in (1) has a free relative *ʔǎn tsáaj lǎawkhám sûɯ nân* 'what Jai Lao Kham bought', a focus phrase *sɤ* 'shirt', and a copula verb *pěn*. Kizu (2005) proposes the structure in (3) for Japanese cleft constructions.

(2) [Hanako-ga [e] atta no] -wa [sono hito-ni] da
Hanako-NOM met NM TOP that person-DAT COP
'It was that person that Hanako met.' (Japanese; Kizu 2005: 3, (6))

(3) $[_{CP}\ Op_i\ [_{IP} \ldots t_i \ldots]\ no_i]$ -wa NP_i-(CASE) da
NM TOP COP

The morpheme *tê* is associated with contrastive focus, which is a feature often associated with cleft constructions (Destruel et al. 2018). There are also three implications associated with cleft constructions, which Halvorsen (1977) first identified. For the English example in (4), there is the PREJACENT INFERENCE in (4a), the EXISTENTIAL INFERENCE in (4b), and the EXHAUSTIVITY INFERENCE.

IMPLICATIONS OF CLEFT CONSTRUCTIONS (Destruel et al. 2018: (1))

(4) It's David who smiled.
   a. David smiled.
   b. Someone smiled.
   c. No one other than David smiled.

These implications are compatible with the equative analysis of pseudoclefts by Heycock & Kroch (1999). According to them, the free relative portion of the pseudocleft denote and individual rather than a set. A definite interpretation of the free relative is compatible with both the existential and exhaustivity inference.

The IHRC constructions, and the nominal left-located constructions more generally, do seem to have the prejacent inference and existential inference, since the constructions in these positions are definite. It is not clear that the exhaustivity inference applies to all kinds of IHRC constructions. Additionally, they do not all involve a copula structure. According to Kizu (2005), Japanese cleft and topicalization constructions both involve A′ movement, for cleft constructions this is the operator movement inside of the free relative and in topicalization constructions, this is movement from the argument position.

There are two significant ways in which this construction differs from e-type anaphora: (1) there is only one head of the IHRC construction and (2) the head is not quantificational. I will propose an e-type analysis for examples of quantificational (donkey) anaphora later in this chapter. However, in the case of the IHRC constructions, the co-indexed matrix clause argument is either a variable or a regular pronoun. The effect of the left-located IHRC construction is to restrict the situation under discussion.

Now that I have described how the IHRC construction contributes to the meaning of the sentence, I will go over how the meaning of the IHRC is built up. To do this I will use the two examples (527), repeated in (623), to represent non-subject heads and (531), repeated in (624), to represent an adjunct free relative cleft construction.

(623) [ʔăn phujíŋ ʔăw táaŋkǐn$_i$ lêŋ méw nân]$_i^{IHRC}$ pěn săŋ
COMP woman take food feed cat that COP what
'The food that the women fed to the cats is what?' (IHRC)

(624) [ʔăn phujíŋ ʔăw táaŋkǐn lêŋ méw nân] pěn ti-lăj
COMP woman take food feed cat that COP place-which
'Where is it that the women fed food to the cats?'

(ADJUNCT Q)

I am going to assume that the demonstrative *nân* 'that' has normal demonstrative semantics and takes a something of type $\langle e, t\rangle$ as its argument. To derive the IHRC, the relative operator will originate in the nominal phrase that contains the head. A $\lambda x.x = y$ trace is left by the operator when it moves to SpecCP in order for the trace to be able to combine with the overt internal head through predicate modification. Then, Derived Predicate Predication takes place at the verb level as usual. For this sort of analysis, there would need to use a different trace type and assume that there can be movement out of the nominal expression.[20]

[20]Another option would require that the verb and internal head combine without Derived Predicate Predication, from chapter 2, taking place. Existential closure would take place after the verb and internal head had composed. Then, instead of existential closure taking place at the VP level, there is an operator there that moves from the SpecVP to SpecCP. The operator leaves behind a variable trace that saturates the argument associated with the complement of the verb. Then, at the CP level, that variable is lambda abstracted over. This is somewhat similar to the Choose

For the adjunct free relative, the operator originates in an adverb phrase adjoined to the vP level that specifies that the event takes place at the specified location. Here, the operator leaves an *e* type trace as is typical of relative clauses. That trace combines with this null adverb phrase semantics. The operator moves to SpecCP and abstracts over the variable trace.

(625) shows the derivation of (623), and (626) shows the derivation of (624). The semantics given for these demonstrates certain assumptions about the serial verb construction *ʔǎw táaŋǩǐn lêŋ mɛ́w* 'take food (and) feed cats'. For examples like (623)–(624), where there are multiple non-subject arguments, it might be important that each of these arguments be the object of a verb in a serial verb construction in order to be a possible IHRC head. I am not making any claims about how serial verb constructions work at this point. The way it is presented is that each verb is transitive and takes one object, existentially closing over that object argument when they combine. Then, the two verbs combine by predicate modification.

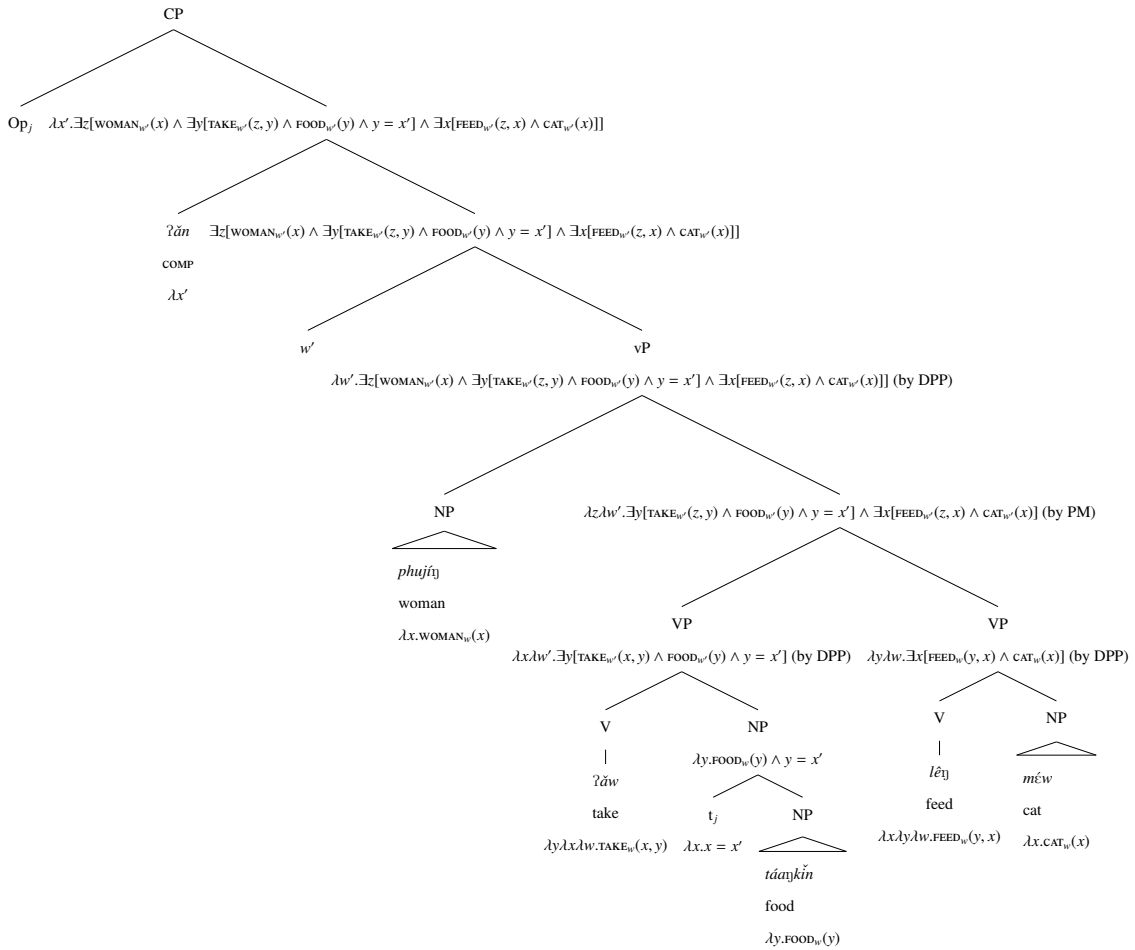
For (623), the internal head is the nominal *táaŋǩǐn* 'food'. In the derivation of (623) in (625), after the $\lambda x.x = y$ operator trace combines with the nominal meaning 'food', they combine with the verb meaning 'take', and Derived Predicate Predication takes place. Then, after the situation argument has been saturated by situation variable or pronoun, the trace is abstracted over, creating a predicate of type $\langle e, t\rangle$ which can combine with the demonstrative.

For (624), the relative clause head is the location of the event. In the derivation of (624) in (626), the operator trace appears in the AdvP the specifies the location of the event. When this trace is abstracted over, it creates something of type $\langle e, t\rangle$ again. For both the derivation in (625) and (626), a situation variable/pronoun must combine with the proposition denoted by the relative clause before the operator variable is abstracted over. For the IHRCs, this situation pronoun might

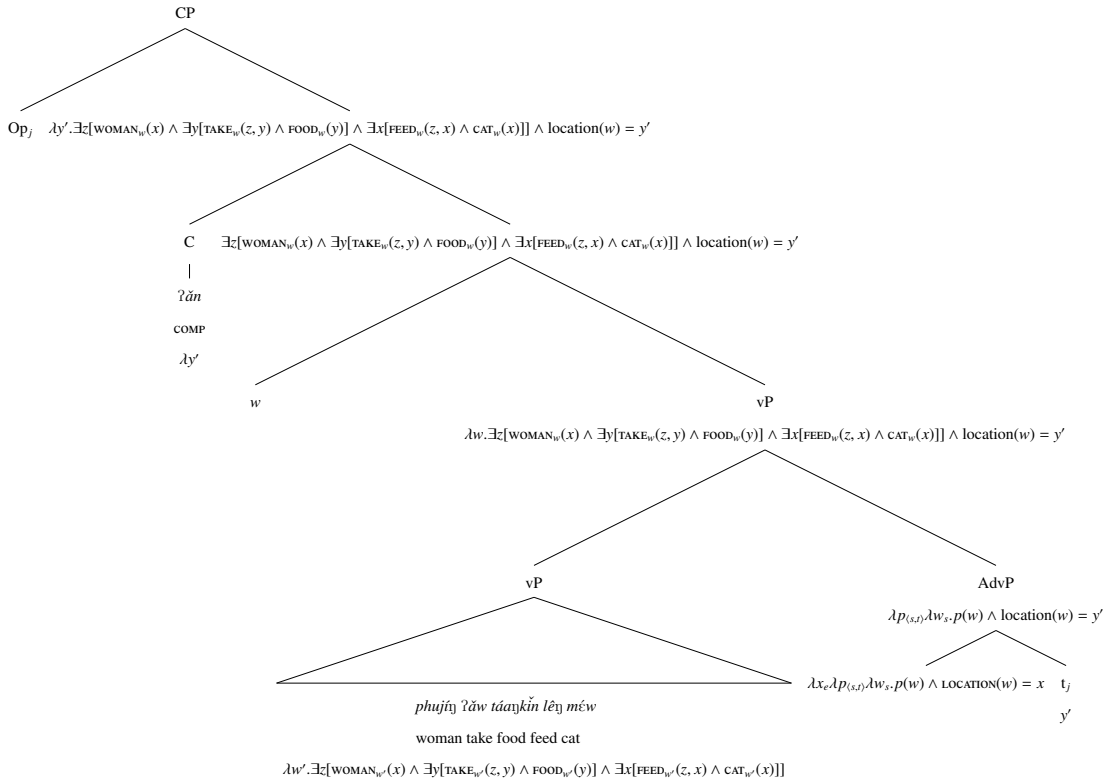
Role operator from Grosu (2010); Grosu & Landman (2012); and Landman (2016b). This operator combines with the VP of an internally headed relative clause and chooses the internal head from the event denoted by the VP. This operator also involves movement from that position to SpecCP to abstract over the internal head. The analysis differs significantly from the quantificational disclosure analysis in that the operator takes the place of existential closure over a particular argument of the VP. As a result, the operator is not choosing a semantic role from the event. The internal head would be identified by where the operator merges.

be connected to the obligatory definiteness of the expression denoted by the IHRC, so it might be that IHRCs can only combine with situation pronouns, while other relatives might combine with a variable or a pronoun, depending on whether it is presupposed or not.

(625) IHRC



(626) FRC



Then the demonstrative would combine with the $\langle e, t\rangle$ type constituent to be able to refer to that discourse referent specified by the IHRC. That individual is then used to restrict the domain of the matrix clause.

5.4 The particle *k*ɔ

There are two particles that commonly appear between left-located constituents and the matrix clause. These are *t*ɛ̂ as in (607), repeated in (627), which is used for contrastive focus, and *k*ɔ which seems to be associated with focus and quantification. Here, I will mostly be talking about *k*ɔ. In Moeng (1995), *k*ɔ is defined as 'also, then, as well as' or 'although', and in Cushing (2000) it is defined as 'although'. For both *t*ɛ̂ and *k*ɔ, the old/given information appears to the left of the particle and the new information appears to the right.

(627) náaŋ lăawsɛ̆ŋ$_j$ ʔàm lajtsăj ʔàan pâplik kójkaa [pâplik lɔŋ tŏ-mán-náaŋ$_j$]$_i$ tê mán$_j$
FEM Lao_Saeng NEG like read book but book about BODY-3-FEM CONTR 3

lajtsăj ʔàan t$_i$ têtê
like read really

'Nang Lao Saeng doesn't like to read books, but books about herself, she really likes to read.'

There are examples such as (628) where the particle *k*ɔ seems to signal that additional information is being given about the individual described by the nominal to the left of *k*ɔ. Here, *kâw tŏ nɯŋ* 'one owl' is introduced in the first clause. The following clause refers back to the owl with *kâw tŏ nân* followed by the particle *k*ɔ.

(628) [kâw tŏ nɯŋ]$_i$ máa tsáp nɤ̆ tonmâj ʔăn mí hímtsăm hɤ́n háw nân sě
owl CLF.ANML one come perch on tree COMP have near house 1 that and

[kâw tŏ nân]$_i$ kɔ mɛ́w háw ʔàm lajtsăj mán$_i$
owl CLF.ANML that PRT cat 1 NEG like 3

'An owl perched in a tree near my house. That owl, my cat doesn't like it.'

Looking at (628), it is possible to see that the particle *k*ɔ is clearly structurally lower than the left-located constituent *kâw tŏ nân* since it appears to the right of that constituent. This use of kɔ seems to be somewhat similar to *yě* 'also' or dōu 'all' in Chinese as described by Badan & Del Gobbo (2011) and Paul & Whitman (2017).

In addition to this use of *k*ɔ, there is also a quantificational use, as in (629). Here *k*ɔ is connecting two clauses rather than a nominal expression and a clause. The first clause is *mán-náa*ŋ *ts*ɔ*j mɤ́-lăj* 'whenever she helps (him)', and the second is *tsáajkhám jíntsómkh*ɔ*ptsăj th*ŭŋ *mán-náa*ŋ 'Jai Kham thanked her'. The expression *mɤ́-lăj* is an indeterminate pronoun expression. When it appears in front of the particle *k*ɔ it has the interpretation 'whenever, whichever time' in other contexts it can have the interpretation of a question ('at which time?'). The particle *k*ɔ seems to function as a universal quantifier of situations, mapping the clause to the left into the restrictor and

the matrix clause to the right into the nuclear scope of the quantifier. The types of expressions with this quantificational interpretation are not nouns but clauses or indeterminate pronouns.

(629) tsáajkhám$_i$ lɛ náaŋʔìŋ$_j$ kwàa tsɔ́m kăn mɤ-lăj kɔ jâw, náaŋʔìŋ$_j$ tsɔj
Jai_Kham and Nang_Ing go go.along together time-which PRT PRF Nang_Ing help

tsáajkhám$_i$ tàasè. [$_{CP}$ mán-náaŋ$_j$ tsɔj mɤ-lăj] kɔ tsáajkhám$_i$ jíntsómkhɔptsăj
Jai_Kham always 3-FEM help time-which PRT Jai_Kham thankful

thŭŋ mán-náaŋ$_j$
towards 3-FEM

'Jai Kham and Nang Ing get along together all the time. Nang Ing always helps Jai Kham. Whenever Nang Ing helps Jai Kham, Jai Kham thanks her.'

(630) is a generic version of (629), which uses indefinite nominals rather than names. This simply demonstrates that the quantification involved with the particle *kɔ* can also be used in donkey anaphora. Example (630) involves quantification over situations and the associated woman and man in each situation. The 'help' situation is the restrictor of the 'thanks' situation.

(630) [$_{CP}$ pɔ́ phujíŋ kô nɯŋ tsɔj phutsáaj (kô nɯŋ) mɤ-lăj] kɔ phutsáaj kô
if woman CLF.HUM one help man CLF.HUM one time-which PRT man CLF.HUM

nân jíntsómkhɔptsăj thŭŋ phujíŋ kô nân tàasè
that thankful towards woman CLF.HUM that always

'If a woman helps a man, that man is always thankful to the woman.'

In (631), the left-located construction *méw nâj* indicates that the sentence is about general properties of cats. The if-clause introduces a restriction of situations where cats see a mouse (*nŭ tŏ lăj* 'which mouse'). Then, *nŭ* 'mouse' in the matrix clause refers the particular mouse in a given situation of mouse-seeing.

(631) [méw nâj]$_i$ [pɔ́ t$_i$ hăn nŭ tŏ lăj] kɔ t$_i$ tĕ lɯplám nŭ tàasè
cat this if see mouse CLF.ANML which PRT IRR chase mouse always
'Cats, if they see whichever mouse, will always chase the mouse.'

A significant feature of donkey anaphora is that it involves pronouns that refer back to antecedent that have been quantified over. These constructions differ significantly from topicalization constructions in that a matrix subject and object can both refer to arguments of the subordinate clause. These arguments also quantificationally co-vary with one another.

If we take seriously the idea that null objects are always variables and not pronouns, we would need to explain the data in (632) by saying that donkey anaphora examples can bind multiple variables in the matrix clause without having moved from there.[21] This seems like a desirable position to take given that these examples have a generic interpretation suggesting that there is quantification over instances of the mouse. The if-clause in (632) introduces a mouse, and a null argument in the matrix clause refers back to the mouse in a given situation.

(632) [pɔ́ [méw tǒ nân]$_i$ hăn [nŭ (tǒ nɯŋ)]$_j$ năj] t$_i$ tě lɯplám t$_j$ tàasè
if cat CLF.ANML that see mouse CLF.ANML one then IRR chase always
'If that cat sees (one) mouse, (it) will chase (it) always.'

(633) and (634) show the differences between left-located constructions with and without kɔ. In (633), *ʔăn jíŋ lăawŋún laat nân* 'what Ying Lao Nguen said' refers to a specific things that Ying Lao Nguen said. In contrast, *jíŋ lăawŋún nâj laat săŋ kɔ...* in (634) means 'whatever Ying Lao Nguen said'. It seems to universally quantify over the things that Ying Lao Nguen said.

(633) tsáaj lăawkhám hát tɔ̀pthěŋ waa [$_{DemP}$ ʔăn jíŋ lăawŋún laat nân] ʔàm
Jai Lao_Kham be.bold dispute COMP$_v$ COMP Jing Lao_Nguen say that NEG
thùkmɛn
correct
'Jai Lao Kham argued that what Ying Lao Nguen said wasn't correct.'

(634) tsáaj lăawkhám hát laat waa [$_{CP}$ jíŋ lăawŋún nâj laat săŋ] kɔ ʔàm thùkmɛn
Jai Lao_Kham be.bold say COMP$_v$ Jing Lao_Nguen this say what PRT NEG correct
'Jai Lao Kham dared to say that whatever Ying Lao Nguen says is wrong.'

[21] Alternatively, I could say that null objects in Shan can also be pronouns.

(635) includes an example with two instances of the contrastive marker *tɛ̂*. The first instance is preceded by *ʔăn jíŋ lăawŋúun sûu sǐ lăm nân* 'the black things that Ying Lao Nguen bought' refers to the specific items that have already been bought. Reference to a specific object includes the demonstrative *nâj* or *nân*. The second instance of *tɛ̂* is preceded by *ʔăn háw hû* 'what I know', which is part of a specificational construction.

(635) kámnân tsáaj lăawkhám tɔ̀p waa [ʔăn jíŋ lăawŋúun sûu sǐ lăm nân] tɛ̂
then Jai Lao_Kham reply that COMP Ying Lao_Nguen buy color black that CONTR

ʔàm hû. [ʔăn háw hû] tɛ̂ mán khăaj sɤ̌ sǐ lɛ̌ŋ sɔ̌ŋ phtŭn lɛ màakhŏ
NEG know COMP 1 know CONTR 3 sell clothes color red two CLF.FLAT and hat

sǐ khăaw săam hòj
color white three CLF.RND

'TLK replied, "The black things YLNg bought I do not know. What I know is she sold two red shirts and three white hats."'

This section has demonstrated that there is a syntactic position below topics in the left periphery that is associated with focus and quantification over situations. The particle *kɔ* separates old information (to the left) from new information (to the right). It also is associated with universal quantification over situations where the information to the left is in the restrictor and the clause to the right is in the nucleus of the scope. The particle *tɛ̂* is in a somewhat similar syntactic position when compared to *kɔ*, but it is associated with contrastive focus.

5.4.1 Particle kɔ: Semantics

I propose that the semantics for the particle kɔ is (636). With this definition, kɔ functions as a connective that universally quantifies over the event argument of the restrictor, which can be a proposition or an individual, and the proposition in its nuclear scope. The portion to the right of kɔ consists of new information.

(636) $[\![\text{kɔ}]\!] = \lambda p_{\langle s,t\rangle}\lambda q_{\langle s,t\rangle}\lambda w.\forall w' \leq w[p(w') \rightarrow q(w')]$

The proposed semantics for kɔ in (636) is essentially identical to the semantics for *if* in English, shown in (637).

(637) STRICT IMPLICATION (Von Fintel & Heim 2011: (100))

$[\![\text{if}]\!]^{w,g} = \lambda p \leq D_{\langle s,t\rangle}\lambda q \leq D_{\langle s,t\rangle}.\forall w'[p(w') = 1 \rightarrow q(w') = 1]$

$\lambda p_{\langle s,t\rangle}\lambda q_{\langle s,t\rangle}.\forall w'[p(w') \rightarrow q(w')]$

(638) and (639) give an approximate semantics for (633) and (634). This demonstrates the difference between the left-located definite in (638) which does not include kɔ and the left-located indeterminate construction in (639) which does include kɔ. For (633), the left-located construction connects the two clauses with a existentially closed situation variable. The definite construction contributes the ι interpretation. For (634), the indeterminate pronoun *sǎŋ* 'what'/'whatever' is under the scope of the universal quantifier over situations that is part of the meaning of kɔ.

(638) (633) $\rightsquigarrow \exists s[\text{wrong}(\iota x[\text{said}(\textit{jing}, x, s)], s)]$

(639) (634) $\rightsquigarrow \forall s[\exists x[\text{said}(\textit{jing}, x, s)] \rightarrow \text{wrong}(\iota x[\text{said}(\textit{jing}, x, s)], s)]$

We see that the left-located constructions more broadly are associated with quantification over situations. If the topic position serves as a frame for the rest of the sentence, it makes sense that a left-located constituent acts as the restrictor of the situation variable for the matrix clause.

If existential closure takes place at the VP level, and words to the left of *kɔ* are above the VP level, we would expect that bare nouns in that position would not be indefinite. Here is one example of a definite bare noun with (640). Here, *pǎlik* 'police' refers back anaphorically to two individuals that have been mentioned already in this story.

(640) pălik kɔ kwàa khaw ti náj hɤ́n nân sě pălik-tsáaj thăam waa…
police PRT go enter at in house that and police-MASC ask that
'The police when into that house and the policeman asked…'

With this in mind, it would be useful to look at example (641). Here, the nouns *hòj* 'round thing' and *kɔ̌ŋ* 'pile' appear immediately before the particle *kɔ*. Therefore, it makes sense that these nouns have a definite interpretation referring to an aspect of the left-located nominal expression *màakkhɔ̌ kɔ̌ŋ nâj* 'this pile of jujube'. The bare nouns *hòj* 'round thing' and *kɔ̌ŋ* 'pile' are interpreted as unique definites. The extension of each property is given with respect to the particular situation. The situation is defined by the left-located constituent *màakkhɔ̌ kɔ̌ŋ nâj* 'this pile of jujube', so the noun *hòj* will refer to the particular jujube within that pile and *kɔ̌ŋ* will refer to the pile itself. If the purpose of the topic position is to define the situation in which the matrix clause is evaluated, it is not surprising that there is a 'super-set/subset relation' between multiple topics (Paul & Whitman 2017), where the topic to the left is the superset of the topic to the right.

(641) màakkhɔ̌ kɔ̌ŋ nâj hòj kɔ hòj jàj kɔ̌ŋ kɔ kɔ̌ŋ jàj
jujube pile this CLF.RND PRT CLF.RND big pile PRT pile big
'This pile of jujube, the jujube are big and the pile is big.'

(642) (641) $\rightsquigarrow$

$\lambda s.\exists s'' \leq s[\exists y'[y = \iota x'[\textsc{jujube}_{s''}(x') \wedge \mu_{\text{pile}}(x') = 1] \wedge \forall w' \leq s''\forall y''[y'' = \iota z[\textsc{rnd.thing}_{w'}(z)] \rightarrow \textsc{big}_{w'}(y'')] \wedge \forall v' \leq s''\forall y[y = \iota z[\textsc{pile}_{v'}(z)] \rightarrow \textsc{big}_{v'}(y)]]]$

There is a situation s'' that contains exactly one pile of jujube, y'. For all sub-situations w' in s'' the jujube are big. For all sub-situations v' in s'' the pile is big.

Looking at (641), the topicalized *màakkhɔ̌ kɔ̌ŋ nâj* 'this pile of jujube' is evaluated with respect to the situation s''. Then, for example, the expression *hòj kɔ hòj jàj* 'the jujube are big' is evaluated with respect to a situation w' which is part of s'' ($w' \leq s''$). As a result, the 'round things' and 'pile' that are part of the main clause are components of the topicalized constituent.

5.4.2 Donkey anaphora

Returning to the donkey anaphora examples from the previous chapter, repeated below, we can apply the semantic account of left-dislocated constituents to these examples. Note that examples (493) and (494)—repeated in (645) and (648—both make use of the particle *kɔ*. (643) uses an if-clause which ends with *nâj-ts*ɯ̌ŋ which is also associated with quantification over hypothetical situation.

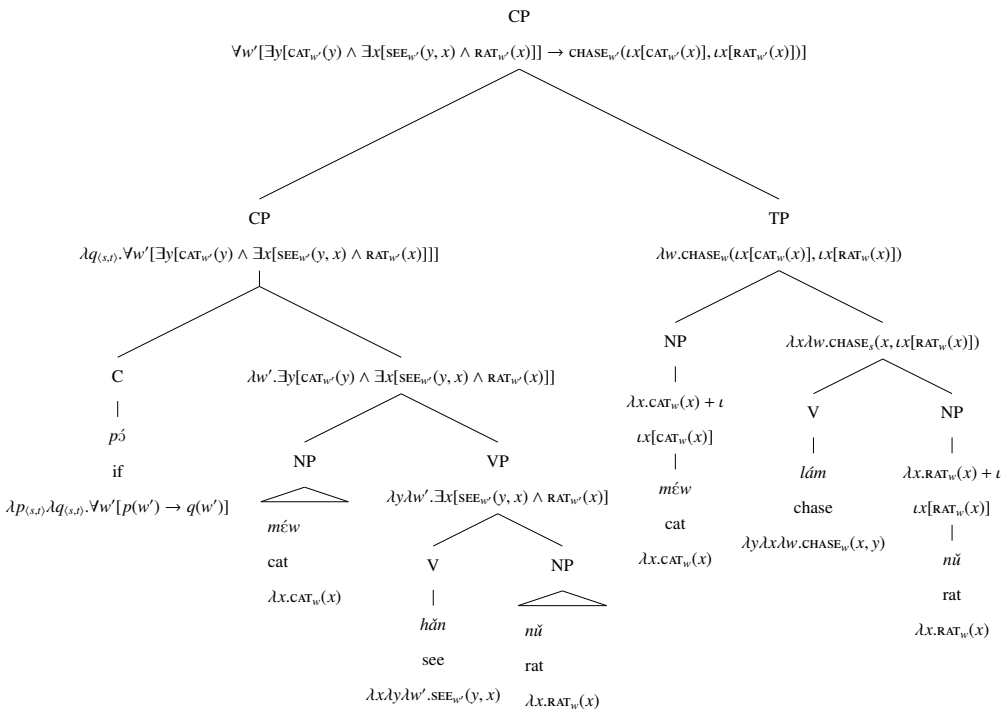
(643) SHAN: CONDITIONAL

mɛ́w lɛ nǔ ʔàm mɛn kǎn. pɔ́ mɛ́w hǎn nǔ nâj-tsɯ̌ŋ, mɛ́w lɯp lám tî
cat and rat NEG right together if cat see rat then cat follow chase grab

njɔ́p nǔ tàasè
snatch rat always

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

(644)



With the combined noun semantics and classifier semantics, the example of donkey anaphora

in (645) involves quantification over situations, dogs in general, and individual cats. This is why the singular definite expression *méw tǒ nân* 'that cat' can be used to refer back to the specific cat in each situation.

(645) Shan: Donkey anaphora

mă̆a nâj hă̆n méw tǒ lă̆j kɔ tě lɯp méw (tǒ nân) tàasè.
dog this see cat clf.anml which prt will follow cat (clf.anml that) always

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

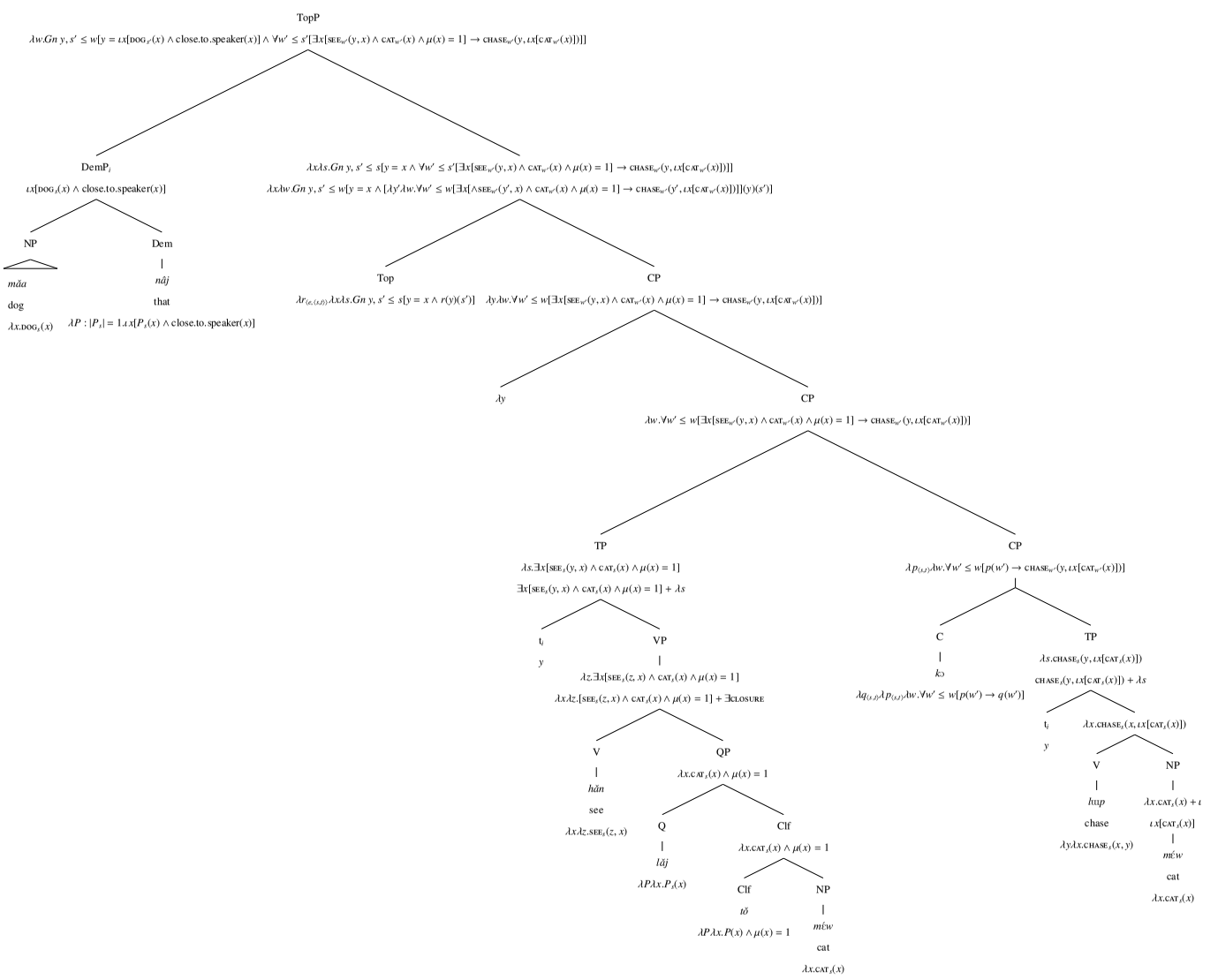
(Moroney 2019a: (20))

(646) (645) $\rightsquigarrow \lambda s.Gn\ y, s' \leq s[y = \iota x[\text{DOG}_{s'}(x) \wedge \text{close.to.speaker}(x)] \wedge \forall w' \leq s'[\exists x[\text{SEE}_{w'}(y, x) \wedge \text{CAT}_{w'}(x) \wedge \mu(x) = 1] \rightarrow \text{CHASE}_{w'}(y, \iota x[\text{CAT}_{w'}(x)])]]$

(647) uses the semantics for topic proposed in a previous section to show the derivation of (645), the complete derived semantics of which is given in (646). This says that for generic situations involving dogs, when there is a sub-situation where the dog or dogs in that situation see a given cat, they will chase that cat. The singular semantics included by the classifier *tǒ* in *méw tǒlă̆j* 'which(ever) cat' means that it is felicitous to include a classifier in the main clause. The derivation here just has *méw* 'cat' in the main clause to refer back to the cat. However, each situation s' will have only one cat, so *méw tǒ nân* 'that cat' is also acceptable.

Here, the position of the topic *mă̆a nâj* 'the dogs' allows it to be bound by generic quantification over situations, since this is a generic/characterizing sentence.

(647)


TopP
λw.Gn y, s′ ≤ w[y = ιx[DOG_{s′}(x) ∧ close.to.speaker(x)] ∧ ∀w′ ≤ s′[∃x[SEE_{w′}(y, x) ∧ CAT_{w′}(x) ∧ μ(x) = 1] → CHASE_{w′}(y, ιx[CAT_{w′}(x)])]]
DemP_i
ιx[DOG_s(x) ∧ close.to.speaker(x)]
NP
Dem
măa
dog
λx.DOG_s(x)
nâj
that
λP : |P_s| = 1.ιx[P_s(x) ∧ close.to.speaker(x)]
λxλs.Gn y, s′ ≤ s[y = x ∧ ∀w′ ≤ s′[∃x[SEE_{w′}(y, x) ∧ CAT_{w′}(x) ∧ μ(x) = 1] → CHASE_{w′}(y, ιx[CAT_{w′}(x)])]]
λxλw.Gn y, s′ ≤ w[y = x ∧ [λy′λw.∀w′ ≤ w[∃x[^SEE_{w′}(y′, x) ∧ CAT_{w′}(x) ∧ μ(x) = 1] → CHASE_{w′}(y′, ιx[CAT_{w′}(x)])]](y)(s′)]
Top
CP
λr_{⟨e,⟨s,t⟩⟩}λxλs.Gn y, s′ ≤ s[y = x ∧ r(y)(s′)]
λyλw.∀w′ ≤ w[∃x[SEE_{w′}(y, x) ∧ CAT_{w′}(x) ∧ μ(x) = 1] → CHASE_{w′}(y, ιx[CAT_{w′}(x)])]
λy
CP
λw.∀w′ ≤ w[∃x[SEE_{w′}(y, x) ∧ CAT_{w′}(x) ∧ μ(x) = 1] → CHASE_{w′}(y, ιx[CAT_{w′}(x)])]
TP
CP
λs.∃x[SEE_s(y, x) ∧ CAT_s(x) ∧ μ(x) = 1]
∃x[SEE_s(y, x) ∧ CAT_s(x) ∧ μ(x) = 1] + λs
λp_{⟨s,t⟩}λw.∀w′ ≤ w[p(w′) → CHASE_{w′}(y, ιx[CAT_{w′}(x)])]
t_i
VP
C
TP
y
kɔ
λq_{⟨s,t⟩}λp_{⟨s,t⟩}λw.∀w′ ≤ w[p(w′) → q(w′)]
λz.∃x[SEE_s(z, x) ∧ CAT_s(x) ∧ μ(x) = 1]
λs.CHASE_s(y, ιx[CAT_s(x)])
λx.λz.[SEE_s(z, x) ∧ CAT_s(x) ∧ μ(x) = 1] + ∃CLOSURE
CHASE_s(y, ιx[CAT_s(x)]) + λs
t_i
λx.CHASE_s(x, ιx[CAT_s(x)])
y
V
QP
V
NP
hăn
see
λx.λz.SEE_s(z, x)
λx.CAT_s(x) ∧ μ(x) = 1
ltɯp
chase
λy.λx.CHASE_s(x, y)
λx.CAT_s(x) + ι
ιx[CAT_s(x)]
mɛ́w
cat
λx.CAT_s(x)
Q
Clf
lâj
λP.λx.P_s(x)
λx.CAT_s(x) ∧ μ(x) = 1
Clf
NP
tŏ
λP.λx.P(x) ∧ μ(x) = 1
mɛ́w
cat
λx.CAT_s(x)

In (648), the indefinite expression *mɛ́w tǒ* nɯŋ 'one cat' introduces the cat with an existential quantifier. For this reason, it is necessary to specify in the anaphoric expression referring back to the cat that there is only one. Otherwise, the cats in situation w might include ones that dog x did not see, making the expression $\iota z[\text{CAT}_w(z)]$ a bad way to refer back to the relevant cat.

(648) SHAN: DONKEY ANAPHORA

mă̌a ku tǒ nâj pɔ́ hă̌n mɛ́w tǒ nɯŋ kój kɔ tě lɯp lám
dog every CLF.ANML this if/when see cat CLF.ANML one only PRT will follow chase

mɛ́w *(tǒ nân) tàasè.
cat *(CLF.ANML that) always

'Every dog, if it sees only one cat, will always chase that cat.' (Moroney 2019a: (21))

(Approximate interpretation:) $\forall w, x[\text{DOG}_w(x) \wedge \mu(x) = 1 \wedge \exists y[\text{CAT}_w(y) \wedge \mu(y) = 1 \wedge \text{SEE}_w(x, y)] \rightarrow \textit{chase}_w(x, \iota z[\text{CAT}_w(z) \wedge \mu(z) = 1])]$

5.5 Chapter summary: Left-located constructions

I have proposed that the constructions that look like internally headed relative clauses in Shan are, in fact, internally headed relative clauses. These IHRC constructions obligatorily appear in left-located positions, which may be because they function to frame the situation in which the matrix clause is evaluated. It seems to be a feature of Shan that clauses that consist of given information must appear at the left side of the clause, and focused information tends to be on the right side of the clause.

I propose that LL constructions are left-dislocations when an overt pronoun must obligatorily co-reference the LL constituent. If we assume that covert object pronouns come from 'topic drop', which involves movement to topic position, then that suggests that some LL constituents are moved.

This chapter has proposed that the function of the left-located position is to restrict the domain of the situation described by the matrix clause. The particle *kɔ* has the function of separating a left-located item that is the restrictor of a universal quantifier over situations. In both cases the left-located construction has the effect of indicating the domain restriction of the clause.

The semantic elements that have been developed over the course of the previous chapters have been brought together here to analyze examples of donkey anaphora in Shan. The semantics of nouns, demonstratives, and classifiers developed over the previous chapters can be brought together with the semantics associated with restriction and quantification over situations to predict the distribution of nominal morphology in these constructions.

This has only skimmed the surface of null arguments and their connection with topicalization and quantification. I noted that the restrictions on null arguments that are co-referential with topicalized nominals does not hold for the donkey anaphora examples. There is more to say about this. The analysis developed here for topics is not consistent with multiple topics, unless those topics are overlapping in reference. Therefore, we would not expect that the multiple null arguments in donkey anaphora examples comes about as a result of multiple topicalization. Perhaps, it is just that the null object is a topic and the subject is a pronoun. The predictions of the analysis developed here with respect to these questions need to be explored and tested further.

CHAPTER 6

CONCLUSION

This dissertation has focused on the syntax and semantics of nominals in Shan, an understudied Tai language of Myanmar. In addition to adding to the body of data on languages by including fieldwork data from an understudied language, this dissertation has emphasized how a careful investigation of the nominal syntax and semantics can give us insight into how reference is constructed by language, both within Shan and cross-linguistically. Chapter 1 gives an overview of the dissertation.

Chapter 2 introduces the main puzzle: how do we account for the variety of interpretations that are available for a particular nominal construction? This chapter focuses on the syntax and semantics of bare nouns. First, I compare Shan bare nouns to bare nouns in Dënesųłiné and bare plurals and mass nouns in English. What we can see there is that bare nominal expressions overlap in certain ways– they allow only low scope existential interpretations and can have be interpreted generically. Additionally, we see that English and Shan bare nouns differ most significantly in two ways: (1) bare nouns in Shan can be used to express definiteness where this requires an overt determiner in English and (2) Shan bare nouns are number neutral where English nouns have a morphologically marked singular and plural distinction. The (c)overtness of nominal morphology has consequences for both the syntactic licensing of arguments and their interpretations. For the semantic interpretation of nouns, I propose to use a hybrid type-shifting and existential/generic closure to account for the distribution of interpretations that are available. I discuss the motivations for saying that a bare noun in Shan is underlying a kind or a predicate and propose to treat them as intensional predicates of type $\langle s, \langle e, t\rangle\rangle$ for derivational simplicity. By assuming a predicate denotation, it is more straightforward for type mismatch to generate a definite or kind interpretation when those are contextually supported. The remainder of the dissertation builds off of the semantics proposed in this account.

Chapter 3 discusses Shan classifiers and what they can tell us about the semantics of count and measure. I present data about several different kinds of classifiers in Shan and propose that the basic semantics associated with the syntactic position of a classifier is one that applies a unit to a noun. By this I mean that it imposes a measure by which nouns can be evaluated in terms of number/amount. This also ties into the semantics of count versus mass nouns in Shan. Count nouns have inherent atomic parts whereas mass nouns do not. We can see this in how stubbornly distributive predicates can combine directly with count nouns but not mass nouns. The addition of a classifier—and therefore a unit for distribution—then allows for distributive predicates to apply to mass nouns. A significant consequence of analyzing classifiers as functions that quantize by unit means that it is possible to propose a combined semantics of count and measure. The difference between the two expressions is whether the thing that is being measured is a count noun or a mass noun. Count nouns are most naturally counted by number whereas mass nouns, due to the fact that they do not have inherent individuation, must be measured by a different scale than numerosity. The count interpretation associated with mass nouns is when a count noun is used to measure the mass noun, bringing in its count semantics.

Chapter 4 discusses the typology of definiteness. It demonstrates that Shan and several other languages that lack definite articles can use bare nouns to express both unique and anaphoric definiteness. Some previous literature had proposed a typology that excludes the possibility of anaphoric bare nouns. The data in this chapter demonstrates that this typology needs to be updated to include bare noun anaphoric definiteness. The chapter also takes up a pragmatic account that says that how easily a nominal expression can uniquely identify its intended referent is important in determining how overtly that nominal expression is specified morphologically. Additionally, this chapter discusses several other factors that contribute to the choice between nominal expressions within a language.

Chapter 5 gives information on left-located constructions in Shan. This chapter demonstrates that left-located constituents in Shan could be topicalized or left-dislocated constructions. These

left-located constituents provide domain restriction for the clause they appear in. The internally headed relative clause constructions are obligatorily left-located constituents. The particle *kɔ* that appears between left-located constituents and the matrix clause separates the restrictor and nuclear scope of universal quantification over situations. Donkey anaphora, which was introduced in the previous chapter, can be analyzed using a combination of all the semantic analyses developed over the previous chapters of the dissertation.

BIBLIOGRAPHY

Ahn, Dorothy. 2019. THAT thesis: A competition mechanism for anaphoric expressions. Harvard University dissertation.

Aikhenvald, Alexandra Y. 2000. Classifiers: A typology of noun categorization devices: A typology of noun categorization devices. OUP Oxford.

Aldridge, Edith. 2017. Internally and externally headed relative clauses in Tagalog. Glossa: a journal of general linguistics 2.

Amfo, Nana Aba Appiah. 2007. Akan demonstratives. Selected proceedings of the 37th annual conference on African linguistics, 134–148. Sommerville: Cascadilla Press.

Arkoh, Ruby, and Lisa Matthewson. 2013. A familiar definite article in Akan. Lingua 123.1–30.

Badan, Linda, and Francesca Del Gobbo. 2011. On the syntax of topic and focus in Chinese. Mapping the left periphery, 63–91.

Baker, Mark C. 2009. Is head movement still needed for noun incorporation? Lingua 119.148–165.

Bale, Alan; Jessica Coon; and Nicholás Arcos. 2019. Classifiers, partitions, and measurements: Exploring the syntax and semantics of sortal classifiers. Glossa: a journal of general linguistics 4.

Barrie, Michael, and Eric Mathieu. 2016. Noun incorporation and phrasal movement. Natural Language & Linguistic Theory 34.1–51.

Bhatt, Rajesh. 1997. Matching effects and the syntax-morphology interface: Evidence from Hindi correlatives. MIT Working Papers in Linguistics 31.53–68.

Bisang, Walter. 1999. Classifiers in East and Southeast Asian languages: Counting and beyond. Numeral types and changes worldwide, 113–185.

Bochnak, M. Ryan, and Lisa Matthewson. 2015. Methodologies in Semantic Fieldwork. Oxford University Press.

Bogal-Allbritten, Elizabeth, and Keir Moulton. 2017. Navajo in the typology of internally-headed relatives. Semantics and Linguistic Theory, ed. by Dan Burgdorf, Jacob Collard, Sireemas Maspong, and Brynhildur Stefánsdóttir, vol. 27, 700–720.

Borer, Hagit. 2005. Structuring sense, vol. I. Oxford, UK: Oxford University Press.

Bošković, Željko, and I-Ta Chris Hsieh. 2012. On word order, binding relations, and plurality within Chinese NPs. Proceedings of the 13th international symposium on Chinese languages and linguistics, National Taiwan Normal University, Department of English Taipei, 19–47.

Bremmers, David; Jianan Liu; Martijn van der Klis; and Bert Le. 2019. Definiteness across Languages: From German to Mandarin. The 22nd Amsterdam Colloquium.

Büring, Daniel. 2003. On D-trees, beans, and B-accents. Linguistics and philosophy 26.511–545.

Carlson, Greg N. 1977. A unified analysis of the English bare plural. Linguistics and Philosophy 1.413–457.

Carlson, Greg N., and Francis Jeffry Pelletier (eds.) 1995. The generic book. Chicago: University of Chicago Press.

Carlson, Gregory N. 1989. The semantic interpretation of generic sentences. Property theory, type theory, and semantics, ed. by Gennaro Chierchia, Barbara H Partee, and Raymond Turner. Dordrecht, Netherlands: Kluwer.

Chao, Yuen Ren. 1968. A Grammar of Spoken Chinese. Berkeley: University of California Press.

Cheng, Lisa L-S, and Rint Sybesma. 2005. Classifiers in four varieties of Chinese. Handbook of comparative syntax, 259–292.

Cheng, Lisa Lai-Shen, and Rint Sybesma. 1999. Bare and Not-So-Bare Nouns and the Structure of NP. Linguistic Inquiry 30.509–542.

Chierchia, Gennaro. 1998. Reference to Kinds across Language. Natural Language Semantics 6.339–405.

Chierchia, Gennaro. 2010. Mass nouns, vagueness and semantic variation. Synthese 174.99–149.

Chomsky, Noam. 1977. On wh-movement. 1977 65.

Chung, Sandra, and William A. Ladusaw. 2003. Restriction and Saturation. MIT Press.

Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its exceptions. Linguistic Inquiry 36.315–332.

Cole, Peter. 1987. The structure of internally headed relative clauses. Natural Language & Linguistic Theory 5.277–302.

Corver, Norbert, and Marjo van Koppen. 2011. NP-ellipsis with adjectival remnants: A micro-comparative perspective. Natural Language & Linguistic Theory 29.371–421.

Cushing, Josiah. 1887. Grammar of the Shan Language. Rangoon: American Baptist Mission Press.

Cushing, Josiah Nelson. 2000. Cushing's Shan-English dictionary: A phonetic version. Tempe, Ariz.: Arizona State University, Program for Southeast Asian Studies, Monograph Series Press.

Dayal, Veneeta. 1996. Locality in wh-quantification: Questions and relative clauses in Hindi.

Dayal, Veneeta. 2004. Number Marking and (in)Definiteness in Kind Terms. Linguistics and Philosophy 27.393–450.

Dayal, Veneeta. 2011. Bare noun phrases. Semantics: An International Handbook of Natural Language Meaning 33.1087–1108.

Dayal, Veneeta. 2012. Bangla classifiers: Mediating between kinds and objects. Rivista di Linguistica 24.195–226.

Dayal, Veneeta. 2014. Bangla plural classifiers. Language and Linguistics 15.47–87.

Dayal, Veneeta, and Julie Jiang. to appear. The Puzzle of Anaphoric Bare Nouns in Mandarin: A Counterpoint to Index! Linguistic Inquiry.

Deal, Amy Rose. 2017. Countability distinctions and semantic variation. Natural Language Semantics 25.125–171.

Deal, Amy Rose, and Julia Nee. 2018. Bare nouns, number, and definiteness in Teotitlán del Valle Zapotec. Proceedings of Sinn und Bedeutung, vol. 21, 317–334.

Despić, Miloje. 2011. Syntax in the absence of determiner phrase. University of Connecticut dissertation.

Despić, Miloje. 2019. On kinds and anaphoricity in languages without definite articles. Definiteness across languages 25.259.

Destruel, Emilie; David Beaver; and Elizabeth Coppock. 2018. Clefts: Quite the contrary! Proceedings of Sinn und Bedeutung, vol. 21, 335–346.

Diesing, Molly. 1992. Bare plural subjects and the derivation of logical representations. Linguistic Inquiry 23.353–380.

Doetjes, Jenny Sandra. 1997. Quantifiers and selection. On the distribution of quantifying expressions in French, Dutch and English. HAG.

Dong, Sicong, and Sam Yin Wong. 2020. Haplology and lexical entries: A study based on cross-linguistic data from Sinitic languages. Lexicography 7.59–77.

Doughty, Catherine. 1991. Second language instruction does make a difference: Evidence from an empirical study of SL relativization. Studies in second language acquisition 13.431–469.

Downing, Bruce T. 1978. Some universals of relative clause structure. Universals of human language 4.375–418.

Eberhard, David M.; Gary F. Simons; and Charles D. Fennig. 2021. Shan. Ethnologue: Languages of the World.

Fernández-Sánchez, Javier, and Dennis Ott. 2020. Dislocations. Language and Linguistics Compass 14.e12391.

Fodor, Janet Dean, and Ivan A Sag. 1982. Referential and quantificational indefinites. Linguistics and philosophy 5.355–398.

Frege, Gottlob. 1892. über Sinn und Bedeutung [On sense and reference]. Zeitschrift für Philosophic und philosophische Kritik 100.25–50.

Gil, David. 2000. Riau Indonesian: A VO language with internally-headed relative clauses. Snippets 1.6–7.

Givón, Talmy. 2017. The Story of Zero. Amsterdam: John Benjamins Publishing Company.

Gorbet, Larry. 1974. Relativization and complementation in Diegueño. San Diego.

Greenberg, Joseph H. 1972. Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. Working Papers on Language Universals 9.1–39.

Grimm, Scott. 2012. Number and individuation. Standford University dissertation.

Grosu, Alexander. 2010. The status of the internally-headed relatives of Japanese/Korean within the typology of definite relatives. Journal of East Asian Linguistics 19.231–274.

Grosu, Alexander, and Koji Hoshi. 2019. Japanese internally-headed and doubly-headed relative constructions, and a comparison of two approaches. Glossa: a journal of general linguistics 4.128.

Grosu, Alexander, and Fred Landman. 2012. A quantificational disclosure approach to Japanese and Korean internally headed relatives. Journal of East Asian Linguistics 21.159–196.

Haegeman, Liliane. 2003. Conditional clauses: External and internal syntax. Mind & Language 18.317–339.

Halvorsen, Per-Kristian Giaever. 1977. Syntax and Semantics of Cleft Constructions. United States – Texas: The University of Texas at Austin Ph.D.

Hankamer, Jorge, and Line Mikkelsen. 2002. A morphological analysis of definite nouns in Danish. Journal of Germanic Linguistics 14.137–175.

Harley, Heidi. 2009. Compounding in distributed morphology. na.

Hawkins, John A. 1978. Definiteness and indefiniteness: A study in reference and grammaticality prediction. Routledge.

Heim, Irene. 1982. The semantics of definite and indefinite noun phrases.

Heim, Irene. 1990. E-type pronouns and donkey anaphora. Linguistics and philosophy 13.137–177.

Heim, Irene. 1991. Artikel und definitheit. Semantik: ein internationales Handbuch der zeitgenössischen Forschung, 487–535.

Heim, Irene. 1998. Semantics in generative grammar. Malden, Mass.: Blackwell.

Heusinger, Klaus Von. 2019. Indefiniteness and Specificity. http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199687305.001.0001/oxfordhb-9780199687305-e-9.

Heycock, Caroline, and Anthony Kroch. 1999. Pseudocleft connectedness: Implications for the LF interface level. Linguistic inquiry 30.365–397.

Hiraiwa, Ken. 2005. Dimensions of symmetry in syntax: Agreement and clausal architecture. Massachusetts Institute of Technology dissertation.

Hoonchamlong, Yuphaphann. 1992. Some issues in Thai anaphora: A government and binding approach.

Hoshi, Koji. 1995. Structural and interpretive aspects of head-internal and head-external relative clauses. University of Rochester dissertation.

Huang, C-T James. 1984. On the distribution and reference of empty pronouns. Linguistic inquiry, 531–574.

Huang, Nick. 2020. "Noun-less" nominal expressions in Mandarin Chinese.

Hundius, Harald, and Ulrike Kölver. 1983. Syntax and semantics of numeral classifiers in Thai. Studies in Language. International Journal sponsored by the Foundation "Foundations of Language" 7.165–214.

Ionin, Tania, and Ora Matushansky. 2006. The composition of complex cardinals. Journal of semantics 23.315–360.

Irani, Ava. 2016. Two types of definites in American Sign Language.

Irani, Ava. 2019. On (in) definite expressions in American Sign Language. Studies in Diversity Linguistics.

Iwasaki, Shoichi, and Preeya Ingkaphirom. 2005. A Reference Grammar of Thai. Cambridge University Press.

Jenks, Peter. 2011. The hidden structure of Thai noun phrases. Cambridge, MA: Harvard University dissertation.

Jenks, Peter. 2012. Three ways to modify classifiers in Thai. 86th Meeting of the LSA, January.

JENKS, PETER. 2013. Quantifier Float, Focus, and Scope in Thai. Annual meeting of the Berkeley linguistics society, vol. 39, 90–107.

JENKS, PETER. 2015. Two kinds of definites in numeral classifier languages. Semantics and Linguistic Theory 25.103–124.

JENKS, PETER. 2018. Articulated definiteness without articles. Linguistic Inquiry 49.501–536.

JIANG, LI. 2012. Nominal arguments and language variation. Unpublished doctoral dissertation, Department of Linguistics, Harvard University, Cambridge, Massachusetts.

JIANG, LI JULIE. 2018. Definiteness in Nuosu Yi and the theory of argument formation. Linguistics and Philosophy 41.1–39.

KADMON, NIRIT. 1987. On unique and non-unique reference and asymmetric quantification. University of Massachusetts, Amherst dissertation.

KADMON, NIRIT. 1990. Uniqueness. Linguistics and Philosophy 13.273–324.

KEENAN, EDWARD L, and BERNARD COMRIE. 1977. Noun phrase accessibility and universal grammar. Linguistic inquiry 8.63–99.

KIM, KYUMIN, and PAUL B MELCHIN. 2018. On the complementary distribution of plurals and classifiers in East Asian classifier languages. Language and Linguistics Compass 12.

KITSOMBAT, PORNTHIP. 1981. [The usage of thii, sin and \textipa{P}an]. Master's thesis, Chulalongkorn University.

KIZU, MIKA. 2005. Cleft constructions in Japanese syntax. Springer.

KRIFKA, MANFRED. 1995. Common nouns: A contrastive analysis of English and Chinese. The generic book, ed. by Gregory N. Carlson and Francis J. Pelletier, 398–411. Chicago University Press.

Krifka, Manfred, and Renate Musan. 2012. Information structure: Overview and linguistic issues. The expression of information structure 5.1–44.

Kuroda, S-Y. 1974. Pivot-independent relativization in Japanese (I). Journal of Japanese Linguistics 3.61–96.

Landman, Fred. 2011. Count nouns-mass nouns, neat nouns-mess nouns. Baltic International Yearbook of Cognition, Logic and Communication 6.12.

Landman, Fred. 2016a. Iceberg semantics for count nouns and mass nouns: The evidence from portions. The Baltic International Yearbook of Cognition Logic and Communication 11.1–48.

Landman, Fred. 2016b. Japanese internally headed relatives: A hybrid analysis with Kuroda functions. Glossa 1.1.

Langdon, Margaret. 1977. Syntactic change and SOV structure: The Yuman case. Mechanisms of syntactic change, 2255–90.

Lee, Han-Gyu. 1992. The pragmatics and syntax of pragmatic morphemes in Korean. University of Illinois at Urbana-Champaign dissertation.

Li, Xu-Ping. 2011. On the semantics of classifiers in Chinese. Bar Ilan University dissertation.

Li, XuPing, and Walter Bisang. 2012. Classifiers in Sinitic languages: From individuation to definiteness-marking. Lingua 122.335–355.

Li, XuPing, and Susan Rothstein. 2012. Measure readings of Mandarin classifier phrases and the particle de. Language and Linguistics 13.693.

Lima, Suzi. 2014. The Grammar of Individuation and Counting. University of Massachusetts Amherst dissertation.

LINK, GODEHARD. 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. Meaning, Use and Interpretation of Language, ed. by R. Bäuerle, C. Schwarze, and A. von Stechow. Berlin: de Gruyer.

LITTLE, CAROL-ROSE; MARY MORONEY; and JUSTIN ROYER. to appear. Classifying classifiers. Proceedings of Semantics of Under-Represented Languages in the Americas 11 (SULA).

LITTLE, CAROL-ROSE, and EKARINA WINARTO. 2019. Classifiers and the definite article in Indonesian. Proceedings of the 49th North East Linguistics Society, ed. by Maggie Baird and Jonathan Pesetsky, vol. 2, 209–220. Amherst, MA: GLSA.

LOBECK, ANNE C. 1995. Ellipsis: Functional heads, licensing, and identification. Oxford University Press on Demand.

LÖBNER, SEBASTIAN. 1985. Definites. Journal of Semantics 4.279–326.

LÜ, SHUXIANG. 1980. Xiandai hanyu babai ci [Eight hundred words in modern Chinese]. Beijing: Shangwu Yinshuguan.

LÜ, SHUXIANG. 1990. Ge zi de yingyong fanwei, fu lun danweici qian yi zi de tuoluo [Scope of the uses of ge and omission of yi in front of classifiers]. Lü Shuxiang Wenji [Collected Works of Lü Shuxiang] 2.144–175.

MENDIA, JON ANDER. 2020. Reference to ad hoc kinds. Linguistics and Philosophy 43.589–631.

MENDIA, JON ANDER, and HANA FILIP. inprep. Characterizing Genericity and Epistemic Commitments. 36.

MILSARK, GARY. 1974. Existential sentences in English. Cambridge, Massachusetts: Massachusetts Institute of Technology dissertation.

MOENG, SAO TERN. 1995. Shan-English Dictionary. Dunwoody Press.

Moltmann, Friederike. 2004. Two kinds of universals and two kinds of collections. Linguistics and Philosophy 27.739–776.

Moroney, Mary. 2019a. Definiteness with Bare Nouns in Shan. At the Intersection of Language, Logic, and Information, ed. by Jennifer Sikos and Eric Pacuit, Lecture Notes in Computer Science, 108–123. Berlin, Heidelberg: Springer.

Moroney, Mary. 2019b. Inconsistencies of the Consistency test. Proceedings of the Linguistic Society of America 4.55–1–10.

Moroney, Mary. 2019c. Interpreting Quantification within Internally Headed Relative Clauses. Proceedings of the 12th Generative Linguistics in the Old World (GLOW) in Asia and the 21st Seoul International Conference on Generative Grammar: Universal Grammar and Its Cross-linguistic Instantiations, 359–368. The Korean Generative Grammar Circle.

Moroney, Mary. 2021. Updating the typology of definiteness: Evidence from bare nouns in Shan. Glossa: a journal of general linguistics 6.1–28.

Nomoto, Hiroki. 2013. Number in classifier languages. University of Minnesota dissertation.

Partee, Barbara H, and Vladimir Borschev. 2012. Sortal, relational, and functional interpretations of nouns and Russian container constructions. Journal of Semantics 29.445–486.

Paul, Waltraud, and John Whitman. 2017. Topic prominence. The Wiley Blackwell Companion to Syntax, Second Edition, 1–31.

Percus, Orin. 2000. Constraints on some other variables in syntax. Natural language semantics 8.173–229.

Piriyawiboon, Nattaya. 2010. Classifiers and determiner-less languages: The case of Thai. University of Toronto Doctoral dissertation dissertation.

Postal, Paul M. 1969. Anaphoric islands. Proceedings from the Annual Meeting of the Chicago Linguistic Society, vol. 5, 205–239. Chicago Linguistic Society.

Prince, Ellen. 1998. On the limits of syntax, with reference to topicalization and left-dislocation. The limits of syntax 29.

Roberts, Craige. 2003. Uniqueness in Definite Noun Phrases. Linguistics and Philosophy 26.287–350.

Ross, John R. 1982. Pronoun deleting processes in German. Presented at the annual meeting of the Linguistic Society of America, San Diego, California.

Ross, John Robert. 1967. Constraints on variables in syntax.

Rothstein, Susan. 2016. Counting and measuring: A theoretical and crosslinguistic account. Baltic International Yearbook of Cognition, Logic and Communication 11.8.

Rothstein, Susan. 2017. Semantics for Counting and Measuring. Key Topics in Semantics and Pragmatics. Cambridge: Cambridge University Press.

Royer, Justin. 2019. Domain restriction and noun classifiers in Chuj (Mayan). Proceedings of the Forty-Ninth Annual Meeting of the North East Linguistic Society, vol. 3, 87–97.

Russell, Bertrand. 1905. On Denoting. Mind 14.479–493.

Saito, Mamoru, and Keiko Murasugi. 1990. N'-deletion in Japanese: A preliminary study. Japanese/Korean Linguistics 1.285–301.

Sauerland, Uli; Jan Anderssen; and Kazuko Yatsushiro. 2005. The plural is semantically unmarked. Linguistic evidence: Empirical, theoretical, and computational perspectives, 413–434.

Schwarz, Florian. 2009. Two types of definites in natural language. University of Massachusetts, Amherst dissertation.

Schwarz, Florian. 2013. Different types of definites crosslinguistically. Language and Linguistics Compass 7.534–559.

Schwarz, Florian. 2019. Weak vs. strong definite articles: Meaning and form across languages. Studies in Diversity Linguistics.

Schwarzschild, Roger. 2007. Stubbornly distributive predicates. Talk delivered at UMass Amherst, December 7.

Schwarzschild, Roger. 2011. Stubborn distributivity, multiparticipant nouns and the count/mass distinction. Proceedings of NELS, vol. 39, 661–678. Graduate Linguistics Students Association, University of Massachusetts.

Sennet, Adam. 2016. Ambiguity. The stanford encyclopedia of philosophy, ed. by Edward N. Zalta, spring 2016 edn. Metaphysics Research Lab, Stanford University.

Shee, Naw Hsar. 2008. A Descriptive Grammar of Geba Karen. Master's thesis, Payap University.

Shimoyama, Junko. 1999. Internally headed relative clauses in Japanese and E-type anaphora. Journal of East Asian Linguistics 8.147–182.

Simpson, Andrew. 2005. Classifiers and DP structure in Southeast Asia. The Oxford handbook of comparative syntax, 806–838.

Simpson, Andrew. 2011. Floating quantifiers in Burmese and Thai. JSEALS 4.115–146.

Simpson, Andrew. 2017. Bare classifier/noun alternations in the Jinyun (Wu) variety of Chinese and the encoding of definiteness. Linguistics 55.305–331.

Simpson, Andrew, and Priyanka Biswas. 2016. Bare nominals, classifiers, and the representation of definiteness in Bangla. Linguistic Analysis 40.167–198.

Simpson, Andrew, and Binh Ngo. 2018. Classifier syntax in Vietnamese. Journal of East Asian Linguistics 27.211–246.

Simpson, Andrew; Hooi Ling Soh; and Hiroki Nomoto. 2011. Bare classifiers and definiteness: A cross-linguistic investigation. Studies in Language. International Journal sponsored by the Foundation "Foundations of Language" 35.168–193.

Sio, Joanna. 2006. Modification and reference in the Chinese nominal. The Leiden University Centre for Linguistics (LUCL), Faculty of Arts dissertation.

Srinivas, Sadhwi, and Kyle Rawlins. 2020. Definiteness and the bare nominal in Kannada. Talk presented at The 94th Annual Meeting of the Linguistic Society of America (LSA).

Strawson, P. F. 1950. On Referring. Mind 59.320–344.

Tellier, C. 1989. Head-internal relatives and parasitic gaps in Moor\'e 6.298–318.

Ueda, Yasuki, and Tomoko Haraguchi. 2008. Plurality in Japanese and Chinese. Nanzan Linguistics: Special Issue 3.229–242.

Von Fintel, Kai, and Irene Heim. 2011. Intensional semantics.

Wespel, Johannes. 2008. Descriptions and their domains: The patterns of definiteness marking in French-related creoles. Stuttgart: University of Stuttgart dissertation.

Wilbur, Ronnie. 2017. Internally-headed relative clauses in sign languages. Glossa: a journal of general linguistics 2.

Wilhelm, Andrea. 2008. Bare nouns and number in Dëne Sųłiné. Natural Language Semantics 16.39–68.

Wilhelm, Andrea. 2014. A radically [+ arg,-pred] language. Semantics of Under-Represented Language in the Americas (SULA) 8.

Yang, Rong. 2001. Common nouns, classifiers, and quantification in Chinese. Rutgers, The State University of New Jersey dissertation.