**The Syntax of Bora Subject Clitics: Anaphora and Long Distance Binding**

by

Marcus Paul Berger

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Linguistics)
in the University of Michigan
2019

Doctoral Committee:

Professor Acrisio Pires, Chair
Professor Marlyse Baptista
Assistant Professor Savithry Namboodiripad
Associate Professor Teresa Satterfield

Marcus Paul Berger

marcberg@umich.edu

ORCID iD: 0000-0002-9501-4565

## DEDICATION

Dedicado a los hablantes del idioma Bora, en especial para todas las personas que directamente compartieron su idioma y su cultura conmigo, y en memoria de uno ya fallecido.

## ACKNOWLEDGMENTS

In writing an acknowledgment section, I run the risk of failing to thank everyone who has been instrumental to the completion of this dissertation. To be sure, this would not have been possible without the guidance and mentorship of my advisor, Dr. Acrisio Pires. Through the entire process of planning and carrying out fieldwork, to finishing writing the dissertation, he has worked tirelessly to make sure that I had all the support necessary to succeed. Thanks also to my other committee members, Dr. Marlyse Baptista, Dr. Teresa Satterfield, and Dr. Savi Namboodiripad for their comments and guidance. Special thanks are due as well to Dr. Sam Epstein, whose dedication and support were central to the outcome of the project and the final product.

I owe a deep debt of gratitude as well to Dr. Lev Michael and Dr. Christine Beier, who provided me with my first opportunity to conduct fieldwork in South America. The experience and connections I made while working with them directly benefitted the project that became this dissertation, and I am truly thankful to them for it.

I am grateful for comments I have received through the years that have formed the ideas in this work, especially to the Syntax-Semantics group at the University of Michigan. Thanks as well to Dr. Frank Seifart, whose expertise on the Bora languages helped inform my own understanding of the Bora language.

Fieldwork for this dissertation was funded through the University of Michigan Linguistics Department and the Rackham International Research Award, for which I am very grateful. I am also indebted to Alfonso López Vela and Martha Luz Coral de López for inviting me into their home, and caring for me as if I were one of their own children. They have truly become a second family to me.

Graduate school can be a trying time, and I am thankful for all the emotional support I have received over the years. Thanks to my fellow graduate students for always being there to listen, to Kim Waters for our weekly chats that helped keep me sane, and to my parents, siblings, and the rest of my family, for always believing in me through the years, even when I did not.

Finally, this dissertation would truly never have been possible without those speakers of the Bora language who shared their language, culture, and stories with me. I am in their debt, and this dissertation is dedicated to them. *Tehdújuco*!

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF ABBREVIATIONS

| | |
|---|---|
| 1 | 1st person |
| 2 | 2nd person |
| 3 | 3rd person |
| 3COR | 3rd person coreferent |
| ABL | Ablative |
| ABS | Absolutive |
| ACC | Accusative |
| AGR | Agreement |
| ALL | Allative |
| AML | Animal gender |
| AN | Animate |
| AUG | Augmentative |
| CAUS | Causative |
| CF | Counterfactual |
| CL | Classifier |
| CMP | Comparative |
| CON | Connective pronoun |
| COND | Conditional |
| CONJ | Conjunction |
| COP | Copula |
| DAT | Dative |
| DEM | Demonstrative |
| DIM | Diminutive |
| DU | Dual |
| EMPH | Emphatic |
| ERG | Ergative |
| EVID | Evidential |
| FOC | Focus |
| FRS | Frustrative |
| FUT | Future tense |
| GEN | Genitive |
| HAB | Habitual |
| IMP | Imperative |
| IN | Inanimate |
| IND | Indicative |
| INS | Instrumental |
| IRREAL | Irrealis |
| LOC | Locative |
| LOG | Logophor |
| M | Male |
| MTR | Multiple action transitive |
| NEG | Negative |
| NMLZ | Nominalizer |
| NOM | Nominative |
| NWIT | Non-witnessed |
| OBL | Oblique |
| PL | Plural |
| PN | Pronoun |
| POSS | Possessive |
| PRED | Predicative |
| PRES | Present tense |
| PRF | Perfect |
| PRT | Particle |
| PST | Past tense |
| PUR | Purpose |
| QUOT | Quotative |
| REC | Recent past |
| RECIP | Reciprocal |
| REFL | Reflexive |
| REM | Remote past |
| REP | Repeated action |
| RPT | Reportative |
| SAP | Speech act participant |
| SE | Simplex element anaphor |
| SG | Singular |
| SIN | Single action intransitive |
| SOC | Sociative |
| STR | Single action transitive |
| SUB | Subject |
| SUBJ | Subjunctive |
| VAR | Variety |
| Y/N | Yes/no question |

# ABSTRACT

This dissertation analyzes subject pronouns in Bora, an endangered indigenous language of the Amazon. Bora uses a series of overt subject clitics to express the subject of clauses, which indicate cross-clausal coreference in many cases. Using data collected during personal fieldwork trips, I investigate the distribution of these clitics, and analyze their properties with respect to theories of control and binding. This accomplishes important research goals of (i) performing research on the Bora language, especially given its status as an endangered language, and (ii) using data from understudied languages to inform formal linguistic theory.

After establishing a basic analysis of Bora syntactic structure, I review literature on binding theory and control structures, with the goal of showing how and whether the Bora data fit into existing linguistic theory. Regarding binding, I review canonical binding theory in generative grammar, as well as an alternative proposal that relies on reflexivity as a means of licensing anaphora. Regarding control clauses, I review analyses of PRO serving as a null, or in exceptional cases, an overt instantiation of a controlled subject. I compare this analysis with a theory of control that dispensed with PRO and instead analyzed control clauses as the result of syntactic movement.

I conclude that, although Bora subject clitics appear in embedded clauses that would constitute control in other languages, such clauses in Bora are not control structures. Rather, every clause is a finite clause with an overt subject, with very few exceptions. I go on to show that overt subject clitics have anaphoric properties. For 3rd person subjects, an embedded proclitic *i*= (3COR) indicates coreference with the structurally next highest clause subject. The Speech Act Participant (SAP) clitic *me*= attaches to any verb with a 1st or 2nd person non-singular subject, and also indicates coreference with a higher clause subject when it appears without an overt subject NP.

While the 3COR marker's antecedent must appear in a higher clause, the SAP can take either a local or non-local antecedent, depending on its coreferent properties. The non-local nature of the antecedents of the 3COR marker and, in some cases, the SAP suggest that they constitute instances

of long distance anaphora. This differs from the properties that I show for the 1st and 2nd person singular proclitics, which I argue to have properties of pronominals.

After comparing the Bora data to instances of long distance anaphora in other languages, I determine that the Bora data has many similarities to these other languages. I argue that the ability of some of these clitics to be bound by non-local antecedents is similar to analyses of long distance anaphora in other languages. I first establish that Bora proclitics share properties with other types of long distance anaphors (being monomorphemic, occurring in restricted environments, having subject antecedents, and being subject to a Blocking Effect). I then provide an analysis based on similarity with Mandarin Chinese whereby the Bora anaphoric proclitics undergo raising at Logical Form in order to be bound by their antecedent, which must be the structurally next highest c-commanding subject in the sentence.

# CHAPTER I

## Introduction

This dissertation aims to investigate the syntax of co-reference in the Bora language, an endangered Amazonian language with very few speakers remaining, most of whom live in Peru, though some live in Colombia. This dissertation especially considers overt subject proclitics found in this language. Bora uses these clitics to indicate the subject of a wide range of clauses.[1] With very few exceptions, these clitics are obligatorily overt, and in some cases co-occur with another overt realization of the subject in the same clause. The syntactic properties of these subject clitics have not been explored in significant detail in the already sparse literature on the Bora language.

This dissertation aims at determining the syntactic status of these clitics. In doing so, I analyze what properties these clitics represent when they are used to indicate coreference. I also analyze whether the appearance of these clitics in embedded complement clauses meets criteria for obligatory control. I go on to determine whether the clitics are overt realizations of pronouns[2] (anaphors or pronominals), obligatory control subjects, or whether they are a type of agreement marker. Finally, arguing that these clitics are in fact instantiations of pronominals and anaphors (depending on their person and number features), I analyze whether the clitics are beholden to binding theory.

In the chapters that follow, I explore the properties of these clitics with respect to control and binding theories. The fact that the clitics in question appear in both main and embedded clauses

[1] The clitic status of these subjects has been assumed throughout the literature on Bora, especially by Thiesen and Weber (2012). I review their reasoning for this classification in Chapter 2.

[2] For the remainder of this work, I specifically refer to anaphors and pronominals separately from pronouns. Other works on binding theory occasionally use the terms ‘pronoun’ and ‘pronominal’ interchangeably. To avoid confusion, I take pronouns to be a class of words that includes both anaphors and pronominals.

(depending on many factors, such as the phi-features of the subject, whether the clitic is coreferent with a subject in a higher clause, etc.) raises the question of how Bora manifests a language-specific requirement for overt subjects in most environments where PRO subjects are typically argued to occur in other languages. Notably, Bora does not seem to allow PRO subjects (or null subjects generally, with few exceptions), but instead uses the same subject proclitics in cases where the identity of the subject is the same as that of the higher clause, showing co-reference across clauses. Additionally, the fact that these clitics often (sometimes necessarily) indicate coreference with a higher clause subject is suggestive of a type of anaphora. As a result, I also explore theories of anaphora and whether they are applicable to these subject clitics.

The overall goals of this dissertation are twofold: to determine the distribution and the empirical and theoretical status of preverbal subject proclitics in Bora, and to use these results to advance linguistic theory generally. The importance of contributing empirical data from a significantly understudied language to the discussion of theoretical issues in syntax is central to this project. Much of the literature on co-reference, binding, and control, especially much of the foundational work, is based on English and a few other widely attested languages, though further research continues to add analyses of new languages within these theories. Bora utilizes multiple typologically interesting phenomena that differ from some of the properties of other languages, especially Romance and Germanic languages. Many of these properties of Bora are detailed throughout Chapter 2. While the Bora data do not necessarily call into question the correctness of different theories, the data provide new evidence from an understudied language that can be used to apply linguistic theories on a more universal scale, or else at least determine how these theories might be modified in order to better develop our understanding of the human language faculty and human language knowledge.

### 1.1. Ethnography of Bora

The Bora language is an endangered language of the Boran family, with roughly 1,350 speakers remaining, most of whom live in Peru, mainly in the communities of Puca Urquillo and Brillo Nuevo, though there are some living in the city of Iquitos, as well as in southern Colombia (Eberhard, Simons, & Fennig 2019). The community is undergoing a shift to Spanish, and the Bora language is at a critical state for documentation. Most children of Bora speakers currently grow up speaking only Spanish, and not learning their native Bora.

The Bora language was first attested in 1820 (Martius 1867), when a group of Bora speakers was found on the Caquetá River. Later that century, the Boras, as well as many other indigenous groups in the area, were exploited by colonizers to work in the rubber industry. In exchange for long hours of hard work in the harsh jungle environment, the workers were given goods, including blankets, machetes, etc., as well as guaranteed protection from other groups. Called a patron system, this system continually indebted the Bora people to the 'patrons' whom they worked for, and over time, degraded many of the cultural traditions of the Bora people, and increased the contact with Spanish speakers (Berger & Seifart, forthcoming).

Bora is a language in the Bora language family (Berger & Seifart, forthcoming). Although Bora was previously thought to pertain to the Witotoan family (Aschmann 1993), more recent work has shown that there is not evidence to establish a genetic linguistic relation between the Bora family and the Witotoan family (Seifart & Echeverri 2015, Echeverri & Seifart 2015, Vengoechea 2015). Bora itself has a closely related dialectal variant, Miraña, of about 400 speakers (Seifart 2005), spoken in Colombia along the Caquetá River. The only other Boran language, Muinane, had only 175-200 as of 1997 (Walton et al. 1997). The number of speakers of each of the Boran languages continues to decline as the speakers shift to Spanish.

The Bora language is relatively understudied. The earliest linguistic account was done by Martius (1867). After that, a significant amount of work was carried out by a husband and wife missionary team, Wesley and Eva Thiesen, from 1952-1998. Their work produced a grammar of the language in Spanish (Thiesen 1996), as well as a dictionary of the language (Thiesen & Thiesen 1998). Wesley Thiesen later partnered with David Weber to create a more thorough grammar of the language, focusing heavily on the tonal system of the language (Thiesen & Weber 2012). Frank Seifart has also published extensively about both Miraña (Seifart 2005) and Bora (Seifart 2010, 2015a, 2015b, and other work). There has been no work on the Bora language (or any of its related languages or dialects) that focuses on a theoretical analysis of binding or control in Bora, outside of descriptive and typological work.

### 1.2. Data Collection and Methodology

Unless otherwise cited, the Bora data in this dissertation were collected via my personal fieldwork with Bora speakers in Peru, during the course of three trips to Iquitos, Peru between 2015-2018. The data were collected from four native speakers of Bora, all of whom were bilingual in Spanish,

which was used as the metalanguage for my fieldwork. One speaker grew up speaking Bora, and learned Spanish later in life, while the other three grew up as simultaneous bilinguals in Spanish and Bora. One speaker also speaks Huitoto (Witotoan), another indigenous language of the region that is in heavy contact with Bora in the community of Puca Urquillo.

The primary modes of data collection were sentence elicitation and narratives. Elicited sentences included structures that were identified before each session, focusing on syntactic constructions that included the subject proclitics under investigation. During one-on-one sessions with the speakers, I provided them with sentences orally in Spanish and asked them to give the equivalent sentence in Bora. After going through one set of sentences, usually thirty at a time, I repeated the sentences back to them in Bora, asking for a back translation into Spanish to ensure that the original meaning of the sentence had been understood and maintained through translation. Many of these sentences targeted different types of embedded clauses, in which I tested for both co-reference and disjoint reference between the main and embedded clause subjects.

There are many problems associated with attempting to collect negative data in fieldwork. When asking for the grammaticality of a sentence, some sentences may be too difficult for a speaker to parse, some sentences may be judged inconsistently depending on the context in which they are presented, or sometimes a naïve speaker may not be able to grasp specific grammatical distinctions that a trained linguist may be trying to elicit. These and other reasons are outlined in Chelliah (2001). For these reasons, I do not include many examples of negative Bora data, and the negative data that is presented is simple in nature and was judged to be ungrammatical in the same way by multiple speakers. While the presence of negative data may lend more credence to the syntactic arguments being presented, I feel comfortable that the quality of the positive data I have collected is more conducive to making theoretical linguistic contributions.

The narratives I collected were of two main types: guided and free narratives. The guided narratives involved some kind of interactive task that was designed to collect data from all speakers in the same environment. Some of these tasks were also targeting specific grammatical phenomena. These included, for example, having the speakers describe a word to another speaker in Bora without using that word, a task which was partially aimed at eliciting relative clauses and nominal classifiers. The guided narratives also included a set of three picture stories: *Frog Where Are You, A Boy a Dog and a Frog*, and *A Boy a Dog a Frog and a Friend*, (Mayer 1969, 1978,

Mayer & Mayer 1992). In this task, the speakers were presented with the pictures in each of the three picture books and asked to tell the story of what was happening in the pictures. I also conducted a sociolinguistic interview with each of the speakers, which consisted of demographic questions and questions about Bora culture and daily life.

I also had each of the speakers give free narratives, sometimes with a prompt (e.g. talk about some advice that you would give to your children or grandchildren on how to lead a good life), and sometimes whatever they felt comfortable telling (e.g. folkloric stories, stories of their families, or how to do or make something typical of Bora culture). Each of these stories was then repeated in Spanish to have a general idea of the topic during later translation. I also recorded 45 minute conversations between speakers to capture more natural speech, as well as to capture phenomena that may appear only within a larger conversational discourse, such as topic continuity.

For each of the free narratives, after recording them, I sat with the speakers to have them help with a translation. For each sentence, the speaker and I listened through headphones, and I had them repeat the sentence slowly for transcription purposes. The speaker then provided a translation for that sentence in Spanish.

The structure of this dissertation is as follows: this Chapter has introduced the Bora language, providing brief ethnographic information about the language and its speakers, and indicating previous and current work that has been done on the language, including personal fieldwork for this dissertation project. In Chapter 2 I elaborate on the grammar and morphosyntactic properties of Bora, including the phenomena in question regarding the distribution of subject clitics. Chapter 3 provides a preliminary overview of some current theoretical literature to be considered in detail in this dissertation, regarding theoretical phenomena that are potentially relevant for an analysis of these Bora clitics, including theories of binding and control. In Chapter 4, I analyze the Bora data in terms of the theories outlined in Chapter 3. In doing so, I determine that the Bora subject proclitics behave differently depending on the person and number of the subject. I also determine that the proclitics corresponding to 3$^{rd}$ person and 1$^{st}$/2$^{nd}$ person non-singular behave anaphorically, sometimes being locally bound and sometimes being bound at a distance, depending on the clitic and the syntactic structure. I also show that the 1$^{st}$ and 2$^{nd}$ person proclitics behave as pronominals, unlike the other proclitics. Chapter 5 concludes the work, reviewing the major contributions with respect to Bora syntax, especially the domains in which the Bora proclitics must be bound (if they

in fact need to be bound), and how this compares to other languages with similar types of long distance binding. This final Chapter also highlights the importance of this dissertation in using data from an understudied language to inform linguistic theories of binding and control.

# CHAPTER II

## Grammatical and Morphosyntactic Properties of Bora

This chapter focuses on the description of some of the linguistic properties of Bora that are relevant for the analysis of the preverbal subject proclitic pronouns in the language to be pointed out in later Chapters. I first describe the distribution of subjects in Bora, not only in terms of the subject proclitics being investigated, but also other ways that the subject can be expressed, including through overt nouns and nominal classifiers. Because the possible ways of expressing the subject vary given the person and number of the subject, I describe how each class of subjects is expressed. For the purposes of this dissertation, I take the term 'subject' to refer to the external argument of a predicate.[3]

I then discuss how embedded clauses are expressed in Bora, especially with regard to (what has been referred to as) finiteness in Bora (Thiesen & Weber 2012), and I point out a distinction between matrix and embedded clauses. Following this, I show some of the distribution of the reflexive and reciprocal morphemes in Bora, and how those morphemes interact with different environments in which co-reference among nominal arguments takes place. I then discuss how an apparent lack of finiteness in Bora (contra Thiesen & Weber 2012) has consequences for an analysis of Bora clausal subjects. I also illustrate the distribution of possessive pronouns in Bora; although these may not have direct relevance to the type of co-reference shown by subject pronouns, the clitics used for possession in Bora show striking similarity to the subject pronoun clitics, and also raise questions about how they can be involved in co-reference. Finally, I discuss other miscellaneous properties of Bora grammar which have direct relevance to the analysis presented in later chapters.

[3] While I recognize that there is a great body of literature discussing the exact syntactic position of the external argument of a predicate, this does not have a bearing on my analysis, and I assume the *v*P-internal subject hypothesis, in which external argument of a VP predicate to Merge into the derivation in the Spec, *v*P position.

Many of the basic descriptive grammatical properties of Bora have been presented in Thiesen and Weber (2012). Generally, the writing system for Bora follows the International Phonetic Alphabet (IPA), with the following exceptions:

| Bora orthography | IPA | Bora orthography | IPA |
|---|---|---|---|
| ⟨c, k⟩ | [k][4] | ⟨v⟩ | [ß] |
| ⟨ch⟩ | [ɟʰ] | ⟨f⟩ | [ɸ] |
| ⟨ñ⟩ | [ɲ] | ⟨w⟩ | [kʷ][5] |
| ⟨h⟩ | [ʔ] | ⟨y⟩ | [j] |
| ⟨j⟩ | [h] | ⟨u⟩ | [ɯ] |
| ⟨ll⟩ | [ɟ] | ⟨e⟩ | [ɛ] |

Table 1: Differences between Bora orthography and IPA

Additionally, Bora distinguishes aspirated and unaspirated voiceless stops, representing unaspirated stops with the voiced variant of that stop in IPA, and aspirated stops with the voiceless variant (e.g. ⟨p⟩ - [pʰ], ⟨b⟩ - [p]).[6] Long vowels are indicated by a simultaneous sequence of identical vowels (in some cases, a vowel can be lengthened by an adjacent affix as a property of that affix). High tone is marked with an acute accent, while low tone is not marked. On long vowels, each mora can carry tone individually, such that a high or low tone can appear on the first mora, the second mora, or both.

Some of the basic phonological properties in Bora include progressive palatalization conditioned by a preceding *i* (1)[7], and limited vowel harmony in which *i* harmonizes to *ɨ* when the following syllable contains an *ɨ* (2).

(1) *oohí**mye***
*oohí-**me***
dog-AN.PL (Thiesen & Weber 2012: 37)
'dogs'

(2) ***tɨ**mɨ́ɨ́heé*
***tí**-mɨ́ɨ́heé*
2SG.POSS-skin
'your hide/skin/fur' (Roe 2014: 7)

[4] In the orthographic system for Bora, *c* is used before the vowels *a*, *o*, and *u*, while *k* is used before *i*, *ɨ*, and *e*.
[5] Thiesen and Weber (2012) analyze this sound as a labial-velar stop [k͡p].
[6] The aspiration distinction found in Bora does not extend to its sister language, Miraña, nor to Bora's dialectal variant, Muinane, both of which have a true voicing distinction.
[7] In some cases, a preceding *a* can also condition palatalization. Aschmann (1993) and Seifart & Echeverri (2015) analyze this as proceeding from [ai] historically, which caused palatalization of the following vowel with later loss of the conditioning environment.

Bora nouns conform to a nominative-accusative alignment. The nominative is not marked with any overt morphology, while accusative case is marked only on animate nouns with the suffix *-ke*.[8]

(3) a. *o=ájtyumɨ́ wájpii-ke*
1.SG=see man-ACC
'I saw the man.'
b. *o=ájtyumɨ́ jaá*
1.SG=see house
'I saw the house.'

Several other oblique cases are also marked on nouns, including ablative, allative, sociative, instrumental, and benefactive.

### 2.1. Bora Tone

Bora has a complex tonal system which uses two tones: high and low. Every syllable in Bora carries tone, with the high tone being unmarked (default), such that every syllable without an existing lexical or grammatical tone will receive high tone. Tone can be lexically marked on a word, or can indicate one of several grammatical functions that are marked by tone in Bora, whether being associated with a grammatical affix or indicating a grammatical function all its own. Tone is so pivotal to the Bora language that Bora signal drums of different pitches, called *manguaré*, can be used to effectively communicate, using drumbeats that emulate the language (Seifart et al. 2018).[9] The primary function of tone in Bora appears to be grammatical, with very few lexical minimal pairs that are distinguished purely by tone (Berger & Seifart, forthcoming), though Thiesen & Weber (2012: 55) point out some examples.

Among the grammatical features encoded by tone in Bora are genitive constructions (4), embedded clause verbs (5), and imperatives (6-7).

[8] The animacy requirement for the accusative marker does not consider body parts to be animate.

(i) *aa-né=vá=a tsaápi-lle ɨ́ɨté-cunú eh-du nehwáyu néjuwa*
CON-CL:IN=RPT=REM one-CL:3F.SG see-MTR that-like crab arm
'One woman saw the arm of a crab.'

[9] Other features of the language besides tone are also conveyed by *manguaré*, including mora and vowel length.

The genitive construction in Bora is a right-headed complex consisting of the possessor (dependent) and the possessed (head). At the juncture of these items is what Thiesen & Weber term a floating genitive low tone. The genitive low tone will appear on the final syllable of the dependent if the head is mono- or bi-syllabic (4a). If the head is more than two syllables (4b), then the genitive low tone will appear on the first syllable of the head. The affected tones in (4-7) below are bolded.

(4) a. *méé**n**í-m**u*** *ja*
pig-CL:AN.PL house
'pigs' house'
b. *tá* *h**a**jchóta*
1SG.POSS height
'my height' (Thiesen & Weber 2012: 254, 264)

Embedded clauses in Bora are marked by a high tone on the initial syllable of the embedded clause verb. It is likely the case that this high tone was historically caused by a marked low tone on the the first syllable of the embedded clause subject, which then subsequently caused a high tone on the following syllable due to the LLX constraint (described later in this section). Synchronically, however, the high tone on embedded clause verbs is not dependent on the appearance of a subject, and will appear on the verb when no overt subject is present (see Berger & Seifart, forthcoming). See section 2.9 for cases in which Bora clauses do not have overt subjects.

(5) a. *wajpi* *m**a**jchó-hi*
man eat-PRED
'The man is eating.'
b. *wajpi* *imíllé* *i=m**á**jcho-ne*
man want 3COR=eat-CL:IN
'The man wants to eat.'

According to Thiesen & Weber (2012: 334), imperatives in Bora bear what they term the 'nonfinite low tone'. While I argue in section 2.6 that this tone outside of imperatives signifies nominalization rather than nonfiniteness, it does not seem to indicate nominalization in imperatives.[10] (Positive) imperative verbs in Bora take one of two subject proclitics, depending on whether the subject is

[10] It is also possible that the low tone on imperatives is licensed by something other than nonfiniteness.

singular or plural.[11] Imperatives with singular subjects take the proclitic *di*= if the verb is monosyllabic (6a), and *d*= if the verb is polysyllabic and begins with a vowel (6b). Polysyllabic verbs beginning with consonants have no overtly expressed subject in the imperative when there is a singular subject (6c). Imperatives with non-singular subjects in Bora take the Speech Act Participant proclitic *me*= (7). Thiesen & Weber (2012) describe the tone of imperatives as a low tone which must occur as early as possible in the word, without preceding the antepenultimate syllable, and without appearing on the initial syllable of the verb.

(6) a. ***dí****=dyoó*
2SG.IMP=eat.meat.EMPH
'Eat (meat)!'
b. ***d****=ímibájchoó*
2SG.IMP=fix.EMPH
'Fix it!'
c. *májcho*
eat
'Eat!'

Thiesen and Weber (2012: 334)

(7) a. *mé=m**a**jcho*
SAP=eat
'Eat!' (pl)
b. *mé=ím**i**bájcho*
SAP=fix
'Fix (it)!' (pl.)

Roe (2014) argues that Bora uses a mixed tone/stress system in which the first syllable of mono- and bi-syllabic words receives stress, and words with three or more syllables will receive stress on the antepenultimate syllable. Roe also claims that Bora assigns low tone to stressed syllables, which indicates that, if Bora does indeed have a mixed tone/stress system, that tone is dependent on stress. However, Roe also states that tone is the only indicator of stress in Bora. Roe's acoustic analysis of vowel intensity in Bora is used to show that there is a significant difference in the intensity of high and low toned short vowels. Until more acoustic information can be gathered to show a difference between stressed and unstressed low-tone vowels, it is unclear that there is a stress distinction in Bora.

[11] Negative imperatives, or prohibitions in Bora do not take a subject proclitic, but rather take the prohibitive suffix *–hdi*.

Acoustically, Thiesen and Weber (2012) note that, in a series of high tones, the latter will rise slightly in pitch higher than the former. Similarly, in a series of low tones (when it is allowed; see the OCP constraint below), the latter will drop slightly in pitch lower than the former.

Though high tone is generally considered the default tone in Bora, appearing on any syllable that does not already have a lexical or grammatical tone, there is an exception for syllables at the end of a tonal phrase, which receive final default low tone (FDLT). Lexical tone in Bora, when it appears on a syllable, is also generally low tone. Other syllables without lexically specified tones receive their tone either by the default high tone, by FDLT, or through tone sandhi and a series of ordered tone spreading rules.

One important tonal restriction on Bora tone is an Obligatory Contour Principle (OCP), which Thiesen and Weber (2012) label the LLX constraint. This constraint disallows two consecutive syllables with low tones[12], except at the end of a word, where maximally two consecutive low tones can occur. The LLX constraint also holds across words within a tonal phrase, such that a word ending in a low tone cannot be followed by a word whose first syllable also has a low tone if they are within the same tonal phrase. In addition to these constraints within a tonal phrase, the last syllable of a tonal phrase must bear a low tone. Violation of the LLX constraint can be avoided in multiple ways, including by tone docking (8), blocking (9), or delinking (10).

(8) a. *úmehe*
tree
'tree'
b. *úméhé-coba*
tree-AUG
'big tree'
c. *úméhé-cobá-ne*
tree-AUG-PL
'big trees'
d. *úmehéé-né-coba*
tree-PL-AUG
'many (big) trees' (stressing plurality) Thiesen and Weber (2012: 69, 206)

The base word in (8a) shows the lexical high tone on the first syllable, and the two consecutive low tones which are unaffected by the LLX constraint due to their final position in the word. In

[12] A bimoraic syllable with one low tone and one high tone does not constitute a low-toned syllable with regard to this constraint.

(8b), the augmentative suffix *–coba* has two floating high tones that dock onto the ultimate and penultimate syllables of its host, and also has a lexical low tone on its own first syllable (*co*). That is, whenever *–coba* attaches to a host, high tone is transferred to the two syllables preceding it. By adding a suffix onto *úmehe*, the two low-toned syllables are no longer in final position, creating a violation of the LLX constraint. The high tones docked onto the host by the *–coba* suffix ameliorate this violation. (8d) follows the same pattern as (8b), with an additional morpheme preceding the augmentative suffix.

Adding a further suffix after the augment forces a further change. While the augment in (8c) still docks its two high tones onto the host, the *–ne* plural suffix causes a high tone on the second syllable of the augment *–coba*, not through docking, but through tone sandhi so as not to violate the LLX constraint.

(9) *íhveté-tso-té-roó-be*
fix-CAUS-go.do-FRS-CL:M.SG
'In vain did he go to make it stop.' Thiesen and Weber (2012: 64)

In (9), tonal rules apply cyclically as suffixes are added to the words. Each of the suffixes added carries a low tone that docks onto the final syllable of the word it is added to. The low tone that the causative suffix *–tso* would dock onto the stem is instead blocked by the lexical low tone on the second syllable of *íhveté*, since docking that low tone onto the final syllable would then violate the LLX constraint. This in turn creates the environment for the following suffix *–te* to dock its low tone onto the causative suffix *–tso*. The frustrative suffix *–ro* is thus blocked from docking its low tone onto the preceding syllable, since doing so would violate the LLX constraint. This forces *–te* to carry high tone by tone sandhi to avoid violating the LLX constraint. Finally, the classifier *–be* then docks its low tone onto the first mora of the preceding syllable.[13]

While the example in (9) shows that a preexisting conflicting tone will block other tones from docking, there are some classifiers which dock their tones by delinking, or undoing in a sense, conflicting tones.

(10) *ímíbájchó-tu-mútsi*
fix-NEG-CL:M.DU
'They (dual masc.) did not fix it' Thiesen and Weber (2012: 84)

[13] The masculine singular classifier also lengthens the preceding syllable.

In (10) the negative suffix *–tu* docks a low tone onto the host word. Following that, the classifier *–mutsi* also docks a low tone onto the previous syllable, in this case the negative suffix *–tu*. The combination of both docked low tones would violate the LLX constraint, with low tones both on the final syllable of *–ímíbájcho* and on the suffix *–tu*. However, rather than the low tone on *–tu* imposed by *–mutsi* being blocked, the low tone docked by *–tu* on the host word is delinked, replaced by the default high tone, so that *–mutsi* may dock its low tone. The detailed process for the form shown in (10) is shown in (11) below.

(11)

| | | | | | | |
|---|---|---|---|---|---|---|
| *í* | *mi* | *baj* | *cho* | | | |
| | | | Ⓛ | *-tu* | | |
| | | | | Ⓛ | *-mu* | *tsi* |
| | | | **X** | | | |
| | Ⓗ | Ⓗ | Ⓗ | | Ⓗ | Ⓛ |
| *í* | *mí* | *báj* | *chó* | *tu* | *mú* | *tsi* |

'They (dual masc.) did not fix it'

'to fix'
NEG
CL:DU.M
LLX (X = delinked tone)
FDLT, Default High Tone

Thiesen and Weber (2012: 84)

### 2.2. Expression of Clausal Subjects in Bora

Bora, with very few exceptions (see section 2.9), obligatorily expresses the subject of every clause. The subject of a sentence in Bora may be expressed in multiple ways: by an overt noun or personal pronoun (12a), by a postverbal classifier (12b), or by a series of subject proclitics in embedded clauses, which is the primary focus of this dissertation (12c).

(12) a. ***wáhtsaji*** *tsá-juco*
carachama come-PRF
'The carachama came.'

b. *májcho-**lle***
eat-CL:3SG.F
'She eats.'

c. *ó* *imíllé* [ ***o***=*májcho-ne* ]
1SG want 1SG=eat-CL:IN
'I want to eat.'

Any clause in Bora has the option of having an overt noun of the type in (12a) as a subject, but there are some cooccurence restrictions between the different types of subject expressions. I have no evidence that classifier subjects, as in (12b), can co-occur with either overt nouns or pronouns,

or with subject proclitics, though other researchers on Bora disagree on this (see section 2.3.3). Classifiers perform a number of functions in the language, which are outlined in Seifart (2005) for Miraña, a close dialectal variant of Bora. Aside from expressing (or, in Miraña, cross-referencing) the subject, classifiers are obligatorily added to relative clauses (see section 2.9.3 for more on the structure of Bora relative clauses), they can be productively used as derivational morphemes on nouns, they provide person and number information on demonstratives and possessive pronouns, and they track nominal referents in discourse. For detailed descriptions of these uses, see Seifart (2005). Further examples are given in (15-17) below. To my knowledge, no formal theoretical syntactic analysis exists of classifiers used as subjects of sentences in Bora.

There are some instances when overt nouns and pronouns must co-occur with subject proclitics (e.g. 1st and 2nd person non-singular subjects), and some instances when overt nouns and pronouns do not co-occur with subject proclitics (e.g. 3rd person overt subjects in main clauses cannot co-occur with any subject proclitics, and the 3COR marker in embedded clauses does not co-occur with overt nouns or pronouns). These restrictions fall out from the properties of the subject proclitics described in sections 2.3 and 2.5.

Except for classifier subjects, Bora subjects cannot occur postverbally.

(13) a. *wajpi* *majchó*
man eat
'The man is eating.'
b. **majchó* *wajpi*
eat man
intended: 'The man is eating.'

Bora also has what has been analyzed as a predicative marker, as in (14a), which is optional, but can only be used when there is a preverbal subject. Although the optionality of the predicative marker raises questions as to its distribution, Thiesen and Weber (2012) claim that use of the predicative marker is ungrammatical if the subject of the clause is expressed by a postverbal classifier. There seem to be other restrictions on its use as well, as shown by the data in (14b-c) in that co-occurrence of the predicative morpheme with negation is not permitted, but further research will need to be done to say more about the predicative marker's distribution. Seifart (2015a) notes that the predicative marker is often used when a new participant is introduced in the discourse with an overt NP.

(14) a. *amoóbe-ke* *ó=majchó-hi*
fish-ACC 1SG=eat-PRED
'I am eating fish.'
b. * *dóhmeba* *tsá* *o=májcho-tú-hi*
aji negro NEG 1SG=eat-NEG-PRED
intended: 'I am not eating aji negro.'
c. *dóhmeba* *tsá* *o=májcho-tu*
aji negro NEG 1SG=eat-NEG
'I am not eating aji negro.'

Personal pronouns in Bora can be used to express most subjects, with the exception of 1st and 2nd person singular and 1st person plural inclusive subjects, which are often expressed by proclitics, unless the clause contains a second position enclitic. Bora pronouns have 3-way distinctions in person (1st, 2nd, 3rd), number (singular, dual, plural), and gender (masculine, feminine, inanimate). Animate Bora personal pronouns are given in the table below. There are also inanimate 3rd person pronouns, which are formed by combining the stem *té-* with classifiers and number markers, e.g. *teé-ne* (inanimate 3rd singular) *té-neé-cu* (inanimate 3rd dual) *té-ne-hɨ* (inanimate 3rd plural) (Berger and Seifart forthcoming).

| | Singular | | Dual | | Plural |
|---|---|---|---|---|---|
| | Masculine | Feminine | Masculine | Feminine | |
| 1st | *(o=)*[14] | | *muhtsi* | *muhpɨ* | *muúha / (me=)*[15] |
| 2nd | *(u=)* | | *ámuhtsi* | *ámuhpɨ* | *ámuúha* |
| 3rd | *diíbye* | *diílle* | *diityétsi* | *diityépɨ* | *diitye* |

Table 2: Bora Animate Personal Pronouns

As shown in (12b) above, the subject of a clause in Bora can also be expressed by a postverbal classifier. Bora has a rich system of nominal classifiers, most of which are shape- or alignment-based. These can be used to signify agreement between nouns and modifiers (15), to

[14] The full form of the 1st and 2nd person pronouns are *oó* and *uú*, but preverbal personal pronoun subjects in Bora undergo vowel shortening.
[15] The choice between the two 1st person plural subject expressions depends on an inclusive/exclusive distinction, described further below.

compositionally change the meaning of a word (16), or to identify a 3rd person subject of a clause (17).

(15) *úméhe-wa* *átéré-wa*
tree-CL:slab worthless-CL:slab
'a worthless plank' Thiesen and Weber (2012: 174)

(16) *dsɨ́ɨ́dsɨ* / *dsɨ́ɨ́dsɨ-jɨ* / *dsɨ́ɨ́dsɨ-hɨ́ámɨ*
money / money-CL:disk / money-CL:leaf
'money 'coin' 'bill' Thiesen and Weber (2012: 200)

(17) *i-íhjyu* *íllu-re* *píllaríhcyuu-be* *caáme-vu* *baá-vu*
3POSS-mouth like.this-only roll.up-CL:M.SG up-ALL down-ALL
'He had rolled up his mouth (lips) upwards and downwards.'

Notably, multiple classifiers of the type in (16) can be used on a single word to form a compositional meaning, as in (18). Seifart's (2005) dissertation describes, in great detail, the classifier system of Miraña.

(18) *mútsɨ́ɨ́tsɨ-he-hɨ́ámɨ*
caimito[16]-CL:tree-CL:leaf
'leaf of a caimito tree' Thiesen and Weber (2012: 180)

The small set of 'general' classifiers that can be used to express a clausal subject, of the type illustrated in (17), indicate natural gender and number of the subject. Only main clauses in Bora can express the subject with classifiers; this is disallowed in embedded clauses (Thiesen & Weber 2012). These include *-be* for masculine[17], *-lle* for feminine, *-mutsi* for dual masculine, *-mupɨ* for dual feminine, *-me* for plural, and *–tsɨ* for a child. These general classifiers can be suffixed to the verb to indicate the subject of a clause in the third person, in which case they are used instead of the preverbal subject, as was seen in (12b) and (17). These general classifiers are largely used to

[16] A caimito (*pouteria caimito*) is a tropical fruit with a sweet taste and a sticky resin.
[17] Thiesen and Weber (2012) point out some cases in which the *–be* classifier can indicate something other than 3rd person singular masculine, including 1st person (i) and 2nd person (ii).

(i) *áá-be-ke* *táá-tsɨ́ɨ́ju* *ɨ́ɨ́cú-ve-é-hi*
CON-CL:M.SG-ACC 1.SG.POSS-mother serve-SIN-FUT-PRED
'My mother will serve me food.' Thiesen and Weber (2012: 399)

(ii) *muhdú* *májchóó-be-jɨ́ɨ́* *ú=pe-é-hi*
how eat-CL:M.SG-deny 2SG=go-FUT-PRED
'How is it that you are going without eating?' Thiesen and Weber (2012: 329)

track nominal referents throughout discourse, especially by attaching them to discourse connective particles.[18]

(19) a. *pee-múpɨ* *mújcó-ju-ri*
go-CL:F.DU harbor-CL:tube-LOC
'They (two females) went to the harbor.'
b. *aa-múpɨ=váa* *pámaúcu-jíjcya-rá* *tee-ne* *nújpacyo*
CON-CL:F.DU=QUOT.REM carry-REP-FRS PN-CL:IN water
'And they (two females) tried to carry water.' Seifart (2010: 902)

### 2.3. Bora Subjects: Main Clauses

#### 2.3.1. 1st and 2nd Person Singular, Main Clause Subjects

In Bora, 1st and 2nd person singular pronouns referring to the clausal subject are generally expressed as subject proclitics attached to the verb.

(20) a. *ó=pɨaabo* *tá-cááni-ke*
1SG=help 1SG.POSS-father-ACC
'I am helping my father.'
b. *ú=majchó-hi*
2SG=eat-PRED
'You are eating.'

In some cases, suffixes can be added to the subject pronoun. In these cases, the suffix attaches to the clitic, and the clitic is 'repeated' on the verb, so that there is always an expression of the subject cliticized to the verb in cases of 1st or 2nd person singular subjects.

(21) *oó=ne-cu*[19] *ó=ájtyumɨ* *wájpii-ke*
1SG=REC-STR 1SG=see man-ACC
'I saw the man.'

[18] The connector pronoun *aa-* in Bora connects a sentence to previous discourse. It can provide referential continuity by agreeing in noun class and number with an antecedent, and can also carry temporal, evidential, and other event-related information that connects the clause to previous discourse. For a detailed description and analysis of the connector pronoun in Bora, see Seifart (2010).

[19] As described in footnote 13, the full form of the 1st person pronoun is *oó*, which is the form attached to the second position tense clitic. The preverbal 1st person singular subject has undergone vowel shortening.

### 2.3.2. 1st and 2nd Person Non-Singular, Main Clause Subjects

For 1st and 2nd person non-singular subjects (dual and plural), in addition to the personal pronoun from Table 2, the verb carries an additional mandatory proclitic *me=* (22). This has been referred to in previous research on Bora as the Speech Act Participant (SAP) marker. If this proclitic is used without a personal pronoun from Table 2, the subject can be interpreted as 1st person plural inclusive (23a), where 'inclusive' is interpreted as including the addressee. This is not to say that the presence of an overt personal pronoun *muúha* is necessarily interpreted as 1st person plural exclusive, rather that lack of a pronoun triggers an inclusive interpretation. As such, I only include (incl.) in glosses for SAP proclitics that indicate inclusivity by appearing without an overt pronoun, but outside of this subsection I do not gloss exclusivity in this way. This same SAP proclitic (or perhaps a homophonous form of it) is also used for impersonal subjects (23b) (leading to an ambiguity in the interpretation of the subject in some cases, as in (23)). Finally, it is also used as the subject of weather verbs (24).

(22) *muhá=ne-cu* *me=tsaa* *iquito-vu* *wákimyéi-vu*
1PL=REC-STR SAP=come Iquitos-ALL work-ALL
'We came to Iquitos to work.'

(23) *ímí* *me=cúwá-hajchíí* *tsá* *me=chémé-i-tyu-ró-hi*
good SAP=sleep-COND NEG SAP=be.ill-FUT-NEG-FRS-PRED
a. 'If we (incl.) sleep well, we are not likely to get sick.'
b. 'Whoever sleeps well is not likely to get sick.' (Thiesen and Weber 2012: 127)

(24) *me=allé-hi*
SAP=rain-PRED
'It's raining.'[20]

As indicated above in (23), a general property of the SAP marker in both main and embedded clauses is that, when it occurs without an overt personal pronoun or noun phrase, the sentential subject can receive a 1st person plural inclusive reading, whereas an overt 1st person plural pronoun with the SAP will often receive an exclusive reading. The relevant pair showing this distinction is given in (25-26).

[20] The same meaning can also be expressed by an inanimate classifier rather than the SAP.

(i) *allé-ne*
rain-CL:IN
'It's raining.'

(25) *tee-ne* *újcuu-be* *me=májcho-ki*
that-CL:IN get-CL:M.SG SAP=eat-PUR
'He got that in order that we (incl.) eat it.' Thiesen and Weber (2012: 233)

(26) *tee-ne* *újcuu-be* *muha* *me=májcho-ki*
that-CL:IN get-CL:M.SG 1.PL SAP=eat-PUR
'He got that in order that we (excl.) eat it.' Thiesen and Weber (2012: 233)

Notably, the SAP's use as a marker of coreference can further be distinguished from the 1st person plural inclusive usage when used in object position, in that 1st person plural exclusive objects (27b) will be expressed with the personal pronoun from Table 2, while 1st person plural inclusive objects will be expressed with the SAP (27a).

(27) a. ***me-ke*** *ɨ́ɨte-lle*
SAP-ACC look-CL:F.SG
'She looks at us (incl.).' (Thiesen & Weber 2012: 233)

b. *diitye* ***múha-ke*** *ájtyumɨ*
3.PL 1.PL-ACC see
'They saw us (excl.).'

Because of this distinction, it will be important to distinguish the use of the SAP marker as a mandatorily cooccurring morpheme with a 1st or 2nd person non-singular subject (in main clauses and coreferent subjects of embedded clauses, to be described in section 2.5.2), as in (22) above, and its use as a 1st person plural inclusive pronoun.

### 2.3.3. 3rd Person, Main Clause Subjects

3rd person subjects in Bora main clauses can be expressed using either overt noun phrases (28) or classifier subjects (29).

(28) *wajpi* *liíñaja*
man hunt
'The man is hunting.'

(29) *cúúvénetu* *taa-híjcyaa-be* *cójɨhajchóta* *wayéeve-túu-be*
early cry-HAB-CL:M.SG all.day rest-NEG-CL:M.SG
'From the morning he would cry all day without stopping.'

Classifier subjects in Bora are unique to 3rd person subjects. Thiesen & Weber (2012: 129) note that a classifier subject cannot co-occur with a preverbal subject NP, what they call their PREVERBAL SUBJECT CONSTRAINT, though a classifier subject can be followed by an overt NP.

Seifart (2015a) claims, to the contrary, that both preverbal and post-verbal NPs are allowed when there is a classifier subject. My own data does not have any evidence in support of one theory or the other.

(30) a. *nááni* *tsáá-be*
my.uncle come-CL:M.SG
'My uncle came.'
b. *tsáá-be* *nááni*
come-CL:M.SG my.uncle
'My uncle came.' (Seifart 2015a: 1771)

### 2.4. Reflexives and Reciprocals

Both object reflexives (31a) and reciprocals (31b) in Bora are expressed by suffixes on the verb, which have been described in the literature as valency-changing suffixes[21] (see e.g. Seifart 2015b, Berger & Seifart forthcoming). These suffixes delete the object of a verb, with the agent and patient semantic roles falling together, reducing the valency of the verb by one argument.

(31) a. *wáhdaí-nú-**meíí**-bye*
cut-MTR-**REFL**-CL:M.SG
'He cut himself.' Thiesen and Weber (2012: 435)
b. *tá-ñá-hbe-mútsí* *méénú-**jcatsí**-hi*
1.SG.POSS-sibling-CL:M.SG-CL:DU.M hit-**RECIP**-PRED
'My two brothers are hitting each other.' Thiesen and Weber (2012: 148)

In Bora, the suffix *–mei* has been described as the reflexive marker (see e.g. 31a), but it is also used in expressions corresponding to the passive (Thiesen and Weber 2012). The reflexive is classified as one of the valency-changing suffixes of Bora, by reducing the valency of the verb through reduction of the patient argument. As such, both the passive and reflexive readings of the *–mei* suffix reduce valency in a similar way. The passive reading is achieved by adding the *–mei* suffix to the causative suffix (the sole valency increasing suffix in Bora, which adds a causer to the predicate, see also (34) below), as in (32).

(32) *í-ñáhbe-dí-vú* *méénú-**tsá-meíí**-bye*
3COR-brother-AN-ALL hit-**CAUS-REFL**-CL:M.SG
'He$_i$ provoked his$_i$ brothers to beat him$_i$ up.' Thiesen and Weber (2012: 145)

[21] Valency-changing morphology is an areal feature of the Amazon, and has been described in other languages (see e.g. Aikhenvald 2012).

lit. ‘He$_i$ allowed himself to be beaten up by his$_i$ brothers.’

Interestingly, if a verb would take two coordinated complement noun phrases, the complements would surface as in (33a), with the coordination marked by lengthening the final vowel of each of the elements being coordinated.[22] However, if one of those noun phrases is the reflexive anaphor, the reflexive is marked on the verb, while the other argument is expressed with an adverbial modifier (33b). The other noun phrase is still acting as an argument of the verb, evidenced by the case marking it receives. However, because the reflexive marker is not, itself, a noun phrase, it does not fit into the coordination strategy of lengthening the final vowel of the conjoined noun phrases.

(33) a. *í-ñá-hbe-kee* *í-ñáá-lle-kee*
3.POSS-sibling-CL:SG.M-ACC.CONJ 3.POSS-sibling-CL:SG.F-ACC.CONJ
*íí-tsíímе-kee* ***íjcyá**-me-ke* *tsajtyéé-be* *té-hullé-vu*
3.POSS-children-ACC.CONJ be-CL:AN.PL-ACC take-CL:SG.M that-CL:yonder-ALL
‘He took his brother, his sister, and his children yonder.’
BORA (Thiesen and Weber 2012: 209)

b. *wajpí-ñe* *ííte-meí* *wállee-ké-ne* *idye*
man-CL:IN see-REFL woman-ACC-CL:IN in.addition
‘The man saw himself and the woman.’ (lit. ‘The man saw himself, the woman also.’)

As for other valency changing suffixes, Bora’s reciprocal marker, the suffix *–jcatsi* (31b), has also been described as a morpheme which reduces the valency of the verb in the same way the reflexive marker does, by suppressing the patient argument. Meanwhile, the causative suffix *–tso* increases the valency of the verb by introducing a causer of the action. The causer then acts as the subject of the verb, with the former subject being demoted to an accusative-marked object of the verb. Any object of the underived verb will then become an allative-marked object of the newly derived verb once the causative suffix is added (Seifart 2015b).

(34) *ó=imíllé* *ó=wajácu-tsó-ne* *badsíjcajá-dí-vú* *ovátsa-ke*
1SG=want 1SG=know-CAUS-CL:IN girl-AN-ALL youth-ACC
‘I want to introduce the youth to the girl.’

[22] In the case of (33b), the coordination is also marked by a summation word, here the copula, which bears the same case as the coordinated items.

The two valency reducers in Bora, the reflexive and the reciprocal, are relevant for their binding properties, since both reflexives and reciprocals are typically assumed to be subject to the canonical Binding Principle A. For the analysis presented in this dissertation, it is also relevant that the reflexive and reciprocal markers reduce the valency of the verb with respect to the object, since the Binding Principles, discussed further in Chapter 3, have slightly different properties for subjects and objects, especially regarding long distance reflexives.

In Bora, the subject appears to be obligatorily expressed in each clause, as I will discuss in the course of this dissertation (though see section 2.9 for possible exceptions). The description in section 2.3 of the expression of the subject in various persons and numbers applies to main clauses. However, the expression of the subject in embedded clauses varies, also depending on the person and number of the subject and, in some cases whether or not the embedded clause subject is coreferent with the main clause subject.

### 2.5. Bora Subjects: Embedded Clauses

Bora embedded clauses are marked by a high tone on the initial syllable of the embedded clause verb. Embedded clauses in Bora are formed by a full CP. This is evidenced by the fact that embedded clauses in Bora are finite (see section 2.6 for further detail on this), as well as the overt appearance of a subject in the embedded clause.

Further evidence of full CP clause structure for embedded clauses comes from second position clitics. These clitics, which indicate such grammatical information as tense and evidentials, must appear in the second position of a clause, and I argue in section 2.8 that second position clitics in Bora attach to their host in the Spec, CP position. The fact that these clitics can appear in embedded clauses, shown in (35) by the non-witnessed evidential clitic and the recent past tense clitic, indicates that embedded clauses are formed of a full CP structure.

(35) *tsá* *o=wájácu-tu* *muhdú=**hja**=**né*** *wajpi* *méénu-ne* *mɨ́ɨ́ne*
NEG 1SG=know-NEG how=NWIT=REC man make-CL:IN canoe
'I don't know how the man made the canoe.'

As with main clauses, embedded clauses generally require the subject of the clause to be overtly expressed (again, see section 2.9 for exceptions to this generalization). Embedded clauses in Bora do not ever have the subject expressed by a classifier (Thiesen & Weber 2012), but can express

their subjects with overt noun phrases or preverbal subject proclitics. This section focuses on instances of proclitic subjects of embedded clauses.

### 2.5.1. 1st and 2nd Person Singular, Embedded Clause Subjects

For clauses with 1st and 2nd person singular subjects, there is no difference in how the subject is represented as compared to main clauses. There is no difference between embedded clauses with subjects that are coreferent with the main clause subject (36a-b), and embedded clauses that display disjoint reference (36c).

(36) a. *ó=imíllé* *o=cúwa-hi*[23]
1SG=want 1SG=sleep-PRED
'I want to sleep.'
b. *ú=imíllé* *u=cúwa-hi*
2SG=want 2SG=sleep-PRED
'You want to sleep.'
c. *ó=imíllé* *o-ke* *u=ííte-ne*
1SG=want 1SG-ACC 2SG=see-CL:IN
'I want you to see me.'

As shown, 1st person singular preverbal subjects of embedded clauses will always appear as *o=* (36a), and 2nd person singular preverbal subjects as *u=* (36b). This includes sentences where the valency of the embedded clause verb has been reduced by the reflexive marker. In this case, co-reference across clauses is achieved via identical subject proclitics in the same way as non-reflexive and non-reciprocal clauses (37), since the agent (typically the subject) is not the argument that is reduced by valency-changing suffixes (compare this to the disjointly referent counterpart in (36c) above, as well as to (38) below, in which the valency of the embedded clause verb has not been reduced).

(37) *ó=imíllé* *o=ííte-meí-ñe* *míícumi-ri*
1.SG=want 1.SG=see-REFL-CL:IN mirror-INS
'I want to see myself in the mirror.'
(38) *ó=imíllé* *péjcore* *u-ke* *o=ííte-ne*
1.SG=want tomorrow 2.SG-ACC 1.SG=see-CL:IN
'I want to see you tomorrow.'

[23] One might note that the tone on the subject of the main and embedded clauses is not the same. This is because of a restriction in Bora that disallows two consecutive low tones except at the end of a sentence or tonal phrase. Because of the lexical low tone on the first syllable of *imíllé*, this would be a violation of the restriction. To rescue this, Bora puts a high tone on the matrix clause subject pronoun here. The same issue does not occur in the embedded clause due to the requirement that embedded clause verbs must have a high tone on their first syllable.

As shown in all of the cases in (36-38), no matter whether the embedded clause subject is coreferent or disjointly referent with the main clause subject, 1st and 2nd person singular preverbal subject clitics are still expressed in the same way.

### 2.5.2. 1st and 2nd Person Non-Singular, Embedded Clause Subjects

As for 1st and 2nd person non-singular subjects in Bora, these are expressed in main clauses with the SAP marker, described above and shown in section 2.3.2, in some cases in addition to the subject pronoun (recall that the SAP appearing without an overt pronoun can indicate a 1st person plural inclusive subject or an impersonal reading, and is also used with weather verbs). For cases in which the embedded clause subject is coreferential with the main clause subject, the SAP marker surfaces in both the main and embedded clause, as in (39).

(39) *muúha me=imílle me=íjchi-ñe*
1.PL SAP=want SAP=swim-CL:IN
'We want to swim.'

For cases in which the embedded clause subject forms a part of the group of the higher clause subject, the coreference marker is not used, signaling that full identity coreference is necessary for the coreference clitics to be used.

(40) *muhtsi mé=imíllé o=píjcu-té-ne*
1.DU.M SAP=want 1.SG=fish-go.do-CL:IN
'We want me to go fishing.'

Disjoint reference for 1st and 2nd person non-singular subjects is expressed with an overt pronoun and the SAP marker (41a, 43), or with just the SAP marker, if the embedded clause subject has a 1st person plural inclusive reading (41b). When the main clause and embedded clause subjects are coreferent, an overt subject is ungrammatical in the embedded clause (42).

(41) a. *ó=imíllé muha me=píjcyu-te-ne*
1SG=want 1.PL SAP=fish-go.do-CL:IN
'I want us to go fishing.'
b. *ó=imíllé me=píjcu-té-ne*
1SG=want SAP=fish-go.do-CL:IN
'I want us (incl.) to go fishing.'

(42) *$muha_i$ mé=imíllé $muha_i$ me=májcho-ne*

1.PL SAP=want 1.PL SAP=eat-CL:IN
'We want to eat.'

(43) *muha mé=wajácú ámuha jáá-hañe me=méénu-ne*
1.PL SAP=know 2.PL house-VAR[24] SAP=make-CL:IN
'We know that you all build houses.'

I analyze in Chapter 4 whether the SAP marker fits the criteria to be analyzed as an overt realization of PRO, or whether it should be analyzed as an agreement marker, or as an anaphor.

### 2.5.3. 3rd Person, Embedded Clause Subjects

For 3rd person subjects of all numbers, when the embedded clause subject is coindexed with the matrix clause subject, the subject of the embedded clause will be represented by the proclitic *i=* (44), which has been glossed in the literature as SELF; as I further discuss below, the anaphoric properties of this marker do not appear to align with the SELF morpheme that has been described, for example, in the Germanic languages. To avoid confusion between SELF morphemes that have been described for other languages and the marker being described in Bora, I will refer to the Bora proclitic as 3COR to indicate its properties of coreference with a 3rd person antecedent.

If the embedded clause subject in Bora is not co-referent with the matrix clause subject, the embedded clause subject must be expressed overtly with an overt noun phrase or non-clitic pronoun, as in (45).[25]

(44) a. *imíllé-mé i=májcho-ne*
want-CL:AN.PL 3COR=eat-CL:IN
'They want to eat.'

b. *wajpíí-mu imíllé i=ímivyé-ne jaá i=myeénu-ne*
man-PL want 3COR=complete-CL:IN house 3COR=make-CL:IN
'The men want to finish building this house.'

(45) *imíllé-mé dibye májcho-ne*
want- CL:AN.PL 3.SG.M eat-CL:IN
'They want him to eat.'

[24] VAR is a plural suffix that can be used to indicate variety.

[25] I have not found evidence of embedded clause disjoint reference subjects being marked solely by a postverbal classifier. If this were the case, the embedded clause verb would need to carry two classifiers, something which is found elsewhere in Bora. Further evidence would be needed to make a strong claim as to whether this is allowed in the language.

The 3rd person coreferent subject morpheme in Bora is different from the "self" described in Germanic, in that the German *sich*, generally glossed as "self", can be interpreted as either a reciprocal or a reflexive marker, as in (46). As illustrated above, reflexives and reciprocals in Bora are marked with separate morphemes. Additionally, German does not require a subject for all embedded clauses in the same way that Bora does, and so the distribution of the morpheme differs as well.

(46) *Sie starrten sich an*
3.PL stared self PTCL
'They stared at themselves/each other.'

Relevantly for any comparative analysis of Bora and German, the German *sich* pronoun occurs exclusively in object position, whereas the Bora 3COR marker occurs in subject position. Reuland (2006c) describes in detail the anaphoric systems of Germanic languages. These are reviewed in Chapter 3.

Similarly to the reflexives and reciprocals shown above in section 2.4, coreference for reflexive and reciprocal objects in embedded clauses is expressed with the same suffixes as main clauses, as shown in (47-48). For both reflexives and reciprocals in embedded clauses, the antecedent is still overtly expressed as the subject of the embedded clause.

(47) *dibye imíllé ɨ=ɨíte-meí-ñe mɨ́ɨ́cúmɨ-ri*
3.M.SG want 3COR=see-REFL-CL:IN mirror-INS
'He wants to see himself in the mirror.'

(48) a. *ɨíté-jcátsi=mútsi*
see-RECIP-CL:2.DU.M
'We are seeing each other.'[26]

b. *wálle-e wajpi-e imíllé ɨ=ɨíté-jcatsí=ñe*[27]
woman-CONJ man-CONJ want 3COR=see-RECIP=CL:IN
'The woman and the man want to see each other.'

c. *walle imíllé tsɨ́ɨ́me waajácú-jcatsí=ñe*
woman want child.PL know-RECIP=CL.IN
'The woman wants the children to know each other.'

[26] Interestingly, the SAP marker doesn't seem to be required in (48a). It might be that SAP markers are not required when the subject is expressed as a classifier. More data collection would be needed to shed light on this.
[27] In (48b), a phonological rule causes the 3COR marker not to be pronounced in the embedded clause, with [i] being assimilated to [ɨ].

In sum, the clitics discussed throughout this section are the clitic pronouns whose investigation will primarily inform this dissertation project. They are consolidated below in Table 3.[28]

| | Singular | Dual | Plural |
|---|---|---|---|
| 1st | *o=* | *me=* | |
| 2nd | *u=* | | |
| 3rd | *i=* | | |

Table 3: Embedded clause coreferent subject clitics in Bora

### 2.6. Finiteness in Bora

Thiesen and Weber (2012) claim that there is a tonal finiteness distinction in Bora, whereby non-finiteness is marked by a low tone on the antepenultimate syllable of a verb. If, however, the verb is shorter than three syllables, then this low tone appears on the first syllable. The proposed nonfinite low tone is shown on the first syllable of 'heal' in (49b).

(49) a. *ó=ájtyumɨ́-hi* *táábóó-be-ke*
1SG=see-PRED heal-CL:SG.M-ACC
'I saw him doctoring.'
b. *ó=ájtyumɨ́-hi* *taabóó-be-ke*
1SG=see-PRED heal.NMLZ-CL:SG.M-ACC
'I saw the doctor.' BORA (Thiesen and Weber 2012: 68)

The tonal distinction referred to by Thiesen and Weber, however, seems to be more of a process of nominalization than it is a standard finiteness distinction. A review of the examples given by Thiesen and Weber suggests that Bora does not actually seem to have a finiteness distinction. That is, there does not seem to be any sort of distinction between the forms of verbs in Bora that would have differences in finiteness in other languages. In fact, Thiesen & Weber (2012: 177) note that examples like (49b) are probably agentive nominalizations.[29] Perhaps the most striking verbal property of what Thiesen & Weber call the nonfinite low tone is that verbs with this tone allow for object incorporation. Note that in the case of object incorporation, the object *obeeja* in (50) no longer receives accusative case.

(50) a. *obééjá-mu-ke* ***té****hmeé-be*
sheep-AN.PL-ACC care-CL:M.SG
'He cares for the sheep.'

[28] The imperative marker *d(i)=* is another preverbal subject proclitic. I do not include it in my analysis here since it inherently does not have cross-clausal coreference properties that are the topic of this dissertation.
[29] Seifart (2005: 43) also points this out for Miraña.

b. *obééjá-mú-**te**hméé-be*
sheep-AN.PL-care-CL:M.SG
'shepherd'

It may be the case that verbs with the 'nonfinite low tone' are in fact some kind of mixed projection with both nominal and verbal properties. Importantly though, these verbs, whether nonfinite or nominalized (or some combination of the two) are not analogous to the types of embedded verbs being analyzed in the following sections. I also have found no evidence of a verb with this tone being able to host a subject proclitic (they seem to be potentially limited to classifier subjects), nor have I found evidence of them carrying tense and aspectual information or appearing with adverbs that modify them.

There is also positive evidence that embedded complement clauses in Bora are finite clauses, in that tense can be expressed in complement clauses, whether by second-position clitics (51) or by suffixes (52).

(51) *mé=wajácú diityé=pe amóóbe-ke dóó-ne*
SAP=know 3.PL=REM fish-ACC eat.meat-CL:IN
'We know that they ate fish.'

(52) *wajpi tsá ílli-tyú-né ɨ=dsɨjɨvé-i-yó-ne*
man NEG fear-NEG-CL:IN 3COR=die-FUT-FRS-CL:IN
'The man is not afraid to die.'

There is, however, a distinction between main and embedded Bora clauses, with all embedded clause (non-nominalized) verbs being marked with a high tone on the first syllable of the verb, as described in section 2.1, and the verb often carrying the inanimate classifier *–ne,* as in (53b). This is true both of clauses that are traditionally called finite and nonfinite in other languages. This tonal marking and classifier are not characteristic of main clauses. Main clause verbs do not appear with this classifier, and the verb will only carry lexical or default tone, as in (53a), unless some other type of grammatical tone is also being marked on the verb.

(53) a. *wajpi majchó-hi*
man eat-PRED
'The man is eating.'
b. *wajpi imíllé i=májcho-ne*
man want 3COR=eat-CL:IN

'The man wants to eat.'

The distinction between main and embedded clauses will have consequence for the later analysis I will propose for the clitics presented above in Table 3 as a type of long distance anaphora.

### 2.7. Possessives in Bora

In Bora, there is a series of possessive pronouns which can be either free or bound. These are given in Table 4.

| Bound Possessive Pronouns | Free Possessive Pronouns |
|---|---|
| *ta-* "my" | *tahñe* "mine" |
| *di-* "your" (SG) | *dihñe* "yours (SG) |
| *me-* "our (incl.)" | *mehne* "ours (in.)" |
| *mu-* "of whom" | *muhne* "whose?" |
| *i-* "self's" | *ihñe* "self's" |

Table 4: Bora bound and free possessive pronouns (from Thiesen and Weber 2012: 247)

Notably, the free possessive pronouns are the bound possessive pronouns with the inanimate classifier added, in some cases with palatalization.

The 3COR anaphor mentioned in section 2.5.3, acting as a subject pronoun, seems to have properties that mirror the 3rd person 'self' possessive marker. Also of note, the SAP marker *me=* is the marker for the 1st person non-singular inclusive possessive pronoun. Thiesen and Weber's examples suggest that the default interpretation of an overt subject pronoun with the SAP marker is as exclusive, but the absence of the overt pronoun provides an inclusive plural reading (see the examples in (25-26) above), although they do not explicitly state this. Thiesen and Weber (2012) do not explain why the *me-* possessive marker in their table (my Table 4 above) can only be interpreted as inclusive, but it may be due to the absence of an overt personal pronoun accompanying the possessive pronoun, leading to the same inclusive reading as the SAP marker when it appears without an overt pronoun.

### 2.8. Basic Clause Structure of Bora

An analysis of preverbal subject proclitics in Bora necessarily relies on a basic understanding of the basic syntactic structure of Bora. Generally, Bora has a fairly free constituent order (Seifart 2015a), which makes an analysis of syntactic structure more challenging; but there are generalizations that can be made. Already it has been noted that there is a difference in structure

between main and embedded clauses, with embedded clauses always being verb final (54), and main clauses not necessarily having a fixed word order. This will lead to different analyses for main and embedded clauses.

(54) *imíllé-me* *añú-múnaá-du-re-juco* *i=ñeé-ne*
want-CL:AN.PL shoot-people-like-only-PRF 3COR=say-CL:IN
'They want to talk like white people.'

Possibly of greatest consequence to the analysis of subject proclitics in Bora is the clitic status of these morphemes. Thiesen & Weber (2012) point out several properties that lead to their analysis of these morphemes as clitics, namely: (i) nothing can appear between the subject proclitic and the verb, (ii) the proclitic and the verb cannot both bear a low tone due to the LLX constraint described in section 2.1 (indicating that the clitic and the verb are part of the same tonal phrase), (iii) the singular imperative marker *di-* (described further in section 2.9.1) and the 3COR marker *i=* both cause palatalization (described at the beginning of this Chapter) of a following consonant (55), indicating that the clitic and its host noun are part of the same phonological phrase, and (iv) the form of the singular imperative, whether *di-*, *d-*, or nothing, is dependent on the beginning sound of the host noun, as well as the number of syllables the host contains.

(55) *wajpíí-mu* *imíllé* *i=ímivyé-ne* *jaá* ***i=myéé****nu-ne*
man-PL want 3COR=complete-CL:IN house 3COR=make-CL:IN
'The men want to finish building this house.'

Bora generally is a head-final language (Berger & Seifart forthcoming), which is especially apparent considering the dependent-head order of the genitive construction (as described in section 2.1) and the use of postpositional phrases (56).

(56) *já* *pañe* *ijcyá-ne*
house inside COP-CL:IN
'It is in the house.' BORA (Thiesen and Weber 2012: 130)

The difference in structure between main and embedded clauses provides useful information for the structure of Bora as well. It is especially relevant that, in embedded clauses, the verb follows the object, whereas in main clauses this is not necessarily the case. Also in embedded clauses, because the verb must be final, adverbs also appear to the left of the verb (57).

(57) *ímí* *ditye* *cúwá-hajchíí* *tsá* *ditye* *chémé-ɨ́-tyu-ró-hi*[30]
well 3.PL sleep-COND NEG 3.PL be.ill-FUT-NEG-FRS-PRED
'If they sleep well, they are not likely to get sick.' (Thiesen & Weber 2012: 126)

With the verb following all the lower matter of the clause, I take this to mean that the verb has raised to a right-headed functional projection, which I take to be INFL (the T projection in some theories). The verb raising to this position also allows it to obtain future tense marking, which is realized as a suffix on the verb. This is also in line with the general OV nature of the language (Seifart 2015a).

(58) *ó=neé* *íñe* idioma Bora *tsá* *múijyú* *nɨ́jkevá-**i**-tyú-ne*
1SG=say this 'Bora language' NEG when end-**FUT**-NEG-CL:IN
'I say that this Bora language will never end.'

The position of the subject is then deducible from the position of the verb. Consider the example in (59).

(59) *múha* *mé=wajácú* *ámuha* *jaa-hañe* *me=méénu-ne*
1.PL SAP=know 2.PL house-VAR SAP=make-CL:IN
'We know that you (pl.) build houses.'

I will be arguing in this dissertation (in Chapter 4) that the SAP marker is an anaphor, which requires either a local or non-local antecedent. In both clauses in (59), the SAP marker on the verb requires a c-commanding antecedent. As such, the subject must move to a position higher than the verb in INFL. This is also in line with the generalization from (12) above that, with the exception of classifier subjects, Bora subjects cannot occur post-verbally. Subject movement to the Spec of INFL is common across languages, and appears to be the case here. Although the INFL head itself is right-headed, the Spec of INFL is left-branching. Assuming that the subject markers are in fact proclitics, they must head-adjoin to the verb as the verb raises.

Finally, it is relevant to mention a class of second position clitics in Bora, and how they can provide evidence for Bora syntactic structure. While the exact position of second position clitics in the

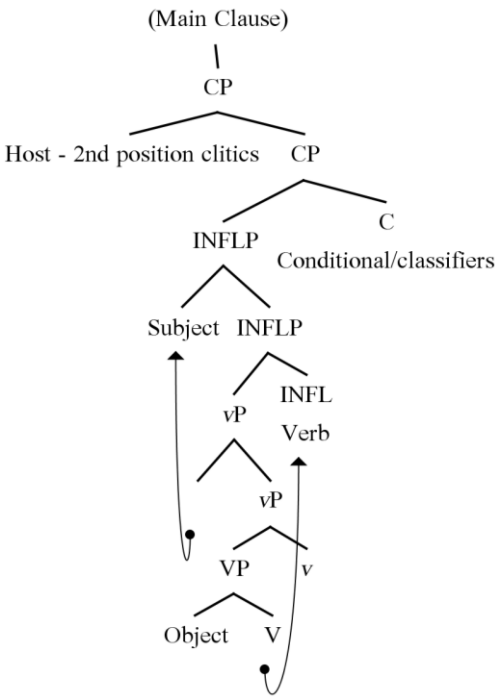
[30] It is not clear in this example why the overt pronoun is present in both the main and embedded clauses. Assuming that the subject of the embedded clause is coreferent with the subject of the main clause, we would expect to see the 3COR marker as the embedded clause subject.

syntax is a topic of debate, often dependent on the language of study, the appearance of other syntactic elements between the host of the second position clitic and the verb with the subject clitic implies movement of the clause-initial syntactic element to a position not adjacent to the subject. Bora has multiple second position clitics that relay tense (including recent and remote past), aspect, mood, and evidentiality. Notably, while future tense is conveyed by a suffix on the verb, recent and remote (60) past are conveyed as second position clitics. Thiesen & Weber (2012) describe second position clitics in Bora as following the first constituent of the clause. I assume for the sake of this analysis that the host of the second position clitic is in the C domain. As such, the element that the second position clitic attaches to will be hosted in the left branching Spec of CP, whereas elements like the conditional suffix (see (57) above) are hosted in the right-branching C head.

(60) *mé=wajácú* *diityé=pe* *amóóbe-ke* *dóó-ne*
SAP=know 3.PL=REM fish-ACC eat.meat-CL:IN
'We know that they ate fish.'

Assuming Bora to be generally right-headed with Specs left-branching, we expect embedded clauses to generally follow the schema represented in (61).

(61)



Turning now to main clauses, there is generally more freedom in word order. While an exact analysis of each possible permutation is not necessary for this analysis, one notable difference is that main clauses are not necessarily verb final.

(62) *o=ájtyumɨ́ wájpii-ke*
1.SG=see man-ACC
'I saw the man.'

(63) *o-ke wajpi ájtyumɨ*
1.SG man see
'The man saw me.'

(64) *diitye múúha-ke ájtyumɨ*
3.PL 1.PL-ACC see
'They saw us.'

The examples in (62-64) show three distinct word orders: SVO, OSV, and SOV. The data in (64) in fact already fall out from the same analysis that has been presented for embedded clauses. Following from this, the data in (63) are attainable by movement of the object to the left periphery. This may be to the Spec of a projection like TopP; however, I do not have data that this movement is related at all to topicalization. For present purposes, I refer to this phrase as $F_1P$.

(65) [$_{F1P}$ *o-ke* [$_{CP}$ [$_{INFLP}$ *wajpi* [$_{INFL}$ *ájtyumɨ* [$_{vP}$ *t* [$_{VP}$ *t t* ]]]]]]
1.SG man see
'The man saw me.'

In order to account for the data in (62), more movement is required. Because the object must follow the verb in this right-headed structure, the same movement seen in (65) must occur. Following the movement proposed there, additional remnant movement of the CP to a higher projection, which I simply refer to as $F_2P$ follows. These movements to $F_1P$ and $F_2P$ are not available to verb final embedded clauses.

(66) [$_{F2P}$ [$_{CP}$ *o=ájtyumɨ* ] [$_{F1P}$ *wajpii-ke* *t* ]]
1SG=see man-ACC
'I saw the man.'

It should be noted that I am not proposing motivations for these movements. Rather, I am using the available data to propose syntactic structure on which to base the remainder of this work. Of greatest consequence to the analysis of subject proclitics that I provide in Chapter 4 are the status of the morphemes in question as clitics, movement of the verb to the INFL head, and movement of the clausal subject (aside from proclitic and classifer subjects) to the Spec of INFL. I also do not take a position here on the properties of, structural position of, or co-occurrence restrictions of classifier subjects. Further research will be able to advance this topic further, including how Bora tonal and phonological phrases interact with and provide clues for syntactic structure.

### 2.9. (Lack of) Null Subjects

Bora generally does not allow null subjects. In most cases, clauses must have an overt expression of the subject of that clause, whether an overt noun or pronoun, a classifier subject, or one of a series of subject proclitics in embedded clauses.

(67) **wajpi imíllé [ májcho-ne ]*
man want [ eat-CL:IN ]
intended 'The man wants to eat.'

The sentence in (67) is ungrammatical because the embedded clause does not have an overt subject. The seeming exceptions to the generalization that all clauses must have overt subjects are some imperatives, some subject wh-questions, relative clauses, and certain conditionals.

#### 2.9.1. Imperatives

Recall from section 2.1 that imperatives in Bora are generally indicated by a proclitic *d(i)=*, where monosyllabic verbs take the full *di=* clitic (68a), polysyllabic verbs beginning with a vowel only carry the reduced *d=* clitic (68b), and polysyllabic verbs beginning with a consonant do not receive a proclitic (68c). Notably, while many of these do take an overt expression of the subject in the form of the imperative marker (68a-b, 68d), the polysyllabic verbs beginning with consonants (68c) do not. Recall that plural subject imperatives in Bora take the SAP as the subject marker, as in (68d).

(68) a. *dí=dyoó*
2SG.IMP=eat.meat.EMPH
'Eat (meat)!'
b. *d=ímibájchoó*

2SG.IMP=fix.EMPH
'Fix it!'

c. *májcho*
eat
'Eat!'

d. *mé=acúúve*
SAP=sit.down
'Sit down

Thiesen and Weber (2012:334, 336)

### 2.9.2. Wh- Questions

Wh-questions in Bora are expressed by wh-words which are fronted in the main clause. When the subject of a main clause is fronted as a result of wh-movement, there is no remnant of that subject attached to the verb. Rather, the only representation of the subject is the wh- element itself (69a). Compare this to the wh- fronting of an object in (69b).

In embedded clauses from which a *wh-* element has been extracted, the clause that the subject originated in often has a representation of the subject remaining, in both coreferent and disjoint referent cases (see the embedded clauses in (70)). When the embedded clause subject is disjointly referent with the main clause subject, the main clause subject is still expressed, while the embedded clause subject is marked by the 3COR marker (70a). If the subject being fronted is the coreferent subject of both the main and embedded clauses, the embedded clause subject is still marked by the 3COR marker (70b) (likely because the wh- element is still being moved from the main clause), while the main clause has no overt expression of the subject aside from the wh- element, the same as in (69a).[31] Because the proclitics in question express only the subject, wh- extraction from an embedded clause object (70c) does not leave any sort of overt remnant.

(69) a. *múha ahdó baajúri*
who buy yuca
'Who is going to buy yucca?'

b. *muu-cá tsiímene iité*
who-ACC child see
'Who did the child see?"

(70) a. *múha ú=itsúcunu i=cúwa-ne*
who 2SG=think 3COR=sleep-CL:IN
'Who do you think is sleeping?'

b. *múha imíllé i=májcho-ne amóme-ke*

[31] Interestingly, wh- extraction from an embedded clause without a coreferent subject also leaves a 3COR marker in the embedded clause. I discuss this and the consequences the 3COR marker has for my analysis in these instances in section 4.2.2.1.

who want 3COR=eat-CL:IN fish-ACC
'Who wants to eat fish?'

c. *ɨ̀ɨ́ná ú=imíllé u=májcho-ne*
what 2.SG=want 2.SG=eat-CL:IN
'What do you want to eat?"

### 2.9.3. Relative Clauses

Strictly speaking, relative clauses in Bora can also be expressed without an overt realization of the subject. Bora relative clauses, like all embedded clauses, are predicate final. The entire relative clause is marked with a classifier, which attaches to the verb and refers to the syntactic element being relativized. If the noun being relativized is overtly part of the main clause, as in (71b), it will generally precede the relative clause.

(71) a. *[ øi jóáa-ke úwááboo ]-be tsáá-hi*
John-ACC teach -CL:M.SG came-PRED
'The one whoi [ øi taught John] came.'

b. *dii-byei [ øi jóáa úwáábo ]-be tsáá-hi*
3-CL:SG.M John teach -CL:M.SG came-PRED
'The one whom [John taught øi] came.' (Thiesen and Weber 2012: 379)

Even when the noun being relatived is part of main clause preceding the relative clause, when it is coreferent with the subject of the relative clause, as in (72), the subject of the relative clause is not expressed with the 3COR proclitic on the verb. This may be because of a co-occurrence restriction with the classifier that is also attached to the verb.

(72) *áá-ne-ri diibye [ ø oohíí-bye-ke dsɨ́jɨ́ve-tsoo ]-be*
that-EVENT-INS 3.M.SG dog-CL:M.SG-ACC die-CAUS-CL:M.SG
*núcójpɨ̀ɨ́-ve-hi*
be.ashamed-SIN-PRED
'The one whoi [øi killed the dog ] became ashamed.' (Thiesen & Weber 2012: 384)

Overall, there is significant variation in relative clauses. Thiesen & Weber (2012: 379-394) provide a broad overview of the different types of relative clauses in Bora and their descriptive properties. Further theoretical work on Bora relative clauses is needed for a more robust analysis.

### 2.9.4. Conditionals

There also seems to be certain conditional clauses that do not have an overt expression of the subject. When the apodosis (consequence) of a conditional precedes the protasis (condition), then the subject of both clauses is overtly expressed (73a). However, when the protasis precedes the apodosis and the two have coreferent 3rd person subjects, the apodosis does not have an overtly expressed subject (73b). It is not apparent why this would be the case. Conditional clauses in Bora require further exploration.

(73) a. *wajpi* *májcho* *ímí* *i=táává-hajchíí*
man eat well 3COR=hunt-COND
'The man will eat well if he hunts.'
b. *wajpi* *táává-hajchíí* *májcho* *ímí*
man hunt-COND eat well
'If the man hunts, he will eat well.'

## 2.10. Wrap-up

This chapter has introduced several basic concepts that are necessary background for the analysis of Bora proclitic preverbal subjects in subsequent chapters. The discussion of reflexive and reciprocal objects in section 2.4 highlight how these structures, often expressed by anaphors in other languages, are expressed differently in Bora. Chapter 3 provides a more in-depth discussion of how reflexives specifically are expressed in other languages, including German, which was alluded to in section 2.5.3 for its use of a SELF anaphor. For the following Chapters, it is important that what Bora has for the 3rd person coreferent clitic is not a SELF marker in this way. The data provided in sections 2.3 and 2.5 will be especially important in establishing an analysis in Chapter 4, especially considering the syntactic structure established in section 2.8.

# CHAPTER III

## Theories of Anaphora and Control

This chapter provides discussion of previous theories related to co-reference phenomena as a background for the analysis of preverbal subject proclitics in Chapter 4. I focus first on theories of binding, in which I give basic information about canonical and more recent approaches to binding theory. Initially, I discuss the binding principles put forth in Chomsky (1981), and subsequent work building on those principles. While I primarily focus on canonical theories of binding, I also briefly describe some alternatives and additions to the canonical binding theory, which account for the same data without necessarily appealing to the concept of being 'bound', but rather appealing to the concept of reflexivity. These include logophoricity and logophoric centers and an alternative set of binding principles laid out in Reinhart and Reuland (1993).

In section 3.1, I focus on the study of anaphora, first in terms of variation across different languages in how they express anaphors, then in terms of canonical theories that have been developed to explain the properties of anaphors. I also specifically discuss long distance anaphors and how they have been described in languages that have been claimed to have them. Relatedly, I briefly discuss how logophoricity has been discussed in the literature on long distance anaphors. In section 3.2, I discuss both pronominals and R-expressions in terms of binding theory as well. Section 3.3 explores how binding domains have been described for anaphors and pronominals. In section 3.4, I discuss Reinhart & Reuland's alternative approach to binding theory and how their theory accounts for the licensing of reflexives. Finally, section 3.5 examines different theories of control, especially movement theories (e.g. O'Neill 1997, Hornstein 1999) and agreement theories (e.g. Landau 2015).

### 3.1. Anaphora

Many early theories of anaphors and pronominals analyze nominal expressions in terms of them being [+/-anaphoric] and [+/-pronominal] (see e.g. Chomsky 1981, 1986). Under this system,

lexical anaphors are [+anaphoric, -pronominal], lexical pronominals are [-anaphoric, +pronominal], and lexical R expressions are [-anaphoric, -pronominal]. Additionally, null elements could be categorized using this binary system as well, with NP trace [+anaphoric, -pronominal], *wh*-trace [-anaphoric, -pronominal], PRO [+anaphoric, +pronominal], and *pro* [-anaphoric, +pronominal]. This theory, while prevalent in early days of Government and Binding theory, has been replaced with the onset of the Minimalist Program (MP) (Chomsky 1993 and subsequent work).

"Anaphoric" and "pronominal" are no longer taken to be primitive features in the lexicon. Once movement of syntactic items in syntactic structure came to be accepted, items like NP traces and *wh*- traces came to be interpreted in terms of their binder. Therefore, the difference between these two types of traces, for example, did not have to do with whether the whether the item was [+/-anaphoric] in the lexicon. Rather, anaphoric and pronominal properties fall out from the syntax, including via c-command (described above), coindexation (though see footnote 33), and movement.

Anaphors are referentially deficient nominal expressions, which rely on an antecedent to obtain their reference in a restricted environment. Chomsky (1981) provides an early and foundational approach to formalizing the properties of binding, which led to the formalization of three Binding Conditions. These conditions predict the distribution of pronouns[32] and noun phrases in language. While the exact formalization of the Binding Conditions has been revised multiple times to account for the binding properties of the languages of the world, a fairly conservative version of them is given in (1).

(1) Binding Conditions
Condition A: An anaphor must be bound within its binding domain.
Condition B: A pronominal must be free within its binding domain.
Condition C: An R-expression must be free.

The Conditions in (1) make reference to elements being bound and free.

(2) Binding

[32] The use of the term "pronoun" varies in the literature. I take the term to mean any referentially deficient nominal expression, to include both anaphors and pronominals.

a. An $NP_1$ binds another $NP_2$ if and only if $NP_1$ and $NP_2$ are coindexed[33] and $NP_1$ c-commands $NP_2$.
b. Any NP that is not bound is free.

(3) C-command

C-command is a relationship between nodes in a syntactic representation, where Node $N_1$ c-commands Node $N_2$ if and only if neither node dominates the other, and the first branching node that dominates $N_1$ also dominates $N_2$.

There are various formal theories of anaphora and how different types of anaphors should be analyzed. These range from approaches of various typologies for different types of anaphors (Déchaine and Wiltschko 2017 propose a five-way distinction of different types of reflexives), to approaches which attempt to reduce anaphora to a system such that anaphors all behave similarly enough to be considered a single formal linguistic unit (Safir 2014). Although I will not consider all these theories in particular below, I now turn to discussing variation in how anaphora is expressed in languages that display a variety of different kinds of anaphora. The languages I describe below provide a relevant range of the types of anaphora present in the world's languages, and show the relevant properties I will be exploring for the Bora subject proclitics.

### 3.1.1. Variation in Anaphora

A generic formal definition for anaphora is hard to come by, except that the meaning (reference) of a given anaphor is dependent on its antecedent, anaphors in their simplest form are underspecified for their phi features, and anaphors are conditioned by Binding Condition A. It is here that anaphors also differ from pronominals, in that pronominals are specified for their phi features. Under these criteria, many languages have expressions that are anaphoric in nature, but the anaphoric behavior of these expressions across languages is not consistent.

This also raises an important distinction between Simplex Element (SE) anaphors and complex anaphors. An SE anaphor is the anaphor in its simplest, monomorphemic form, which is underspecified for phi features. Complex anaphors (also called SELF anaphors), on the other hand, are morphologically complex, often combining the pronominals with elements like *self*, body parts,

[33] Throughout this dissertation, I refer to coindexation to denote co-identity of the reference of nominal elements, and I mark identity in examples with indices. Within the Minimalist framework, the Inclusiveness Condition does not allow for any new features to be introduced to the syntax, and the Interpretability Condition does not allow lexical items to have any features that are not interpretable at the interfaces (Chomsky 1995, 2000). This prohibits indices as a part of the syntax. I use indices throughout purely for ease of exposition and do not take them to be present in the syntax.

etc. (Reuland 2001). The fact that complex anaphors have pronominals as part of their morphological makeup means that these elements are not devoid of phi features in the same way that SE anaphors are, but rather *are* specified for phi features (e.g. *himself* and *herself* in English). Although complex anaphors often have a pronominal as part of their makeup, they have distributional properties of anaphors when it comes to the Binding Conditions. However, as will be discussed further in section 3.3, SE anaphors, despite being anaphors, exhibit some properties of pronominals when it comes to the domain they are bound in.

A commonly noted property of complex anaphors is that they have a reflexivizing property which is not shared by SE anaphors (this will be especially important when considering Reinhart & Reuland's (1993) alternative to Binding Theory discussed in section 3.4, where it will be formally defined). Additionally, SE anaphors have a subject orientation property, such that they can only be bound by subjects. This property of SE anaphors will be important when considering the binding properties of the Bora preverbal subject proclitics in Chapter 4.

Even across Germanic, there is a variety of anaphoric systems, as outlined extensively in Everaert (1986) and Reuland (2011), among others. Dutch uses a three-way system, distinguishing SE anaphors, complex (SELF) anaphors, and pronominals. Dutch uses the SE anaphor *zich* (4a). In Dutch, as described earlier, the SE anaphor is underspecified for gender and number, only appearing in the third person, without distinguishing between singular and plural. The corresponding complex anaphor in Dutch is *zichzelf* (4b). The choice between the SE and complex anaphors has to do with whether the predicate is inherently reflexive (this is important for Reinhart & Reuland's (1993) theory of anaphors, and is discussed in detail in section 3.4). The environment that licenses these anaphors does not allow for a pronominal (4c). The predicate in (4b) is inherently reflexive, allowing for the complex anaphor in object position, while (4a), which is not inherently reflexive, uses the SE anaphor.

(4) a. *Willem$_i$* *schaamt* *zich$_i$*
William shames SE
'William shames himself.'

b. *Willem$_i$* *bewondert* *zichzelf$_i$*
William admires SEself
'William admires himself.'

c. **Willem$_i$* *schaamt* *hem$_i$*
William shames him

‘William shames himself.’ DUTCH (Reuland 2011: 62-63)

To understand the complex anaphor’s distribution in Dutch compared to the simplex anaphor, it is helpful to compare Dutch to other Germanic languages.

While German and Dutch have many similarities with respect to anaphoric and pronominal elements, there are some key differences, not least of which being that German lacks the complex form found in Dutch. In non-dative positions, German only uses the SE anaphor *sich* (5), syncretizing the Dutch three-way distinction into a binary distinction between the SE anaphor and pronominals.

(5) *Max$_i$ hasst sich$_i$*
Max hates himself
‘Max hates himself.’ GERMAN (Reuland 2011: 274)

Although it appears that German has only a two-way distinction between pronominals and anaphors, Reuland (2011) points out that there are environments in which German *sich* can be stressed and others where it cannot, and these environments seem to line up with the environments of the Dutch complex anaphor and SE anaphor respectively, indicating a potential further split in German anaphors.

Frisian, also unlike Dutch, uses a two-part anaphoric system, including an anaphor (*himsels*) and a pronominal (*him*). However, the distribution of these elements does not match that of English. Frisian has locally bound 3rd person pronominals. These are visible in lexically reflexive verbs, wherein the predicate inherently licenses a reflexive interpretation. Consider the English example in (6a), where even without the overt anaphor, the verb *wash* can be interpreted reflexively, which is not the case in (6b) with the verb *admire*.

(6) a. *John$_i$ washed Ø$_i$ / himself$_i$.*
b. *John$_i$ admired *Ø$_i$ / himself$_i$.*

In Frisian, a distinction similar to English holds, but rather than the lack of a pronoun that English shows in (6a), Frisian uses the pronominal form as the object of lexically reflexive verbs (7a). Compare this to the verb *admire* in (7b) which is not lexically reflexive, but uses the anaphoric pronoun to achieve a reflexive interpretation.

(7) a. *Willem$_i$* *wasket* *him$_i$*
William washes him
'William washes himself.'
b. *Willem$_i$* *bewûnderet* *himsels$_i$ / *him$_i$*
William admires himself / him
'William admires himself.' FRISIAN (Reuland 2011: 269)

In Frisian generally, bound pronominals like those in (7a) have the same distribution as the simplex anaphor *zich* in Dutch.

Additionally, in Exceptional Case Marking (ECM) constructions, Frisian allows both bound pronominals and complex anaphors to act as the embedded subject.

(8) *Jan$_i$* *fielde* *him$_i$ / himsels$_i$* *fuortglieden*
Jan felt him slip-away
'Jan felt himself slip away.' FRISIAN (Reuland 2011: 269)

The analysis of these different types of anaphors (as well as the distinction between them and pronominals) lays the basis determining whether the Bora preverbal subject proclitics should be analyzed as one of these types of anaphors. In order to establish whether Bora proclitics behave like one of these types of anaphors, we will look at whether the syntactic properties of the proclitics match the syntactic properties of those anaphors. I now turn to the formal properties of long distance anaphora.

### 3.1.2. Long Distance Anaphora

Of particular interest for this analysis of Bora will be long distance anaphora. This refers to anaphors which are bound outside of a local domain, and instead have a non-local antecedent. While long distance anaphors still meet the qualifications of dependence on an antecedent and lack of phi features which characterize anaphors, they differ from reflexive anaphors in that reflexive anaphors must be arguments of the same binding domain as their antecedents. Long distance anaphors, by definition, are not subject to this same restriction, appearing in a non-local domain to the antecedent.

Huang (2000) gives examples of long distance anaphora across several typologically distinct languages, and in several different environments: binding out of an NP, out of a small clause, across an infinitival clause, across a subjunctive clause, across an indicative clause, across sentence

boundaries, and across turns in a conversation. Notably, not all languages that exhibit long distance anaphors allow for them to be bound in all of these different environments, though there appears to be a hierarchy whereby if a language allows binding of a long distance anaphor in one environment, it will be allowed in certain others. For example, if a language allows long distance binding of an anaphor across a subjunctive clause, it would also allow for long distance binding across an infinitival. For details on this hierarchy, see Huang (2000: 93).

Long distance anaphora has been described in several languages of the world, and the properties of anaphors in each language are not identical. However, Cole, Hermon, & Huang (2006) cite several properties that many languages exhibiting long distance anaphora have been shown to have in common, namely (i) the anaphoric forms are monomorphemic, (ii) the antecedents that the anaphors take are subjects (or at least c-command the anaphor), and (iii) the occurrence of long distance anaphora is often restricted to environments in which the anaphor and antecedent must occur in a specific domain, such as infinitivals or subjunctives. A fourth property that has been shown to hold for some languages is what has been analyzed as a Blocking Effect (Cole, Hermon, & Sung 1990), which prevents binding of a long distance anaphor if a subject of a different person intervenes between the anaphor and its antecedent.

The monomorphemicity generalization can be seen especially when comparing long distance anaphors (9a) to their local anaphor counterparts (9b), as seen in Italian. Notably, this generalization is for SE anaphors, since SELF type anaphors are necessarily complex.

(9) a. *credo* *che* *Mario*$_i$ *sostenga* *che* *tu* *abbia* *parlato* *de* ***se***$_i$
believe.1SG that Mario claims that 2.SG have spoken of self
*e* *della* *sua* *famiglia* *in* *TV*
and of.the his family on TV
'I believe that Mario claims that you spoke about him and his family on TV.'

b. *Gianni*$_i$ *pensava* *che* *quella* *casa* *appartenesse* *ancora* *a* ***se*** ***stesso***$_i$
Gianni thought that that house belong still to self self
'Gianni thought that that house still belonged to him.'

(Giorgi 1984)

Cole, Hermon, & Huang's second claim, that the antecedents of anaphors are subjects, already holds in the examples given in (9). They argue, however, that based on examples in Mandarin Chinese, objects in preverbal position (10b) can serve as antecedents to long distance *ziji*, whereas

postverbal objects (10a) cannot. They use this as evidence that the restriction on the antecedent of a long distance anaphor is not one of subjecthood, but rather is reduceable to c-command.

(10) a. *Wangwu$_i$ shuo Zhangsan$_j$ zengsong gei Lisi$_k$ yipian guanyu* ***ziji***$_{i/j/*k}$
Wangwu says Zhangsan give to Lisi one about self
*de wenzang*
DE article
'Wangwu says that Zhangsan gave an article about him/himself to Lisi.'

b. *Zhangsan$_i$ yiwei Lisi$_j$ hui ba Xiaoming$_k$ dai hui* ***ziji***$_{i/j/k}$
Zhangsan thought Lisi will BA Xiaoming take self self's
*de jia*
DE home
'Zhangsan though Lisi would take Xiao Ming back to his home.'

(Cole and Wang 1996)

The third claim by Cole, Hermon, & Huang is that long distance anaphora is restricted to certain types of environments, such as infinitival or subjunctive clauses. Icelandic clearly shows long-distance anaphora, as in (11), where the embedded SE *ser* can be bound by *Jón* in a higher clause, with both *Maríu* and *mig* as intervening NPs, or by *Maríu* with *mig* intervening. Notably, all of the NPs intervening between the anaphor and its antecedent are subjects of infinitive verbs. The Latin example in (12) shows a similar long distance relationship between the anaphors (labeled REFL2) and their antecedent (*patres conscripti*), with the other intervening subjects serving as subjects of subjunctive clauses. The relevant anaphors and their antecedent are bolded in (12).

(11) *Jón$_i$ sagði [Maríu$_j$ hafa$_{(inf.)}$ látið [mig þvo$_{(inf.)}$* ***sér***$_{i/j}$ *]]*
John said Mary have made me wash SE
'John said that Mary had made me wash him/her.' ICELANDIC (Reuland 2006d: 549)

(12) ***(patres conscripti)*** *legatos miserunt qui a rege peterent*
senators delegates.ACC sent.3PL who from king ask$_{subj}$
[ *ne inimicissimum* ***suum*** *se-cum haberet* ***sibi*****-*** *que dedere* ]
that.not biggest.enemy REFL$_2$ REFL$_1$-with keep$_{subj}$ REFL$_2$-and give$_{subj}$
'The senators$_i$ sent delegates to ask the king$_j$ not to keep their$_i$ biggest enemy$_k$ with him$_j$ and to give [him$_k$] to them$_i$.'

LATIN (Cornelius Nepos, from Hagège 1974: 289, gloss and translation Reuland 2006b: 2)

Cole, Hermon, & Huang's fourth observation, though not necessarily a diagnostic for long distance binding, is the existence of a Blocking Effect, which prevents binding of an anaphor by a long

distance antecedent if a subject of a different person intervenes. This is shown in (13) in Mandarin Chinese.

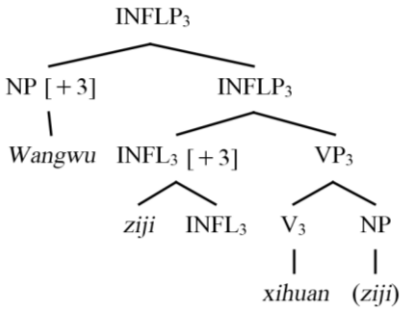
(13) a. *Zhangsan*$_i$ *renwei* *Lisi*$_j$ *zhidao* *Wangwu*$_k$ *xihuan* *ziji*$_{i/j/k}$
Zhangsan think Lisi know Wangwu like self
'Zhangsan thinks Lisi knows Wangwu likes self.'

b. *Zhangsan*$_i$ *renwei* *wo*$_j$ *zhidao* *Wangwu*$_k$ *xihuan* *ziji*$_{*i/*j/k}$
Zhangsan think I know Wangwu like self
'Zhangsan thinks I know Wangwu likes self.' (Cole, Hermon, & Huang 2006: 44)

(13a) contains only 3rd person subjects, which allows the anaphor *ziji* to take any of the 3rd person subjects as its antecedent, either locally or long distance. This contrasts with (13b), where a 1st person subject is intervening between the anaphor and the main clause subject *Zhangsan*. This disallows *Zhangsan* as a possible antecedent for the anaphor *ziji*. This effect, though it has been noted for Mandarin and Malayalam, has not been noted in other languages with long distance reflexives, e.g. Kannada and Icelandic. Cole, Hermon, & Huang suggest that one reason for the existence of the Blocking Effect in a language like Mandarin is the lack of verb agreement (see also e.g. Cole & Wang 1996).

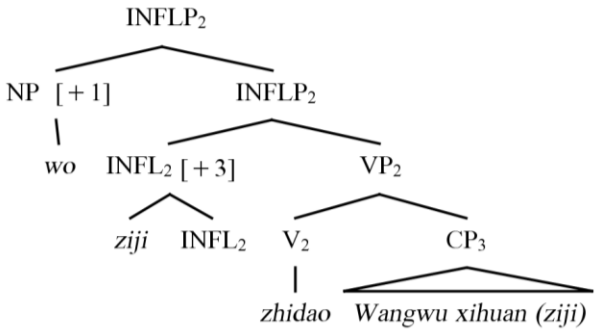
The (lack of) agreement analysis for the Blocking Effect is argued by Cole, Hermon, & Huang to be related to the presence or absence of person features in INFL. In Mandarin, agreement between the Spec and head of INFL is vacuously satisfied (i.e. there are no base-generated person features in INFL), whereas languages with verb agreement will have feature agreement (in this case, person features) between the Spec and head of INFL. While a language with verb agreement undergoes agreement between the element in Spec and the head, INFL in a language like Mandarin will percolate features upward. Consider the following example, adapted from Cole, Hermon, & Huang (2006: 45) building the sentence in (13) above from the most embedded clause, upwards.

(14)



In (14), *ziji* has covertly adjoined to the head of INFL with the 3rd person features it was generated with, and in doing so has percolated its features up, and thus has valued INFL as 3rd person. Spec-head agreement checks the features of the INFL head against that of *Wangwu* and finds no issue, allowing *Wangwu* as a well-formed antecedent for *ziji*.

(15)




Having agreed in features with the Spec of INFL in $CP_3$, *ziji* continues to raise into the next highest clause, adjoining this time to $INFL_2$. In doing so, the 3rd person features percolate up to the head of $INFL_2$, valuing it with 3rd person features. This does not allow Spec-head agreement to carry through with the 1st person features of *wo*. As such, *wo* is not a valid antecedent for *ziji*. Because further movement would lead to ungrammaticality, movement through $INFL_2$ is blocked.

In a language with verb agreement (such as the Romance languages), INFL is already valued by the time a reflexive could raise. As such, when the reflexive adjoins to the head of INFL, it is the head's features that percolate up rather than the reflexive's features via the Feature Percolation Principle (see Cole, Hermon, & Huang 2006: 45 for a formal definition).

If the 3rd person features continue to percolate up, then agreement will be possible at each INFL node along the way, provided no intervening element blocks the reflexive from continuing to raise. If, however, there is an $INFL_2$ head with different person features from the more embedded $INFL_1$ head, then further agreement past that $INFL_2$ projection will not be possible, such as the structure shown in (15). This will be relevant when considering whether Bora preverbal subject proclitics behave like Mandarin long distance anaphors.

A further observation about long distance binding of anaphors by Huang (2000) is that long distance reflexives tend to be referentially optional. The pronoun in (16) in Inuit can either be interpreted as a long distance anaphor or a locally bound anaphor in one case, or as a long distance

anaphor or a pronominal in another case. The form *immi* in (16a) can be either locally bound by *Pavia*, or bound at a distance by *Kaali*. In (16b), however, the form *taassu* can either be bound at a distance, or be a free form, referring outside the sentence.

(16) a. *Kaali*$_i$ *uqar-p-u-q* *Pavia*$_k$ *immi-nit*$_{i/k}$ *angi-nir-u-sinnaa-nngi-tsu-q*
Kaali say-IND-3SG Pavia self-ABL big-CMP-be-can-NEG-PRT-3SG
b. *Kaali*$_i$ *uqar-p-u-q* *Pavia*$_k$ *taa-ssu-managa*$_{i/m}$ *angi-nir-u-sinnaa-nngi-tsu-q*
Kaali say-IND-3SG Pavia DEM-3SG-ABL big-CMP-be-can-NEG-PRT-3SG
'Kaali said that Pavia couldn't be taller than self/him.'
INUIT (Huang 2000: 94, citing Bittner 1994: 147)

It should also be noted that all of these tendencies are exactly that: tendencies. They should not necessarily be taken to be universals of long distance anaphora.

Approaches to analyses of long distance anaphora are varied, but a common starting point for many analyses begins with Bouchard's (1984) proposal that SE anaphors are underspecified for phi-features, and thus need feature specification in order to be interpreted at LF. In order to receive this specification, there is covert head movement of the SE anaphor to AGR (located in INFL in later theories) at LF.

In order to account for long-distance anaphora, Reinhart and Reuland (1991) propose a pied piping analysis, whereby there is covert movement of the SE anaphor to form a constituent with the verb (this is seen in (14-15) above). From there, the verb/SE anaphor complex raises further to the functional projection INFL (the tense head in some approaches), then to a higher Verb Phrase if the anaphor is not yet able to be specified for its phi features, and so on up until the anaphor's phi features are valued. Taking the example from (11) above, the verb/SE anaphor will continue to raise covertly from its base position until it reaches the INFL of the matrix clause in *Jón sagði*.

### 3.1.3. Logophoricity

Not all types of long distance binding fit the theory proposed by Reinhart & Reuland (1991) of anaphors raising out of non-finite embedded clauses due to an underspecification of phi features. There are other anaphors that appear at a distance from their antecedent with finite clause boundaries intervening between them. These pronouns, called logophors, have additional restrictions on their appearance as well, which have been noted since their original descriptions in Hagège (1974) and Clements (1975). Logophoric pronouns, as described in Clements (1975), are

used in reportive contexts in which the attitude of an individual is being expressed rather than the attitude of the speaker or narrator. Clements notes that logophors can be cross-linguistically characterized with the following criteria (Clements 1975: 171-172).

(i) logophoric pronouns are restricted to *reportive contexts* transmitting the words or thought of an individual or individuals other than the speaker or narrator;

(ii) the antecedent does not occur in the same reportive context as the logophoric pronoun;

(iii) the antecedent designates the individual or individuals whose words or thoughts are transmitted in the reportive context in which the logophoric pronoun occurs.

As such, the antecedent of the logophoric pronoun is the individual whose attitude is being expressed, and that antecedent does not occur in the same reportive context as the logophoric pronoun. There is also a strong tendency in languages that display logophoric pronouns for the antecedent of those pronouns to be 3rd person, and often singular. An example of a logophoric pronoun is given in (17b), as compared to the non-logophoric correlate in (17a).While the antecedent of the 3rd person pronoun in (17a) could be *Kofi* or another individual, the logophoric pronoun in (17b) must take *Kofi* as its antecedent, since *Kofi* is the one reporting the leaving. It is *Kofi*'s attitude being expressed, not that of the person making the utterance.

(17) a. *Kofi be e dzo*
Kofi say 3SG leave
'Kofi$_i$ said that he$_{i/j}$ left.'
b. *Kofi be yè dzo*
Kofi say LOG leave
'Kofi$_i$ said that he$_{i/*j}$ left.' EWE (Pearson 2015: 78, adapted from Clements 1975)

Logophors differ from other types of anaphors in that their interpretation is relative to what is known as a logophoric center. That is, each event in an utterance can be characterized with respect to the event's place and time, as well as its participants. Unless specific event times and places are specified in the utterance, those variables are interpreted relative to the utterance. Importantly, the time and place of the utterance do not necessarily align with the time and place of the event. Additionally, the times and places of utterances may be different for events that occur in different clauses, such that the time and place of the event denoted in the matrix clause, and the time and

place of the event denoted in an embedded clause may be different, and might also differ from the time and place of the utterance.

Notably, Sells (1987) suggests that, rather than a single unified notion of logophoricity, there are three roles at play, namely SOURCE (the one making the report), SELF (the one whose mental state is being reported), and PIVOT (the one with respect to whose spatial and/or temporal location the report is being evaluated). Each of these roles describes a different context in which predicates of different types can be interpreted logophorically (though see Reuland 2006b for arguments that sensitivity to PIVOT may not be related to the notion of logophoricity).

Logophors, as a subtype of anaphors, differ from other anaphors in that they are not syntactically bound. Reuland (2006b) argues that this is indicative that there is no requirement for anaphors to be syntactically bound. The syntactic mechanism will always look for a binder for reasons of economy. However, in the case that an anaphor is not syntactically bound, its referent will be determined by discourse context, as is the case for Icelandic *ser* in (18).

(18) *María var alltaf svo andstyggileg. þegar Ólafur$_j$ kæmi segði hún*
Mary was always so nasty when Olaf came said she
*ser$_{i/*j}$ áreiðanlega að fara…*
himself certainly to leave
'Mary was always so nasty. When Olaf came, she would certainly tell himself [the person whose thoughts are being presented – not Olaf] to leave…' ICELANDIC (Thráinsson 1991: 58)

### 3.2. Pronominals and R-expressions

As mentioned above, canonical binding theory proposes that pronominals are governed by Binding Condition B. Pronominals differ from anaphors most notably in that their interpretation is not necessarily dependent obligatorily on an antecedent. Since they are not necessarily dependent on other elements for their interpretation, they are able to (and must) appear free of any binding within their local domain. Condition B, as formalized in canonical binding theory, is given again in (19).

(19) **Condition B**

A pronominal must be free in its binding domain.

The motivation for a difference in binding principles between anaphors and pronominals is characterized by the differences in (20a-b).

(20) a. *John$_i$ likes himself$_i$/him$_{*i/m}$.*
b. *John$_i$ said that Bill$_k$ likes himself$_{*i/k}$/him$_{i/*k/m}$.*
(21) a. *Mary's$_i$ sister$_k$ likes herself$_{*i/k}$.*
b. *Mary's$_i$ sister$_k$ likes her$_{i/*k/m}$.*

It is relevant that, in (20a), *him* cannot take *John* as an antecedent. Following the definitions given in (2) and (3) above, since *John* c-commands *him*, the only way to keep the pronominal *him* free in (20a) is to have *him* refer to a person other than *John* (i.e. *John* and *him* cannot be coindexed). By the same logic, since we saw in the previous section that (simply put), anaphors must generally be bound, the co-reference of *himself* with *John* is required in (20a).

In (20b), however, the anaphor and pronominal in question are appearing in a different clause than their intended antecedent *John*. In this case, the anaphor and pronominal have different possible antecedents within the same sentence. *Bill* is able to bind *himself*, following Condition A, but *John* is not able to. If binding consists purely of c-command and coindexation within a local domain, and *John* in (20b) both c-commands and is intended to be coindexed with the anaphor, then *John* must be outside the local domain of the anaphor. This leaves the pronominal *him* free within its binding domain when it is coindexed with *John*, but leaves the anaphor *himself* bound within its binding domain when it is coindexed with *Bill*.

In (21a), *herself* can only be bound by *sister*, since *sister* is the only noun that c-commands the anaphor. Although *Mary* is still within a local domain, the embedded nature of *Mary* within the subject NP disallows *Mary* from binding the anaphor. For this same reason, however, *Mary* is able to be conindexed with the pronominal *her* in (21b), since this still leaves *her* free within the local binding domain. Of course, coindexation of the pronominal with a referent outside the sentence also leaves the pronominal free in both (20-21).

Finally, Binding Condition C makes reference to R(eferential)-expressions, including all DPs that are not classified as anaphors or pronominals. By nature, R-expressions cannot have an antecedent, meaning that they (generally) cannot be bound by an element in a higher syntactic position (though see Nediger (2017) for discussions of violations of this Condition). For this reason, Condition C is given as follows.

(22) **Condition C:** An R-expression must be free.

Because this dissertation focuses primarily on anaphors and pronominals, I do not go further into Condition C here.

## 3.3. Binding Domains

### 3.3.1. Anaphor Binding Domains

Up until now, I have used the term 'binding domain' loosely without giving a formal definition. Research on binding over the years has uncovered several restrictions on what can and cannot act as a binding domain. Although the clause level may seem to be an intuitive domain for binding, an illustrative example comes in Exceptional Case Marking (ECM) structures, in which there is an NP which is thematically the subject of an embedded clause, but behaves like the object of a higher clause, most notably through accusative case marking, though also through processes of passivization.

(23) a. *The judge$_i$ believes **him**$_{*i/j}$ to be guilty.*
b. *The suspect$_i$ believes **himself**$_{i/*j}$ to be innocent.*
c. *The prosecutor$_i$ believes the suspect$_j$ to hate **himself**$_{*i/j}$.*

There is an interesting juxtaposition between (23b) and (23c). In (23b), the exceptionally marked subject is an anaphor which is bound by the subject of the main clause, indicating that the main clause subject is within the binding domain of the anaphor. Similarly, the embedded clause object in (23c) is bound by the exceptionally marked subject *the suspect*, indicating that those two NPs are also in the same binding domain. However, (23c) also shows that *the prosecutor* cannot bind *himself*, indicating that the binding domain for *himself* cannot extend to the whole sentence. This provides evidence that there are two overlapping binding domains for ECM constructions.

(24) **[** Subj V **[** ECM **]** V Obj **]** (modified from Büring 2005: 47)

Importantly, the ECM subject in (24) appears to belong to two different binding domains. However, we saw in (23b) that an anaphor in the ECM position is bound by a higher clause subject, allowing for a generalization that anaphors must be bound in a domain that includes their case assigner. This is seen most clearly in (23c), where the embedded object anaphor is bound by the embedded clause subject, but cannot be bound by the main clause subject. Meanwhile, the anaphor

in (23b), in the same subject position as the binder in (23c), *can* be bound by the matrix clause subject, leading to the schema presented in (24).

There is also evidence that binding domains are present for nominals as well, shown by so-called "picture NPs".

(25) a. *John saw [ a picture of himself/*him ]$_{NP}$.*
b. *John saw [ Mary's picture of ??himself/him ]$_{NP}$.* (Büring 2005: 50)

The anaphor *himself* in (25a) is bound by a subject outside of the NP, indicating a wider binding domain than (25b), in which *himself* is not bound by *John*. The difference between the NPs in these two examples is that (25b) contains a subject within the NP, while (25a) does not.

At this point, we can reformulate Binding Condition A by identifying the binding domain for anaphors as follows:

(26) Binding Domain: An anaphor NP must be bound within the smallest[34] XP that contains the NP, the NP's case assigner, and an accesible subject.[35]
(27) Accessible Subject: α is accessible to β if and only if β is in the c-command domain of α, and assignment to β of the index of α would not violate the i-within-i condition.
(28) i-within-i Condition: [$_{\gamma}$…δ…] where γ and δ bear the same index.
(Reuland 2006a: 265)

The definition given in (26) becomes less straightforward when considering long distance anaphors. Huang (2000) points out that there have traditionally been two ways of accounting for long distance anaphors in a generative framework: (i) claiming that long distance anaphors are not anaphors, but rather are pronominals, and are thus not subject to Binding Condition A, and (ii) modifying the Binding Conditions such that long distance anaphors are accommodated.

An analysis claiming that anaphors are, in fact, pronominals does not seem to hold since, as Huang (2000) points out, the fact that they would then be subject to Condition B does not obtain. Condition B does not place a restriction on where the pronominal should be free rather than where

[34] I intend "smallest" here to have an intuitive reading. In set theoretic terms, this defintion should be taken to mean the set XP which contains the NP, its case assigner, and a subject, and contains no other set YP which also contains the NP, its case assigner, and a subject.
[35] This definition is adapted from Büring (2005: 66), though Büring notes that languages differ as to whether the relevant requirement for the XP is that it must contain a subject, a coargument, a finite clause, or the entire sentence. The notion of an accessible subject, defined in (27), comes from Reuland (2006a).

it should not be bound. This makes it very hard to explain examples like (13) above, repeated below, which show restrictions on which antecedent can bind the pronouon *ziji*.

(29) a. *Zhangsan*$_i$ *renwei* *Lisi*$_j$ *zhidao* *Wangwu*$_k$ *xihuan* *ziji*$_{i/j/k}$
Zhangsan think Lisi know Wangwu like self
'Zhangsan thinks Lisi knows Wangwu likes self.'

b. *Zhangsan*$_i$ *renwei* *wo*$_j$ *zhidao* *Wangwu*$_k$ *xihuan* *ziji*$_{*i/*j/k}$
Zhangsan think I know Wangwu like self
'Zhangsan thinks I know Wangwu likes self.' (Cole, Hermon, & Huang 2006: 44)

Assuming a theory in which long distance anaphors are treated as pronominals, *ziji* in (29) should fall under Condition B, and thus should be free. This is not the case, as in (29a), *ziji* can potentially be bound by any of the 3rd person subjects in the sentence. Such a proposal also cannot account for the fact that *ziji* and some other long distance anaphors can also be bound locally, making them at least partially subject to Condition A.

A theory that long distance anaphors are actually pronominals would need to find a way to restrict certain antecedents while the pronoun is still free. A similar problem would occur if the pronoun were treated as a pronominal anaphor ([+anaphoric] and [+pronominal] in early Binding Theory terms), since this would make the pronoun subject to both Binding Conditions A and B, requiring it to be both bound and free. This might, in theory, be possible given a theory in which anaphors and pronominals are subject to different binding domains (see section 3.3.2 for environments which allow either pronominals or anaphors, and section 4.3.2 for further discussion of pronominal anaphors), but Huang (2000) points out multiple languages where differences in binding domains for pronominals and anaphors does not obtain.

One of the more widely accepted analyses of binding domains for long distance (and local) anaphors is one in which there is covert Logical Form (LF)[36] movement of the anaphor, such that each dependency of an anaphor on its antecedent is a local dependency. This analysis was inspired by work by Lebeaux (1983) and Chomsky (1986), and has been applied to several other languages. I focus here on Huang and Tang's (1991) analysis of Mandarin Chinese.

[36] The Conceptual-Intentional (CI) interface in some theories.

Motivation for this analysis comes from earlier work by Barss (1986), who analyzed such dependencies as they relate to the 'reconstruction problem'. Consider the sentences in (30).

(30) a. *John knows that Bill likes pictures of himself.*
b. *John knows that, pictures of himself, Bill likes.*
c. *Pictures of himself, John knows that Bill likes.*

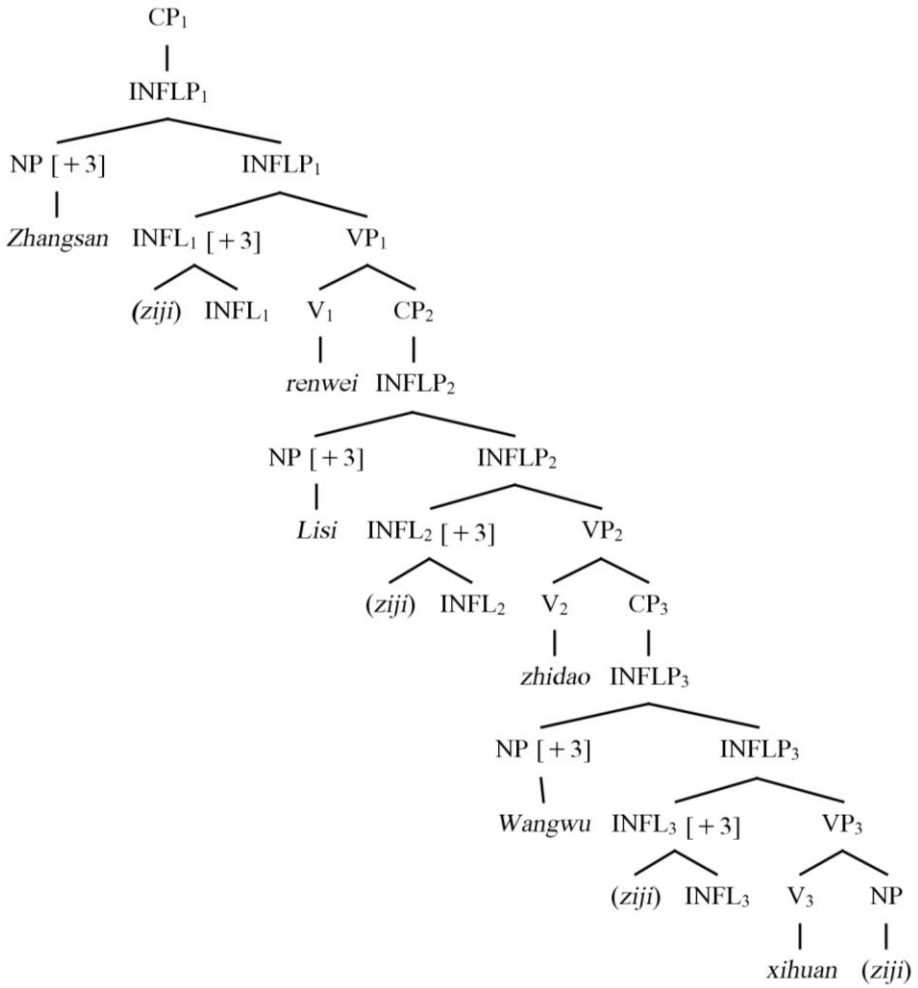
(Huang & Tang 1991: 272)

In (30a), *Bill* is the only possible antecedent of the anaphor, in a typical case of Condition A binding. In (30b) and (30c) however, either *John* or *Bill* is a possible antecedent for the anaphor. Recall from section 3.1 that binding is characterized by the binding element c-commanding and being co-indexed with the bound element. In (30b), only *John*, and not *Bill*, c-commands the anaphor, and in (30c), neither *John* nor *Bill* c-command the anaphor. Barss (1986) however, makes a claim that *himself* is actually 'chain bound' by *John* or *Bill* in these sentences, by virtue of binding the trace of the anaphor.

The LF movement analysis relies on head-to-head movement of the anaphor between INFL projections. Huang and Tang (1991) note that for the long distance anaphor *ziji*, while the antecedent may ultimately fall outside the traditional binding domain for anaphors, such binding requires the anaphor to be licensed by a local NP that agrees in person and number to the actual antecedent. In Mandarin, this could be any subject that agrees with the anaphor in person and number, and does not have to be the antecedent. From there, other successive NPs may license each other as long as they maintain agreement in person and number, thus accounting for the Blocking Effect, noted in section 3.1.2. This is illustrated in (32), a syntactic representation of (13a), repeated below. The structure in (32) shows each possible position of the covertly raised anaphor *ziji*.

(31) a. *Zhangsan*$_i$ *renwei* *Lisi*$_j$ *zhidao* *Wangwu*$_k$ *xihuan* *ziji*$_{i/j/k}$
Zhangsan think Lisi know Wangwu like self
'Zhangsan thinks Lisi knows Wangwu likes self.'

(32)




Assuming a reading of (31) in which *ziji* is bound by *Zhangsan* in the matrix clause, *ziji* covertly raises to each intervening INFL projection. In each of those INFL projections, although the NP in the Spec of that INFL is not the intended antecedent, the fact that the NP in that Spec agrees in person and number with the anaphor allows *ziji* to continue raising. This would not be the case in the instance of the Blocking Effect, illustrated in (15) in section 3.1.2.

Crucially, Huang & Tang (1991) note that this INFL-to-INFL movement must happen covertly at LF. Previous work, e.g. by Chomsky (1981) and Barss (1986) has shown that binding theory must also apply in the narrow syntax, but Huang & Tang show that LF application of binding is also necessary for their theory. Application of binding at LF is established as well in other subsequent work on binding and the copy theory of movement. This allows for the ambiguity in (33a), given the underlying structure in (33b). Since there are copies of *which picture of himself* in positions

that can be bound by both *John* and *Fred*, the sentence can be interpreted in either way, similarly to each copy of *ziji* shown in (32).

(33) a. *John$_i$ wondered which picture of himself$_k$ Fred$_{i/k}$ liked.*
b. [$_{TP}$ John wondered [$_{CP}$ [which picture of himself ] [$_{TP}$ Fred liked [which picture of himself ]]]]

(Hornstein et al. 2005: 257)

Huang & Tang's (1991) analysis separates binding in terms of phi-features, which applies in the syntax, and binding in terms of reference, which applies at LF. Consider the Mandarin example in (34).

(34) *Zhangsan$_i$ shuo Lisi$_k$ chang piping ziji$_{i/k}$*
Zhangsan say Lisi often criticize self
'Zhangsan said that Lisi often criticized self.' MANDARIN (Huang & Tang 1991:275)

The binding that applies in the syntax allows for *ziji* to receive phi-features, namely 3rd person singular, masculine in (34). However, binding must apply again at LF, since there are still two possible antecedents in *Zhangsan* and *Lisi*. At LF, the lower copy of *ziji* can be interpreted, receiving its reference and being bound by *Lisi*, or the higher copy of *ziji* can be interpreted, having raised to the higher INFL projection, which allows *Zhangsan* to bind *ziji* locally at LF in one interpretation.

For the purposes of the canonical Binding Conditions in this dissertation, I will be assuming the definition of a binding domain as given in (26) above, though the formalization of an "accessible" subject will not be important to my analysis. Further, I will be assuming that, for instances of long distance anaphora, the anaphor covertly raises, and the copy theory of movement allows for the raised copy of the anaphor to be interpreted, similarly to the representation in (32). Chapter 4 will highlight some differences in the way that this process happens in Bora, since Bora is not subject to the same Blocking Effect as Mandarin, but some aspects of a theory of Blocking Effect still hold.

### 3.3.2. Binding Domains for Non-Anaphors

Pronominals have been shown to have different requirements for binding domains than anaphors. Recall from (19) above that a pronominal must be free in its binding domain. Notably, Bresnan

(1987) was among several authors[37] to note that the binding domain for pronominals does not need to include a subject in the way that anaphors do.

(35) *John$_i$ put the wine in front of him$_{i/k}$.*

In fact, this makes a prediction that there should be domains which license both an anaphor and a pronominal. In PP complements, the object of the PP is in an XP which contains itself and its case assigner, however it cannot contain a subject[38]. This means that the PP domain serves as the binding domain for pronominals, but cannot serve as the binding domain for anaphors.

(36) *[ John$_i$ put the wine [ in front of him$_i$/himself$_i$ ] ].*

Both the anaphor and the pronominal are grammatical in (36) with the same referent, *John*. This is because the PP serves as the binding domain for the pronominal, as the smallest XP which contains the pronominal and its case assigner. Because it is co-indexed with *John*, and *John* is outside of the PP, *him* is free within its binding domain, in accordance with Condition B. The anaphor *himself*, on the other hand, must be bound in the smallest XP which contains itself, its case assigner, and a subject. Since the only available subject is *John*, the binding domain for *himself* must be the whole sentence. This allows *himself* to be bound by *John*, in accordance with Condition A.

Notably, this differs from the ECM constructions discussed in section 3.3.1, repeated below.

(37) *The prosecutor$_i$ believes [the suspect$_k$ to hate him$_{i/*k/m}$/himself$_{*i/k}$ ].*

In the case of ECM, the embedded clause object is bound within a binding domain that fits the definition given in (26) above, bracketed in (37), despite the fact that the embedded clause does not contain a CP. As such, the anaphor and the pronominal in the embedded clause cannot take the same antecedent as its referent. The anaphor must take a referent in its local domain, where *the suspect* was generated as an accessible subject, while the pronominal must be free in that same

[37] See, for example, Hestvik (1991) and references therein.
[38] The PP also cannot contain anything else that Büring (2005) notes can constitute a binding domain: a coargument, a finite clause, or a whole sentence.

domain. As far as ECM constructions are concerned, anaphors and pronominals share the same binding domain, unlike the PP construction shown in (36) above.

### 3.4. An Alternative Binding Theory

Reinhart and Reuland (1993) (RR) note the same overlap in the distribution of anaphors and pronominals shown in (35) and (36) above. However, rather than explaining this overlap by redefining what serves as a binding domain for anaphors and pronominals, they redefine the conditions that govern how anaphors and pronominals are distributed (notably doing away with the notion of *binding* entirely). While canonical Binding Theory consists of structural conditions that license the appearance of pronouns and R-expressions, RR's Conditions are non-structural in nature, though they must also make reference to a Chain Condition, which is structural in nature. Primarily, RR take reflexivity to be a central property for the distribution of anaphors and pronominals. While RR's proposed Conditions still leave some exceptions, they account for those exceptions by appealing to a Chain Condition.

RR also use their analysis to account for apparent violations of Condition A in canonical binding theory, in which anaphors in English (and Dutch) can appear in positions with no sentential antecedent (i.e. no binder) (38-39). They also argue that their analysis does not require different analyses for anaphors and logophors (40).

(38) a. *Physicists like yourself are a godsend.* (citing Ross 1970)
b. **A famous physicist has just looked for yourself.*

(39) a. *She gave both Brenda and myself a dirty look.* (citing Zribi-Hertz 1989)
b. **She gave myself a dirty look.*

(40) a. *It angered him that she… tried to attract a man like himself.* (citing Zribi-Hertz 1989)
b. **It angered him that she tried to attract himself.* (Reinhart and Reuland 1993: 669-670)

In RR's account, a predicate can be marked as reflexive if two of its arguments are coindexed. Given this, the predicate can be reflexive in one of two ways: the predicate can be intrinsically marked as reflexive in the lexicon, or the predicate can be marked as reflexive by having one argument that is a complex anaphor. These two criteria form the basis for two properties that are central to RR's analysis: being *reflexive*, and being *reflexive-marked*.

(41) a. A predicate is *reflexive* iff two of its arguments are coindexed.

b. A predicate formed of (a head) P is *reflexive-marked* iff either (i) P is lexically reflexive[39], or (ii) one of P's arguments is a complex anaphor.

(Reuland 2006a: 279-280, citing Reinhart and Reuland 1993)

Using these definitions, RR propose new conditions for the licensing of anaphors. These conditions, unluckily, share their name with the canonical binding conditions. To avoid confusion, I refer to these as RR Condition A and RR Condition B. The Conditions A and B of canonical Binding Theory should not be taken to be related to RR's Conditions in any way.

(42) a. RR Condition A: A reflexive-marked syntactic predicate is reflexive.
b. RR Condition B: A reflexive semantic predicate is reflexive-marked.

(Reinhart & Reuland 1993)

(43) a. A syntactic predicate formed of (a head) P is P, all its syntactic arguments, and an external argument of P (a subject). The syntactic arguments of P are the projections assigned theta-role or Case by P.
b. The semantic predicate formed of (a head) P is P and all its arguments at the relevant semantic level.

(Reuland 2006a: 279-280)

Notably, RR's Conditions must make a distinction between syntactic and semantic predicates. This is because, for cases like (44), the predicate cannot be reflexive-marked, since *herself* is not syntactically, on its own, an argument of the verb. However, the predicate must be reflexive-marked for the anaphor to be licensed.[40] A semantic representation of the conjunction allows for an interpretation of *herself* to be interpreted as an argument of the predicate, allowing for the predicate to be reflexive-marked, and to license the appearance of the anaphor through RR Condition B, via the definition in (43b).

(44) a. *The queen$_i$ invited both Max and herself$_i$ to our party.*
b. the queen (λx (x invited Max & x invited x)) (RR 1993: 675)

In contrast, RR Condition A, which refers to reflexive-marked syntactic predicates using the definitions in (41), will rule out a sentence like (45), since the predicate is reflexive-marked (see (41b)) by nature of having a complex anaphor as one of the arguments, but the predicate will not be reflexive (see (41a)), since the arguments of the predicate are not coindexed.

[39] I take "lexically reflexive" to refer to a predicate which takes two arguments that are intrinsically coindexed.
[40] This property is not reversible; a reflexive-marked predicate is not necessarily a reflexive semantic predicate, but a reflexive-marked semantic predicate *is* necessarily a reflexive-marked predicate.

(45) **The queen invited myself for tea.* (RR 1993: 675)

It is also consequential that RR take the notion of syntactic arguments of predicates to be not just related to theta-arguments, but also case assignment. Notably, in (46), the anaphor receives a theta-role from the embedded clause, but receives its case from the matrix clause *strike* predicate.

(46) *Max$_i$ strikes himself$_i$ [* t *as clever ]*.
(47) *Lucie$_i$ expects [ herself$_i$ to entertain herself$_i$ ]*. (RR 1993: 679-680)

In (47), the first instance of the anaphor *herself* acts syntactically as part of both predicates, receiving its theta-role from the embedded predicate *entertain*, and its case assignment from the predicate *expect*. Given the definitions in (43) above, this means that *herself* is an argument of both of the predicates, meaning that by the Condition in (42a), *herself* is licensed as reflexive, and by (42b), is therefore reflexive-marked by both predicates. If licensing were purely semantic and not syntactic, *herself* could not be licensed by the matrix clause predicate.

RR must expand their theory in some cases for SE anaphors. Recall from section 3.1.1 that Dutch has a three-way distinction between complex anaphors, SE anaphors, and pronominals. RR's definition of a predicate being *reflexive-marked*, central to their Binding Conditions, is partially dependent on an argument of the predicate being a complex anaphor, given the definition in (41b). Using the definitions given by RR so far, this makes the data in (48) difficult to account for, where an SE anaphor that is not reflexive by RR's definition in (42) is used in Dutch where a reflexive would be used in English.

(48) a. *He$_i$ accidentally assigned himself$_i$/*him$_i$ to himself$_i$.*
b. *Henk$_i$ wees zich$_i$/*hem$_i$ aan zichzelf$_i$ toe*
Henk assigned SE/3SG to himself to
'Henk assigned himself to himself.' DUTCH (RR 1993: 691)

The problem arises in the matrix clause in (48). According to (43a), the SE form *zich* is a syntactic argument of the main clause predicate (rather than the PP predicate) by nature of it receiving both its theta role and Case from the main clause. In order to be licensed as reflexive then, it must either be part of a lexically reflexive predicate (which *wees* is not), or be itself a complex anaphor (which *zich* is not).

RR address the issue brought up in (48) by appealing to a Chain Condition, invoking earlier work by Chomsky (1973) who noted that NP-movement and anaphora were related. Specifically, any environment in which a moved NP can bind its trace is also an environment in which an NP can bind anaphors (either SE or complex), but not non-anaphors.

RR make the case that a chain must be headed by a +R[eferentially independent] element, which excludes anaphors (both SE and complex) and NP-traces, though these elements may serve as intermediate links in a chain. RR take +R to mean that a given NP is fully specified for phi-features, as well as for structural case. RR argue against a requirement for the chain to also form a single theta-argument. This allows the chain to form a syntactic argument rather than a semantic one, since its status is based on the syntactic properties of Case and phi-features rather than semantic interpretation, as was the case in (44b).

The Chain Condition is necessary to account for (48b) above, in which an SE anaphor is used in Dutch where a pronoun cannot be used. The three-place predicate is reflexive based on one of its arguments being a reflexive, which would mean that RR Condition B would allow either a SE or a pronoun. Notably, as shown in (48b), a pronoun is ruled out as an argument of the predicate. To account for this, RR appeal to the Chain Condition. In RR's analysis, pronouns are classified as +R, whereas SE's are classified as –R. As a result, a pronoun in (48b) would be a +R link in a nonhead position in a chain. The SE anaphor *zich*, however, is classified as –R, and is therefore viable as a link in the chain with *Henk* as the head.

Fox (1993) highlights other instances when appealing to the Chain Condition is necessary, given below in (49).

(49) a. **Himself$_i$ likes Bill$_i$.*
b. **Mary$_i$ behaved her$_i$.* (Fox 1993: 2)

In (49a), the predicate is theoretically licensed by RR Condition A, with the predicate reflexive-marked by having an argument that is a complex anaphor. The Chain Condition rules this out because the complex anaphor is the head of a chain, despite being [-R]. Similarly, (49b) is theoretically licensed by RR Condition B, since the predicate is inherently reflexive. Therefore, by RR Condition B, the predicate is reflexive-marked. The Chain Condition rules out (49b) because the chain has two elements that are [+R], which is not permitted in a chain.

Given their proposed theory, RR are able to account for the examples in (38-40) above, repeated below.

(50) a. *Physicists like yourself are a godsend.* (citing Ross 1970)
b. **A famous physicist has just looked for yourself.*
(51) a. *She gave both Brenda and myself a dirty look.* (citing Zribi-Hertz 1989)
b. **She gave myself a dirty look.*
(52) a. *It angered him that she… tried to attract a man like himself.* (citing Zribi-Hertz 1989)
b. **It angered him that she tried to attract himself.* (Reinhart and Reuland 1993: 669-670)

The anaphors in cases like (50a) and (52a) are argued by RR not to be reliant on RR Condition A. Since RR assume, as in the previous section, that the anaphor in these instances raises at LF, the unraised anaphor is no longer subject to the conditions of anaphors. Because binding occurs at LF, and the anaphor has already raised by LF (though its trace remains in the embedded clause in RR's theory), the overtly expressed anaphor no longer reflexive-marks the predicate. Since the predicates in (50a) and (52a) are not reflexive-marked, they are not subject to RR Condition A. The ungrammatical examples in (50b) and (52b) are ruled out because the anaphor is argued not to have raised in these sentences. When the anaphor does not raise, it can no longer satisfy the condition for reflexivity given in (41a), since the coargument of the anaphor in each case is not a valid antecedent. Reflexivity can only be achieved in these examples if the anaphor raises at LF.

The examples in (51a-b) are licensed by RR Condition B, since the anaphor *myself* is interpreted as a semantic argument of the predicate in the same way as is illustrated in (44b) for the anaphor in (44a). The complex anaphor *myself* in (51a) is embedded in the argument of the predicate, and therefore does not reflexive-mark the predicate by virtue of the definition in (41b). The ungrammatical example in (51b) is a result of the predicate being reflexive-marked by means of a complex anaphor. This rules out (51b) by RR Condition A, since the reflexive-marked predicate is not reflexive (i.e. its arguments are not coindexed).

I do not take a position on which of the analyses of anaphors and prononimals (RR's analysis or Chomskyan canonical binding conditions) is more correct. However, in Chapter 4, I consider both types of analyses and what the data from Bora can contribute to either analysis.

### 3.5. Theories of Control

The topic of control clauses has been one of the more intensely debated topics in theoretical syntax in recent years. Cases of linguistic control have been described as the "relation of referential dependency between an unexpressed subject (the controlled element) and an expressed or unexpressed antecedent (the controller)" in which "[t]he referential properties of the controlled element […] are determined by those of the controller" (Bresnan: 1982: 372). This is shown in English in (53).

(53) *John wanted [ __ to leave ]*.

In this example, the embedded clause ___ *to leave* does not have an expressed subject. The subject of *leave*, however, is understood to be *John*, the expressed subject of the higher clause. Because *John* is 'controlling' the unexpressed subject of the embedded clause, that unexpressed subject is understood to have the same referential properties as *John*.

Control clauses may also be further broken down. Consider the cases in (54).

(54) a. *John$_i$ promised us$_j$ [ ___$_{i/*j}$ to come to the show ]*.
b. *John$_i$ persuaded us$_j$ [ ___$_{*i/j}$ to come to the show ]*.

The controlled embedded subjects in (54a-b) show the difference between subject and object control. The verb *promise* in (54a) is a subject control verb, such that the referent of the controlled subject of the embedded clause is determined by the subject of the higher clause. That is, the one who is "coming" in the embedded clause in (54a) is interpreted as *John*, and not *us*. Conversely, the verb *persuade* in (54b) is an object control verb, for which the embedded controlled subject depends on the object of the higher clause for its referent, in that the one "coming" in (54b) is interpreted as *us*, and cannot be interpreted as *John*.

The difference between the control structures and the ECM structures seen earlier is argued to be that the ECM constructions have an element that originates in the embedded clause, but receives accusative case from the matrix clause. In one theory, control structures are argued to have a separate item, PRO, serving as the subject of the embedded clause (to be outlined in section 3.5.1), and the accusative case from the matrix clause is assigned to an object that originates in the matrix clause (though see the movement analysis of control outlined in section 3.5.2 for an alternative

view). This can be seen in that ECM constructions generally have an embedded finite clause counterpart showing that the ECM overt subject indeed originates in the embedded clause. (55) below shows an ECM construction with its finite counterpart which are both grammatical, while (56) shows that a finite counterpart of a control verb in the same way is not possible.

(55) a. *The judge believes [him to be guilty]*.
b. *The judge believes [that he is guilty]*.
(56) a. *John persuaded us [PRO to come to the show]*.
b. **John persuaded [that we come to the show]*.

The example in (54a) above is also an example of Obligatory Control (OC), meaning that the appearance of the control verb in the main clause necessarily indicates a control relationship with the subject of the embedded clause. Other types of predicates similarly have an unexpressed subject that is referentially dependent on a controller, but are not the result of a control verb. Rather, these are the result of a syntactic construction that introduces Non-Obligatory (optional) Control (NOC).

(57) a. *[$PRO_{ARB}$ maintaining innocence] is important in court.*
b. *[$PRO_i$ incriminating $myself_i$] was a mistake.*

In many cases, OC clauses serve as complement clauses, whereas subject and adjunct clauses are cases of NOC. The type of control also ties into (canonical) Binding Theory, in that the subject of OC clauses (assuming a PRO analysis, as described in section 3.5.1 below) must be bound. That is, OC subjects must be c-commanded by an antecedent and must be bound in a local domain. This does not hold for NOC clauses. The controller in NOC clauses does not have to bind PRO, though PRO does have a restriction of being [+human] in NOC clauses. In (57a), PRO is arbitrary and does not have a controller, while in (57b), the controller for PRO is present, but does not bind PRO. For more on NOC clauses and their distinction from OC clauses, see Landau (2013).

### 3.5.1. The PRO/Agreement Theory of Control

In Government and Binding approaches, it has been proposed that the unexpressed embedded subject that is controlled is a linguistic entity called PRO, where PRO is a phonologically null pronoun. In that framework, PRO was necessarily not governed. While Government and Binding theory may have fallen out of widespread use in linguistic analyses, there are still many who have

maintained that PRO is a fundamentally necessary component for theories of control (notably Landau 2015, and other work).

Several syntactic phenomena point to the presence of some kind of null subject in control clauses. Landau (2013) points out several of these phenomena. To point out a few of them here, floating quantifier constructions, partial control constructions, and binding all reveal the presence of a subject in control clauses.

Landau (2013) points out that floating quantifier constructions show agreement with an unexpressed subject that, without the existence of PRO, would be difficult to explain.

(58) a. *They have all gained something.*
b. **Something has all been gained.*
c. *They wanted [PRO to all gain something].*
d. *[PRO to all gain something], they knew, would be a miracle.* (Landau 2013: 73)

The floating quantifier *all* in (58) has an inherent plural agreement with the subject of the clause. In (58b), the suppressed agent of the passive is not capable of that agreement, ruling the sentence out. However, the presence of a PRO subject in the control clauses in (58c) and (58d) does allow for this agreement.[41]

Partial control constructions of the kind illustrated in (59-60) provide additional support for the presence of an unexpressed subject in control clauses. Partial control allows, and sometimes requires that the identity of an embedded PRO properly include its controller. Compare the partial control in (59b) with the exhaustive control counterpart in (59a).

(59) a. **John$_i$ managed [PRO$_{i+}$ to gather at 6].*
b. *The chair$_i$ preferred [PRO$_{i+}$ to gather at 6].* (Landau 2013: 157)

Intransitive uses of verbs like *meet* and *kiss* must be licensed by a semantically plural subject. Semantically singular subjects, like that in (60a) lead to ungrammaticality (cf. (60b)). However, PRO in (60c) is able to provide the semantically plural subject for the embedded clause, being

[41] Note the ungrammaticality of these clauses if the controlling subject is not plural.

i. **He wanted [PRO to all gain something].*

partially controlled by *Mary*. That is, *Mary*, the controller of PRO in (60c), is propertly included in the identity of that PRO. Conversely, in the exhaustive control example in (60d), PRO's identity is determined only by its controller, *Mary*, and not any larger set.

(60) a. **Mary kissed.*
b. *John and Mary kissed.*
c. *John$_i$ felt sorry that Mary$_k$ regretted [PRO$_{i+k}$ kissing the night before]*.
(Landau 2013: 77)
d. *John$_i$ felt sorry that Mary$_k$ managed [PRO$_{k/*i+k}$ to kiss the night before]*.

Landau points out that partial control is an important diagnostic for an Agreement theory of control clauses because PRO is independently specified for feature mereology. This means that PRO will always agree in phi-features with its controller, including in syntactic number, though mereologically they may differ. For example, while *committee* may be syntactically singular, it is mereologically plural. Landau (2003: 493) points out that movement theories of control (see section 3.5.2) cannot account for instances of partial control because "there is no partial raising."

In Chapter 4, I consider some examples of cases of constructions in Bora that have partial control equivalents in English, especially to determine whether the identity of the Bora preverbal proclitic subjects, regardless of whether they represent cases of control, can represent identities of which their controllers or binders are a proper subset.

Perhaps one of the strongest cases that has been presented for the existence of an unexpressed subject is that PRO can bind anaphors.

(61) a. *Mary$_i$ planned [PRO$_i$ to buy herself$_{i/*j}$ a new coat]*.
b. *[PRO$_i$ behaving oneself$_i$ in restaurants] would be necessary.*
(Landau 2013: 75)

In (61a), PRO must be present to act as the binder for the embedded anaphor *herself*. Complex anaphors do not act as long distance anaphors, and so the anaphor must be bound in its local domain. Without PRO in the binding domain, *herself* would remain unbound, violating Condition A. In RR's theory, the embedded clause predicate could not be reflexive-marked without PRO, since there would be no other part of the predicate for *herself* to be co-indexed with. In (61b), *oneself* must be bound in its local domain, also because of Condition A. However, there is no other NP in the sentence, let alone in the local domain, that could serve as a binder for the anaphor. For

this reason, it has been argued that the unexpressed subject PRO is serving as the subject and the binder for the anaphor.

The controller of PRO is always in the immediately dominating clause. In cases where PRO appears as part of an infinitival clause complement to the verb (62a), there is only one choice for a controller. However, cases of subject control (62b) only allow for the subject of the higher clause to act as the controller, while cases of object control (62c) only allow for the object of the higher clause to act as the controller. In the rare case of variable control, (62d), either the subject or the object of the higher clause can act as the controller for PRO.

(62) a. *John tried [PRO to save himself/*oneself]*.
b. *John promised Mary [PRO to save himself/*herself/*oneself]*.
c. *John persuaded Mary [PRO to save herself/*himself/*oneself]*.
d. *John proposed to Mary [PRO to save himself/herself/*oneself]*. (Landau 2013: 124)

Landau's theory proposes that PRO is crucially –R[eferential], meaning that PRO lacks phi-features. Landau refers to this as the anaphoric property of PRO. The [R] feature, Landau [2004] points out, is analogous to [Case] in the MP. That is, whenever the INFL and C projections are specified as +T[ense] and +AGR[eement], they will be specified as [+R]. Any other combination of [–T] [+AGR], [+T] [–AGR], or [–T] [-AGR] is then specified as [-R]. Once the controlled clause enters into an agreement relation with the higher [+R] clause, PRO's [–R] feature is deleted by an agreement relation with its controller, in a way drawing a corollation between binding and control through phi-feature specification. For further details of this analysis, see Landau (2004).

### 3.5.2. The Movement Theory of Control

Other authors, especially those working in the MP who have spent significant time re-evaluating remnants of the government and binding framework, have been skeptical of PRO, notably O'Neill (1997) and Hornstein (1999 and subsequent work). Given a framework like the MP which attempts to reduce analysis of language to only elements and operations that are necessary and simplest (in addition to explaining why that is the case), Hornstein and Polinsky (2010) note that PRO does not make any significant contributions to the grammar. At PF, PRO is phonologically null, and thus makes no phonological contribution. At LF, Hornstein and Polinsky (2010: 5) note that, in cases

of OC,[42] PRO's "only semantic contribution is to provide… a semantic place holder whose value is provided by its antecedent".

Among the problems with PRO that led to the movement analysis was the postulation of a special, phonologically null grammatical case for PRO, which PRO must receive, and which no other linguistic element can receive. Under this analysis, Chomsky and Lasnik (1993), and others, especially Martin (2001), argued that the phonological nullness of PRO could be accounted for.[43] Martin captured the difference between OC clauses and other similar infinitivals (raising and ECM) by claiming that OC clauses actually involved tensed infinitives, the subject of which required null case. This has been argued against, e.g. in Pires (2001, 2006).

Proponents of the movement theory of control have argued that there is no need for the entity PRO, but rather the empty subject position is the result of A-movement of an element, leaving behind an unpronounced trace. The result is an A-chain headed by the moved element, which is then able to control its trace. Hornstein and Polinsky (2010: 10) note the properties that A-chains have in common with OC: all but the head of the chain is phonetically null, only the head of the chain is in a case-marked position, the movement can be successive cyclic, the chain respects locality, they license sloppy readings under ellipsis, and they do not block *wanna* contraction.

In addition to the similarities between control clauses and A-chains, the movement theory also draws on the similarities between control and raising clauses. In a PRO theory analysis, the difference between the raising-to-subject sentence in (63a) and the OC structure in (63b) was taken

[42] Boeckx et al. (2010) note that the analyses provided for a movement analysis of control hold for OC (ia), but not necessarily for NOC (ib). Because the subject of an NOC clause does not have to be bound by its controller (sometimes the controller is not even expressed in the sentence), PRO in NOC clauses does not have a sentential antecedent in the same way that OC clause subjects do.

(i) a. *Frank knows that Sue wants [ ___ to enjoy the weather]*.
b. *Frank knows that [ ___ robbing a bank] is a crime.*

Boeckx et al. propose, following Hornstein (1999), that NOC does not involve an instance of PRO, but rather of *pro*, a null pronoun which is used when movement is not possible, continuing to allow for the elimination of PRO in the syntax of control structures.

[43] Other notable approaches to PRO and case include Hornstein (1999) arguing that PRO is caseless under an analysis of A-movement, and Landau (2013) presenting an analysis in which PRO does in fact receive case just like any other NP, giving evidence of PRO appearing in different case positions, and consequently arguing that, in this respect, PRO is no different than any other lexical NP.

to be that the raising structure involved a trace left from the movement, while the OC structure involved PRO in the same position.

(63) a. *The man$_i$ seemed [ t$_i$ to fly ]*.
b. *The man$_i$ tried [ PRO$_i$ to fly ]*.
(64) *The man$_i$ tried [ t$_i$ to fly ]*.

This is the basis of the movement theory of control: ascribing a movement analysis to sentences like (63b), such that a trace would be left (64), similarly to (63a). Notably, an analysis which analyzes both traditional raising-to-subject and OC as movement had problems in a GB approach, namely with the theta-criterion (Chomsky 1981). That is, *the man* receives a theta role from the main clause verb in (63b), but not in (63a). A movement analysis like that in (64) would go contrary to the restrictions of nouns receiving multiple theta roles.

Hornstein & Polinsky (2010) explain that the MP, by the nature of its structure-building operations, handles this problem on its own. With Merge as the sole structure-building operation, there is no longer a need for a distinction between D(eep)-Structure and S(urface)-Structure. The restriction against an expression bearing multiple theta roles followed in Government and Binding theory as a consequence of D-Structure, since movement operations could not apply in the narrow syntax before S-Structure. Even at S-Structure, theta roles are maintained after transformations apply. However, the elimination of D-Structure (Chomsky 1995) in theory opens the door for the possibility of an expression bearing multiple theta roles.

Hornstein's (1999) initial proposal does not seek to eliminate PRO from syntactic theory, but rather establish its properties as an NP-trace, as shown in (64) above. He illustrates this with the following example.

(65) a. *John hopes to leave.*
b. [$_{IP}$ John [$_{VP}$ John [ hopes [$_{IP}$ John to [$_{VP}$ John leave ]]]]].

(65b) shows all of the intermediate landing sites for *John* as it moves. The movement is motivated by feature-checking, both for Case and theta roles. *John* begins by merging with *leave* to check the verb's theta role, then raises to the SPEC of the embedded INFL to check the D feature of INFLP. Following the assumption that a DP can bear any number of theta roles, this allows *John* to continue raising to check the theta role on *hope* in the main clause. Also, crucially for Hornstein's

analysis, *John* in the SPEC, INFLP of the embedded clause is not in a Case position, and must continue to raise to get Case. This motivates *John*'s final movement to SPEC, INFLP of the main clause.

Boeckx et al. (2010) point out and attempt to solve certain apparent shortcomings of Landau's Agree approach. For example, Landau (2004) places great emphasis on partial control as a death knell for a movement analysis of control since, as stated in the previous section, there is no partial raising. Boeckx et al. point out that Landau's analysis also has shortcomings. Landau gives a detailed list of environments that trigger partial control, including matrix clause predicates like *hope* (66a). However, there must also be properties of the embedded clause at play in order to rule out sentences like (66b).

(66) a. *The chair hoped [ to meet at 6 ].*
b. **The chair hoped [ to sing alike ].* (Boeckx et al. 2010: 21-22)

Rather, Boeckx et al. point out that there is evidence that partial control is licensed by the possibility of the predicate to take a commitative PP (67). If the predicate cannot take (and necessarily select for) a commitative PP, then exhaustive control is possible, while a partial control reading is not (68). This, they point out, is still compatible with a movement analysis of control.

(67) *The chair left (with Bill).*
(68) *The chair preferred to leave at 6.* (Boeckx et al. 2010: 22)

A movement analysis of control also allows for covert movement of an expression to the higher clause, allowing for the possibility of backward control, in which the higher clause has a phonologically null subject (represented by Δ below) which still receives a theta role (contrasted with forward control, the type seen thus far in which the overt element in the higher clause controls a phonologically null element in the embedded clause).

(69) $\Delta_{i/*k}$ *[kid-bā*$_i$ *ziya bišra] yoq-si*
girl-ERG cow.ABS feed.INF begin-PST.EVID
'The girl began to feed the cow.' TSEZ (Polinsky & Potsdam 2002: 246)

The analysis presented by Polinsky & Potsdam (2002) for the case of backward control in Tsez is one of movement of the controlled element to the higher clause, but with only the lower copy pronounced. Polinsky & Potsdam (2006) point out that some necessary properties of the complement clause in cases of backward control are: (i) that the complement clause must be capable of licensing an overt subject, and (ii) that the clause be transparent to A movement. However, if the copy is overtly pronounced in each clause, this yields a resumptive pronoun under the copy theory (copy control). Languages may exhibit different types of control; Haddad (2009), for example, shows that the Telugu language exhibits all three types: forward, backward, and copy control.

### 3.5.3. Control of Phonologically Overt Pronouns

The behavior of control clauses is inherently connected to the analysis of co-reference. This is because the languages whose control structures have been analyzed do not have a phonologically present subject in the embedded clause, but in a subset of cases (obligatory control) that subject must be co-referent with an antecedent (its controller).

However, it has been argued in other languages that there can be overt realizations of controlled subjects. For example, Szabolsci (2009) showed that Hungarian allows overt controlled subject pronouns when that pronoun is modified by *only* or *too* (70), and Madigan (2008) showed that Korean (71), Japanese, and Chinese (see (72) below) also allow phonologically overt controlled subjects.

(70) *Szeretnék* *én* *is* *magas* *lenni*
would.like.1SG I too tall be.INF
'I want it to be the case that I too am tall.' HUNGARIAN (Szabolsci 2009: 10)

(71) *Inho$_i$-ka* *Jwuhi$_j$-eykey* *caki$_{j/*i}$-ka* *cip-ey* *ka-la-ko* *mal-ha-yess-la*
Inho-NOM Jwuhi-DAT SELF-NOM home-LOC go-IMP-C tell-do-PST-DC
'Inho told Jwuhi to go home.' KOREAN (Madigan 2008: 237)

Szabolsci (2009) explains that overt nominative subjects in Hungarian controlled clauses are licensed by Long-Distance Agree (70), but she does not detail what rules this possibility out in other languages.

Notably, however, the controlled element must be bound by a local controller (not necessarily the subject of the higher clause). Also, importantly, the use of an overt controlled element indicates

an exhaustive focus interpretation of the controlled subject. Consider the Mandarin example (72) and its explanation in (73) below, which Madigan (2008) indicates shows a pattern that holds in Korean, Japanese, and Serbo-Croatian as well.

(72) *Zhangsan*$_i$ *bi* *Lisi*$_k$ *PRO*$*_{i/k}$/*ziji*$*_{i/k}$ *xie* *zuoye*
Zhangsan force Lisi PRO/SELF write homework
'Zhangsan forced Lisi to do the homework.'

(73) a. With *ziji*: Lisi does the homework by himself. The focus is on Lisi doing the homework alone.
b. Without *ziji*: Lisi is doing the homework (maybe other people are helping, maybe not)
(Madigan 2008: 256-257)

Madigan (2008) does not provide a formal analysis of how the overt realization of the controlled subject comes to appear in the syntax. Rather, he points out that languages like English, following Landau's (2004) analysis, do not allow for overt controlled subjects because of morphosyntactic reasons related to referentiality ([+/-R] of the embedded subject) and the specification of the embedded clause functional projections as being [+/-T] and [+/-AGR]. Madigan shows that this does not differ in Korean (and by extension, the other languages he compares Korean to). Rather, what allows for the overt controlled subject are semantic reasons, though he does not formalize these reasons.

This Chapter has provided formal generative theoretical overviews of theories of binding and control, both of which are potentially relevant to an analysis of preverbal subject proclitics in Bora. For binding, I have given an overview of the canonical Binding Theory and the Conditions assumed therein, with special attention paid to what constitutes a binding domain for anaphors and pronominals. Relatedly, I discussed theories of logophoricity and long distance binding. I also discuss Reinhart & Reuland's (1993) approach to reflexivity as an alternative to the canonical Binding Theory. With respect to control, I have discussed the Agreement theory put forth by Landau, as well as movement approaches to control of the type argued for by Hornstein. For both binding and control, I have provided background on multiple major analyses. In Chapter 4, I apply these analyses to Bora data, attempting to determine whether the Bora proclitics constitute cases of control or anaphora (or potentially both). In doing so, I expect that the data being provided will inform further research into these theories.

# CHAPTER IV

## Analysis of Co-Reference Phenomena in Bora

Here is a summary of what has been discussed up to this point: in Chapter 2, I have laid out the basic syntax of Bora, especially in terms of a series of preverbal subject proclitics in Table 3. Chapter 2 argued for the clitic nature of these subjects, and showed their distribution in both main and embedded clauses, focusing especially on their use to indicate cross-clausal coreference. It was also shown that reflexive and reciprocal morphemes in Bora perform a separate role from these proclitics. Finally, it was established that embedded complement clauses in Bora are finite clauses, and with limited exceptions, all clauses in Bora have an overtly expressed subject.

Chapter 3 laid out the theoretical framework that will be used as a basis for analysis in this Chapter. In particular, I provided detailed summaries of theories of anaphora and control, both of which I consider in this Chapter as possible analyses of Bora preverbal subject proclitics. Regarding anaphora, I outlined what I term canonical binding theory, tracing back to Chomsky (1981), and extending to present day analyses. This involved outlining Binding Conditions A, B, and C (with particular focus on Condition A), as well as providing formalizations for what constitutes a binding domain, and how that may differ for anaphors and pronominals. Additionally, I provided descriptions of languages that display long distance anaphora, and compared this to anaphors that are locally bound. Based on this, I showed an analysis from Huang & Tang (1991) of how long distance anaphora is achieved in Mandarin Chinese, with covert raising of the anaphor at LF. I also provided a summary of Reinhart & Reuland's (1993) alternative approach to binding, which focuses on reflexivity as a relevant property for licensing anaphors.

In addition to theories of anaphora, I also provided summaries of different analyses for control clauses, since in many cases, the Bora proclitics being investigated here manifest themselves in complement clauses, which often involve instances of control cross-linguistically. In doing so, I provided overviews of two different approaches to control: an Agreement approach, as outlined in

Landau (2004) and other work, and a movement approach, as adopted by Hornstein (1999) and other work.

Before presenting further analysis, below is a reminder of the subject proclitics being investigated in Bora and their distribution by person and number.

| | Singular | Dual | Plural |
|---|---|---|---|
| 1st | *o=* | *me=* | |
| 2nd | *u=* | | |
| 3rd | *i=* | | |

Table 5: Embedded clause coreferent subject clitics in Bora (repeated from section 2.5.3)

This chapter analyzes how the proclitics being investigated here fit into a theory of syntax. I explore in particular analyses of control and anaphora outlined in Chapter 3. In this Chapter, in section 4.1, I compare the distribution of the Bora subject proclitics to outwardly similar datasets in San Martín Peras Mixtec and Modern Greek, and point out the relevant similarities and differences. I then go on in section 4.2 to apply the analyses of control outlined in Chapter 3 to the Bora data, attempting to determine how well embedded clauses with preverbal subject proclitics conform to control clauses generally, and whether an Agreement or movement analysis of control could better account for the data. Section 4.3 then analyzes the same proclitics as anaphors, looking at each of the different clitics separately because of their different distributions (outlined in Chapter 2). Section 4.4 continues the analysis of the proclitics as anaphors, considering what properties have to be considered in terms of binding domains in order for the clitics to be propertly bound. Section 4.5 then points out potential problems for the analyses presented and outlines suggestions for further research.

### 4.1. Cross-Linguistic Comparisons

The proclitics being analyzed in Bora seem to share at least some common properties with the expression of subjects in similar clauses in San Martín Peras (SMP) Mixtec when expressing coreference across clauses. Compare the Bora data in (1) to the SMP data in (2).

(1) ***walle*** *ábájɨɨve-h* *[* ***i****=nuú-ne* *wájyamu ]*
woman forget-PRED 3COR=sew-CL:IN clothing
'The woman forgot to sew the clothes.' BORA

(2) *Nàntŏso* ***nà*** *kan* *[ nakatsya* ***nà*** *míí* *tsyàà ]*
forget.PST she that wash.IRREAL she the clothes
'That woman forgot to wash the clothes!'

SAN MARTÍN PERAS MIXTEC (Ostrove 2018: 32)

Similarly to Bora (1), the data in (2) show that the subject of the matrix clause in SMP Mixtec is also overtly expressed in the embedded clause. The situation in SMP Mixtec is more complicated, however, as Ostrove (2017) shows that there are different types of embedded complement clauses (recall from section 2.6 that Bora only has finite complement clauses). SMP Mixtec allows for fully inflected embedded clauses (that is, these embedded clauses allow for the full range of TAM morphology) (3a), and two distinct types of subjunctive clauses, which Ostrove terms F-subjunctives (3b) and C-subjunctives (3c). Notably, the embedded clause only requires an overt realization of the subject in F-subjunctives, but not in C-subjunctives. It is noted in the examples below that inflection for tense is not allowed in either of the subjunctives (recall from section 2.6 that Bora can express tense in complement clauses).

(3) a. *Káchi* *ñá Maria* *ba'a* *tsyáa* *ñá*
say.PST Maria well write.PRES she
'Maria said that she writes well.'

b. *Kòni* *rà Julio* *kusi/*kìxi* *rà*
want.PST Julio sleep.IRREAL/sleep.PST he
'Julio wanted to sleep.'

c. *Kìxă* *míí* *leso* *taxa'a/*tàxa'a*
start.PST the rabbit dance.IRREAL/dance.PST
'The rabbit started to dance.' SAN MARTÍN PERAS MIXTEC (Ostrove 2017)

Ostrove (2017) characterizes the difference between the two types of subjunctive in terms of properties related to control: the subjects of C-subjunctives must be locally bound, *de re* readings are not available, and strict readings under VP-ellipsis are not available. Meanwhile, the subjects of F-subjunctives need not be locally bound, and both *de re* readings and strict readings under VP-ellipsis are available. Using this as evidence, Ostrove concludes that C-subjunctives are instances of OC, while the subjects of F-subjunctives behave as pronominals. Interestingly, the verbs that Ostrove (2017) lists that introduce C-subjunctive complements in SMP Mixtec are verbs which necessarily have coreferent embedded clause subjects (*start, finish, forget*), whereas F-subjunctive complements are introduced by verbs which either can or must have a disjointly referent embedded clause subject (*want, pray, order*).

Besides differing from Bora in distinctions between finite and subjunctive embedded clauses, SMP Mixtec differs from Bora in that the expression of the subject in SMP Mixtec is morphophonologically identical in F-subjunctives regardless of whether it appears in a main or embedded clause, and regardless of whether or not it is coreferent with another NP. That is, if *ñá* appears in the main clause, *ñá* will also be the form that appears in the embedded clause. While this is the case for the 1st and 2nd person clitics in Bora (*o=, u=, me=*), it is not the case for the Bora 3rd person coreferent marker (*i=*).

In instances of pronoun doubling, whether within a single clause or across clauses, an analysis requires a determination of whether the pronouns in question are in fact pronouns (as Ostrove 2017 argues for subjects of F-subjunctives), or a form of agreement. [44] The remainder of this Chapter will investigate different types of analyses for pronoun doubling in Bora, including the possibility that embedded subject proclitics in Bora are simply agreement markers and not overt representations of nouns.

Baker and Kramer (2018) discuss the importance of this distinction in terms of a difference between agreement and clitic doubling in Amharic. They argue that the clitics they investigate in Amharic are pronouns, with Weak Crossover effects[45] and Condition B violations (4) serving as their diagnostics. In (4a), the clitic *–w* is cannot be coreferent with *Lemma*, the subject of the sentence, indicating that *Lemma* cannot bind *–w*, consistent with Condition B that a pronominal be free within its binding domain (in this case, the sentence). If the object of the clause *is* coreferent with the subject, Condition B cannot be satisfied with respect to the object, since the object would be both coreferent with and c-commanded by the subject. As such, (4c) is grammatical, but (4b), with *–w* indicating the object is not, suggesting again that *–w* is acting as a pronominal.

(4) a. *Lämma gäddäl-ä-w*

[44] Ostrove (2018) determines in his work that there are some instances of 'pronouns' in question in SMP Mixtec that do not meet requirements to be classified as pronouns. Rather he analyzes them as a form of topic agreement in the language. These are separate from the types of examples given in (3) above, although these topics do take the same morphological form as other pronouns in the language.

(i) ***rí*** *ndába* *sâ*
it.AML jump.PRES bird
'The bird is jumping.' (Ostrove 2018: 25)

[45] Because I do not explore Crossover effects as they pertain to the Bora data, due to the lack of relevant empirical data on this phenomenon, I do not go into detail about Baker & Kramer's diagnostic here.

Lemma.M kill.PFV-3M.SG.SUB-3M.SG.OBJ
'Lemma killed him.' (him ≠ Lemma)

b. **Lämma ras-u-n gäddäl-ä-w*
Lemma.M self.M-his-ACC kill.PFV-3M.SG.SUB-3M.SG.OBJ
'Lemma killed himself.'

c. *Lämma ras-u-n gäddäl-ä*
Lemma.M self.M-his-ACC kill.PFV-3M.SG.SUB
'Lemma killed himself.' AMHARIC (Baker & Kramer 2018: 1067-1068)

While the diagnostics used by Baker & Kramer make predictions of how object markers should be analyzed, they do not predict how subject markers should be analyzed, which is necessary for an analysis of Bora. I discuss the Amharic data further in section 4.3.2.1.

The Bora data, on the surface, also seem to share some properties with Greek complementation, as analyzed in Kapetangianni (2010). Kapetangianni points out that Greek, along with other Balkan languages, has for the most part lost the category of non-finiteness, such that embedded clauses are generally finite. Nonetheless, Greek exhibits a distinction between indicative and subjunctive embedded clauses, with indicative clauses introduced by a complementizer *oti* or *pu* (5), and subjunctives introduced by the mood marker *na* (6).

(5) *i maria pistevi oti efige o yanis*
the Mary.NOM believe.3SG.PRES that leave.3SG.PST the John.NOM
'Mary believes that John left.'

(6) *i maria pistevi na efige o yanis*
the Mary.NOM believe.3SG.PRES SUBJ leave.3SG.PST the John.NOM
'Mary wants to believe (hopes) that John left.' GREEK (Kapetangianni 2010: 27)

Kapetangianni notes also that subjunctive clauses with null subjects in Greek are clauses which constitute cases of OC, in which the interpretation of the null subject of the embedded clause is determined by the controller in the matrix clause.

(7) *o yanis kseri na horevi*
the John.NOM know.3SG.PRES SUBJ dance.3SG
'John knows how to dance.' GREEK (Kapetangianni 2010: 27)

There are a few notable differences between Greek and Bora, namely that Greek does not require an overt subject in embedded clauses in the way that Bora does, and Bora does not have a

subjunctive mood that is distinct from indicatives in the way that Greek has. I return to the Greek data in the following section to determine whether any parallels can be drawn to suggest that Bora embedded clauses might be exhibiting some form of control clause.

I have shown in this section that, while other languages seem to display embedded clauses with similar subject properties to Bora, there are key differences that do not allow Bora to be analyzed in the same way as these other languages. I show in the remainder of this Chapter that (i) unlike the data shown for SMP Mixtec, subjects of embedded clauses in Bora are not divisible into two types of clauses, (ii) unlike the data shown for Amharic which is subject to Condition B, Bora embedded clause subjects are subject to Condition A of canonical binding theory, and are thus anaphors, and (iii) unlike the data shown for Greek, there is no distinction between the indicative and the subjunctive in Bora, and that the types of clauses investigated by Kapetangianni (2010) in Greek all behave in the same way in Bora, despite showing different types of behavior in Greek.

### 4.2. A Possible Control Analysis of Bora Subject Clitics

Because Bora subject clitics occur as the subjects of embedded clauses, one attractive analysis would be to treat them as overt instantiations of subjects of control clauses, similar to those mentioned in section 3.5.3 in Hungarian and Korean, repeated here. I will argue in this section that it is not the case that Bora embedded clauses, especially those with preverbal subject proclitics that are coreferent with the main clause subject, are control clauses. In fact, I will argue that the components necessary for control, based on the analyses presented in Chapter 3, are absent in Bora.

(5) *Szeretnék* *én* *is* *magas* *lenni*
would.like.1SG I too tall be.INF
'I want it to be the case that I too am tall.' HUNGARIAN (Szabolsci 2009: 10)

(6) *Inho$_i$-ka* *Jwuhi$_j$-eykey* *caki$_{j/*i}$-ka* *cip-ey* *ka-la-ko* *mal-ha-yess-la*
Inho-NOM Jwuhi-DAT SELF-NOM home-LOC go-IMP-C tell-do-PST-DC
'Inho told Jwuhi to go home.' KOREAN (Madigan 2008: 237)

First, it should be noted that Bora appears to lack the types of predicates that have traditionally been analyzed as raising-to-subject constructions. This does not allow for a direct comparison between raising and OC constructions in Bora that was part of the basis for most approaches to control clauses. The examples in (7-8) illustrate that, when eliciting sentences that would be

analyzed as containing raising predicates in other languages, Bora speakers still provided finite complement clauses with overt (not raised) subjects.

(7) *aalle* *ɨjtsúcunú* *me=álle-i-ñe*
3.F.SG believe SAP=rain-FUT-CL:IN
'She believes it will rain.' (Provided for 'It seems to her that it will rain.')

(8) *ó=ɨjtsúcunú* *táá-tyáá-lle* *chéme-ne*
1.SG=believe 1.SG.POSS-grandparent-CL:F.SG be.sick-CL:IN
'I believe that my grandmother is sick.'
(Provided for 'It seems that my grandmother is sick.')

Even in cases of what would be considered object control, the subject of the embedded clause appears to be expressed in the embedded clause. The syntactic object of the main clause predicate is overtly expressed in the main clause, evidenced by the accusative case, but the embedded clause will have a preverbal subject proclitic which is coreferent with that object noun phrase. The example in (9a) shows this for the SAP marker, while (9b) shows this for cases of 3rd person.

(9) a. *wajpi* *imílle-tsó* *tsɨɨméne-ke* *i=májcho*
man want-CAUS child-ACC SAP=eat
'The man persuaded the child to eat.'

b. *múha* *mé=imílle-tsó-meí* *me=májcho*
1.PL SAP=want-CAUS-REFL SAP=eat
'We persuaded ourselves to eat.'

Interestingly, the embedded clauses in (9a-b) above exhibit the embedded clause tone pattern with the high tone on the first syllable of the verb (see section 2.1), indicating that there are two separate clauses in these examples. However, the embedded clause verb does not have the classifier that is indicative of complement clauses. Thiesen & Weber (2012) suggest that the classifier on complement clauses allows them to act as a nominal argument of the main clause verb. However, the main clause verbs in the sentences in (9a-b) have derived objects that have been 'demoted' from subjecthood of the underived version of the main clause verb by the addition of the causative morpheme.

There are multiple ways the sentences in (9a-b) could be analyzed. They could represent an instance of object control in Bora, in which the accusative-marked object of the main clause is controlling the subject of the embedded clause. However, other sentences like (10) below show

that object control does not occur in other environments in Bora, as the object of the main clause cannot serve as the antecedent for the embedded clause subject proclitic.

(10) *wajpi*$_i$ *neé* *wállee-ke*$_j$ *i*$_{i/*j}$*=ímillé* *i*$_{i/*j}$*=májcho-ne*
man say woman-ACC 3COR=want 3COR=eat-CL:IN
'The man told the woman that he wants to eat.'

Another alternative would be that the accusative-marked arguments in (9a-b) are an example of ECM, acting as the subject of the embedded clause. This is unlikely since the 3COR marker cannot have an overt noun antecedent within the same clause (this is further detailed in section 4.3.2). Alternatively, the embedded clause subject proclitics could be identifying their antecedent as the subject of the underived main clause predicate. Consider a more literal translation of (9) as 'The man made the child want to eat', in which 'child' serves as the subject of 'want'. This translation shows that the underived subject in (9a) is *child*.

The causative constructions in (9a-b) generally require further investigation as far as formal analysis, before one can determine which of these alternatives would be viable (for a descriptive account of the distribution of valency-changing morphology in Bora, see Seifart 2015b). Additionally, further research could reveal other types of constructions that behave similarly to (9a-b).

In analyzing embedded clauses in terms of whether these proclitics constitute the overt subjects of control clauses, it is necessary to specify the types of clauses that these proclitics can appear in. These include a variety of embedded clauses, which Thiesen & Weber (2012) break down into two categories. The first of these are nominal clauses with predicate-final classifiers, which include complements of main clause verbs like *want* (11), which take a *–ne* classifier[46] attached to the verb, as well as relative clauses, which take a classifier attached to the verb which refers to the syntactic element being relativized. Nominal embedded clauses can also be marked with a case

[46] Thiesen & Weber (2012: 359) treat the *–ne* morpheme as ambiguous between denoting a thing or an event. This creates a distinction (and in some cases, an ambiguity) between the readings in (ia) and (ib) below.

(i) *ó=ájtyumɨ* *dibye* *májcho-ne*
1SG=see 3.SG.M eat-CL:IN
a. 'I saw that he ate.' (event)
b. 'I saw that which he ate.' (thing)

marker, which occurs most often for relative clauses. The second type of embedded clauses are those which Thiesen & Weber classify as adverbial clauses, such as purpose clauses (12) and temporal clauses (13), which are not marked by classifiers, but instead by suffixes indicating their function. Other adverbial clauses include conditional and comparative clauses.

(11) a. *wajpi imíllé diibye ááhɨvé-ne*
man want 3.M.SG visit-CL:IN
'The man wants him to vist.'
b. *wajpi imíllé i=ááhɨvé-ne aadí-ke*
man want 3COR=visit-CL:IN other-ACC
'The man wants to visit him.'

(12) *ávyéjuu-bé=vá=a péé-h bájú pañé-vú iyá-me-ke*
reign-CL:M.SG=RPT=REM go-PRED jungle inside-ALL animal-CL:AN.PL-ACC
*i=néhco-ki*
3COR=hunt-PUR
'A chief went into the jungle to hunt for animals.' (Thiesen & Weber 2012: 482)

(13) *ditye tsá-cooca pee-í-myé i-hjyá-vu*
3.PL come-when go-FUT-CL:AN.PL 3.POSS-house-ALL
'When they come, they will go to their house.' (Thiesen & Weber 2012: 373)

Many control constructions that have been analyzed in the literature have been nonfinite clauses; in fact, recall from (64) in section 3.5.2 that the nonfinite nature of OC clauses is crucial for Hornstein's (1999) analysis, which analyzes the movement of the controlled item to the main clause as Case-driven movement. Such an analysis for Bora would likely involve a case of copy control (as mentioned in section 3.5.2), since the element is pronounced in both the main and embedded clauses. Notably however, there are analyses of control into finite clauses that have been described in the literature. However, these are limited to a subset of verbs that introduce control clauses (see section 5 of Holmberg and Sheehan 2010). Landau (2004) notes that finite control in Hebrew not only constrains finite control to certain verbs, but also only to 3rd person and to complements in the future tense. I focus here on a comparison between Bora and examples of finite control in Greek, also introduced in section 4.1 above (though see also Terzi (1993, 1997) for finite control in other Balkan languages).

Recall from section 2.6 that there does not seem to be a finiteness distinction in Bora. Nominative case is not overtly marked in Bora, and so a Case-driven analysis does not have obvious evidence. However, assuming that all clauses in Bora are finite, each subject should be receiving nominative

case from the INFL projection of its clause, as was shown in section 2.8. In terms of the MP, each overt subject in an embedded clause is unspecified for Case, and must be valued for Case in order to be interpretable at the CI-interface. Complementarily, the INFL projection of the embedded clause must agree in phi-features to value uninterpretable features on the INFL head, which are supplied by the NP subject of the embedded clause.[47] If this symbiotic relationship between the subject and predicate of the embedded clause is indeed the case, then the embedded subject would have no impetus to move beyond the embedded clause.

Comparing now the Bora and Greek data, Kapetangianni (2010) notes that, in Greek subordinate OC clauses, the embedded clause subject is not expressed (14a), though in non-control clauses, the subject can be expressed (14b). Compare this to the corresponding Bora data, with the embedded clause with a coreferent subject in (15a), and a disjointly referent subject in (15b).

(14) a. *i maria kseri na kolimbai*
the Mary know.3SG.PRES SUBJ swim.3SG
'Mary knows how to swim.'

b. *i maria pistevi na efige o yanis*
the Mary believe.3SG.PRES SUBJ leave.3SG.PST the John
'Mary wants to believe (hopes) that John left.' GREEK (Kapetangianni 2010:27)

(15) a. *wajpi imíllé i=ááhɨvé-ne aadí-ke*
man want 3COR=visit-CL:IN other-ACC
'The man wants to visit him.'

b. *wajpi imíllé diibye ááhɨvé-ne*
man want 3.M.SG visit-CL:IN
'The man wants him to vist.'

It appears on the surface that there can be a correlation drawn between Greek null subject OC clauses (14a) and Bora coreferent clauses (15a), indicating that cases of OC are expressed differently in Bora, i.e. with an overt embedded clause subject. However, Greek OC subjunctive clauses are restricted in the verbs that introduce them.

(16) Verbs introducing the Greek indicative
a. Assertive verbs (*say, claim, state*)
b. Epistemic verbs (*believe, think*)
c. Factive verbs (*know, be glad, regret*)
d. Semifactive verbs (*remember, discover*)

[47] This raises the possibility that the embedded preverbal subject proclitics may be agreement markers. I return to this possibility in section 4.3.2.1.

e. Fiction verbs (*imagine, dream*)

(17) Verbs introducing the Greek subjunctive
   a. Volitional/desiderative verbs (*want, desire, hope*)
   b. Modal verbs (*must, may, it is possible*)
   c. Directive verbs (*order, advise, suggest*)
   d. Knowledge verbs (*know how, learn*)
   e. Permissive verbs (*allow, forbid*)
   f. Perception verbs (*see, hear*)
   g. Aspectual verbs (*begin, continue*)
   h. Commissive/implicative verbs (*force, manage*) (Kapetangianni 2010: 25-26)

If Bora coreferent embedded clauses (necessarily involving an embedded preverbal subject proclitic) are indeed analogous to Greek OC subjunctive clauses, we would expect to only see Bora coreferent embedded clauses introduced by the types of verbs indicated in (17), and not those in (16) (except for some cases of overlap). This prediction is not born out, as shown by Bora sentences like those in (18), which are introduced by the types of verbs in (16), and those in (19), introduced by the types of verbs in (17).

(18) a. *Nɨjɨho neé i=íjchívye-í-ñe*
   Nɨjɨho say 3COR=leave-FUT-CL:IN
   'Nɨjɨho says he will leave.'
   b. *wajpí=ñe ɨ́tsaavé-h iijyú=ne i=májcho-ne*
   man=REC remember-PRED yesterday=REC 3COR=eat-CL:IN
   'The man remembered that he ate yesterday.'
   c. *Wajco wajácu i=wájtsɨ-í-ñe péjcore*
   Wajco know 3COR=leave-FUT-CL:IN tomorrow
   'Wajco knows that he will arrive tomorrow.'

(19) a. *walle tujkénú i=tyáá-ne*
   woman begin 3COR=cry-CL:IN
   'The woman began to cry.'
   b. *táá-cáání imílle i=wáhtsɨ-ne*
   1.POSS-father want 3COR=dance-CL:IN
   'My father wants to dance.'

The examples in (18-19) shows that, in cases which would be predicted to show a difference between subjunctive and indicative, Bora does not make a distinction with regard to embedded clauses. For 3rd person coreferent embedded subjects, the same preverbal subject proclitic is used in both cases, the embedded clause verb will always have a high tone on the first syllable of the verb, and the same inanimate classifier is added to the embedded clause verb, as they are all

nominal verbal complements (see the description of the difference between adverbial and nominal embedded clauses earlier in this section).

Recall from Chapter 3 that Hornstein's (1999) movement analysis of control, having dispensed with the Theta Criterion in its traditional sense, involves movement of the controlled element to the matrix clause for Case-related reasons. I argued earlier in this section that this does not account for the Bora facts, since the finite nature of Bora embedded clauses allows for the subjects of embedded clauses to receive their Case without movement. In order to apply an analysis of movement-driven control, Bora would need to exhibit embedded clause subjects that cannot receive Case within their own clause. Though (9-10) above could potentially be analyzed in this way, the evidence broadly does not lend itself to such an analysis. This suggests that, given a movement theory of control, Bora does not have the components for control present in the language.

Consider now a PRO approach to control. Many approaches to PRO in OC clauses require PRO to be case-marked, whether with null case as proposed by Chomsky and Lasnik (1993), or with standard case, the same as other DPs (Landau 2006). Alternatively, an approach to control involving PRO does not require Case-motivated movement in the way that the movement analysis does.

Landau's (2004) agreement analysis of OC clauses requires the controlled clause to be negatively specified for AGR or T(ense), in addition to being –R (lacking in phi features). This would mean that an analysis of PRO in Landau's terms would require the OC clause to be either –AGR or –T in order to license PRO. Embedded clauses in Bora, however, do not seem to fit these criteria, since they are not tenseless, as indicated by the future tense in (20) and the past tense in (21).

(20) *wajpi tsá ílli-tyú-né ɨ=dsɨjɨvé-i-yó-ne*
man NEG fear-NEG-CL:IN 3COR=die-FUT-FRS-CL:IN
'The man is not afraid to die.'

(21) *ó=illíjkivyé oó=ne o=neé-ne*
1SG=regret 1SG=REC 1SG=say-CL:IN
'I regret saying (that).'

In order to have an analysis of Bora that would constitute control under the agreement analysis, Bora would need to exhibit clauses that are not specified for AGR or T. Because agreement is not manifested overtly in Bora (see section 4.3.2.3 for further evidence of this claim), such an analysis

would include clauses which would be unable to be specified for tense. Having no evidence that such clauses exist in Bora[48], the components for control seem to be lacking in Bora.

Finally, one of the hallmarks of a PRO analysis is that PRO is phonologically null. Recall from Chapter 3, however, that Madigan (2008) and Szabolsci (2009) (among others) have proposed phonologically overt PRO in languages like Korean and Hungarian, repeated earlier in this section, and again here.

(22) *Szeretnék* *én* *is* *magas* *lenni*
would.like.1SG I too tall be.INF
'I want it to be the case that I too am tall.' HUNGARIAN (Szabolsci 2009: 10)

(23) *Inho$_i$-ka* *Jwuhi$_j$-eykey* *caki$_{j/*i}$-ka* *cip-ey* *ka-la-ko* *mal-ha-yess-la*
Inho-NOM Jwuhi-DAT SELF-NOM home-LOC go-IMP-C tell-do-PST-DC
'Inho told Jwuhi to go home.' KOREAN (Madigan 2008: 237)

The argument put forth by Madigan (2008) is that phonologically overt subjects in these languages involve an exhaustive focus interpretation of the controlled subject. This is not the case in Bora. In Bora, there is no other way to express an embedded clause subject without using the clitics in Table 5 (with the exception of 3rd person disjoint referent subjects, which cannot use any of those clitics). This means that, according to a theory like Madigan's, if we were to analyze Bora embedded complement clauses as cases of OC, then each instance of an overt embedded subject should be involve an exhaustive focus interpretation of that subject. Since I have no evidence of embedded clauses without overt subjects besides those outlined in section 2.9, an analysis of Bora embedded clauses as cases of OC with overt subjects is an unattractive analysis.

In sum, the combination of these factors suggests that the clitics in Table 5 do not form part of control clauses in Bora; in fact, it seems that Bora lacks any kind of control structures, based on the fact that Bora embedded clauses do not seem to fit into any traditional analysis of control: a movement analysis of control is ruled out by the lack of nonfinite clauses conforming to Case-related movement, a PRO agreement analysis is ruled out since Bora embedded clauses do not fit Landau's (2004) –AGR and –T criteria, an analysis akin to Kapetangianni's (2010) analysis of Greek OC subjunctive clauses is not feasible since the types of clauses that Greek subjunctives are restricted to do not match the Bora data, and an analysis of Bora embedded clause subjects as overt

[48] Imperatives are a possible exception.

representations of PRO does not match Madigan's (2008) analysis since the embedded clause subject is not interpreted as exhaustive focus of the controlled subject.[49] For these reasons, I do not analyze embedded complement clauses in Bora as cases of OC.

In this section, I have shown that embedded clauses in Bora with preverbal subject proclitics do not conform to analyses of control. I have applied different theories of control, which are outlined in Chapter 3, to the Bora data, and found that Bora lacks the necessary components for control to be present in the language. I now turn to a possible anaphora analysis of the Bora data.

### 4.3. A Possible Anaphoric Analysis of Bora Subject Clitics

Assuming that the clitics from Table 5 are not subjects of control clauses, I explore the possibility that these Bora subject clitics have properties of anaphors. I have provided evidence in Chapter 2 suggesting that these clitics can co-refer with their antecedent in a higher clause (see also (24) below).[50] It is also the case that the antecedent of these clitics must always c-command the clitic. Recall that from Chapter 3 that canonical binding theory requires co-indexation[51] and c-command as requirements for an anaphor to be bound. With this in mind, I explore whether these clitics behave in other ways that are similar to other anaphors cross-linguistically. Because of partial differences in the distribution requirements for each of the clitics from Table 5, I analyze each in turn.

#### 4.3.1. 1st and 2nd Person Singular

Recall from Chapter 2 that 1st and 2nd person singular pronouns in Bora have the same form in both main and embedded clauses. Although the same form is used across clauses, the tone on the embedded clause proclitic appears different because of the high tone on the embedded clause verb's first syllable (see section 2.1 for more on the LLX constraint that restricts this).

(24) a. *ó=imíllé* *u-ke* *o=ííte-ne*
1SG=want 2SG-ACC 1SG=see-CL:IN
'I want to see you.'

[49] Following Thiesen and Weber's (2012) analysis of a nonfinite tone in Bora, there may be nonfiniteness, but not at the clausal level. That is, the cases which they analyze as having a nonfinite tone are never clausal, but instead operate at the word level.
[50] This is not necessarily the case with the 1st and 2nd person non-singular SAP marker. However, this Chapter will primarily focus on instances when the SAP marker is coreferent with the main clause subject.
[51] As noted in Chapter 3, I use the term "co-indexation" here for expository purposes only, not to assert the existence of indices as part of syntactic knowledge.

b. *ó=llíanú* *o=májcho-ki*
1SG=hunt 1SG=eat-PUR
'I hunt in order to eat.'

c. *ú=piívyete* *u=íjchi-ñe*
2SG=be.able 2SG=swim-CL:IN
'You can swim.'

(25) *múu-be-ámi* *o-ke* *náni-híjcya* *tá-mutsɨ́tsɨ*
who-CL:M.SG-DISGUST 1SG-ACC steal-HAB 1.SG.POSS-caimito
'Who has been robbing me of my caimitos?!'

The 1st person singular subject clitic will always appear as *oó*, and the 2nd person singular as *uú*, with some variations in length and tone that are unrelated to the lexical item itself. The stand-alone form of the 1st person singular is *oó*, and the 2nd person as *uú*, but these are shortened to *o=* and *u=* respectively when they are cliticized to a verb, or when they take a case marking. The form of the pronoun is also the same regardless of whether it is a preverbal subject ((24a-b, 27) for 1st person singular, (24c, 27a) for 2nd person singular) or takes an overt case marker ((25) for 1st person singular, (24a, 27a) for 2nd person singular). However, Thiesen and Weber (2012) note that in single word responses to questions, when the pronoun is overtly case-marked (recall that nominative is unmarked in Bora, and inananimate nouns cannot receive accusative case marking), the vowel remains long, as in (26, 27a).

(26) a. *oó-ke*
1SG-ACC
'to me'

b. *meé-ke*
SAP-ACC
'to us (incl.)' Thiesen and Weber (2012: 234)

When a second-position clitic appears in the clause (like those described in section 2.8), the pronoun that the clitic attaches to is not shortened, but the pronoun is also repeated as a pro-clitic attached to the verb. (27b) shows the 1st person pronoun attached to the second position recent past clitic, after which the 1st person preverbal subject proclitic is repeated on the verb *wajacu*. When the pronoun is shortened before a verb, it has high tone. However, in embedded clauses, there is a marked low tone on the embedded clause subject.[52]

(27) a. *u=ímíllé-hajchíí* *ó=péé-i-yá* *úú-ma*
2SG=want-COND 1SG=go-FUT-FRS 2SG-SOC

[52] This marked low tone is likely the cause of the high tone on the first syllable of embedded clause verbs, which would have arisen to avoid violating the LLX constraint (see section 2.1).

'If you want, I will go with you.'

b. *oó=ne* *me=álle-ne* *o=wájácu-ca* *tsa* *o=péé-í-tyu-ro*
1SG=REC SAP=rain-CL:IN 1SG=know-CF NEG 1SG=go-FUT-NEG-FRS
'If I had known that it would rain, I would not have gone.'

Due to the 1st and 2nd person pronouns appearing as the same form in all positions, with the only differences being the tone and length differences described above, there does not seem to be evidence that these clitics have the properties of an anaphor. The referent of the clitics is not dependent on c-command or coindexation, as would be the case with an anaphor, nor is the referent dependent on the clitic appearing in some sort of local domain. Additionally, there is no difference between the form that appears when an embedded 1st or 2nd person singular subject is coreferent with (24), or disjointly referent (27a) with the main clause subject (compare also (28a-b), which is not the case for the other clitics from Table 5.

(28) a. *ú=imíllé* *amóó-me-ke* *u=dóó-ne*
2SG=want fish-CL:AN.PL-ACC 2SG=eat.meat-CL:IN
'You want to eat fish.'

b. *ó=imíllé* *o-ke* *u=ííte-ne*
1SG=want 1SG-ACC 2SG=see-CL:IN
'I want you to see me.'

That is to say, the 1st person singular pronoun in Bora will always be *oó*, and the 2nd person singular pronoun will always be *uú*, with differences in length and tone as described above. Since the referent of these pronouns is not based on c-command, coreference, or a controlling NP, the Bora 1st and 2nd person pronouns do not fit any criteria for binding or control as described in Chapter 3. Rather, the referent for these clitics is purely discourse-oriented, dependent entirely on the identity of the speaker and the adressee. This indicates that the 1st and 2nd person pronouns act as pronominals, and are thus subject to Condition B.[53]

[53] Since Condition B states that a pronominal must be free within its binding domain, the only instance where we would expect these pronouns to be disallowed would be in object positions where they are coreferent with the subject of the same predicate. Recall, however, that this instance is expressed with the reflexive marker described in section 2.4.

(i) *ó=imíllé* *o=ɨíté-meí-ñe*
1.SG=want 1.SG=see-REFL-CL:IN
'I want to see myself.'

### 4.3.2. 3rd Person

The 3rd person coreferent clitic behaves still differently than the other clitics analyzed so far. Consider the following sentence from Bora with a contrasting English sentence.

(29) *Maria illí i=dsɨ́jɨ́ve-i-jo-ne*
Maria fear 3COR=die-FUT-FRS-CL:IN
'Maria$_i$ is afraid that she$_{i/*j}$ will die.'
(30) *Maria$_i$ is afraid that she$_{i/j}$ will die.*[54]

In (30), *she* can refer to either *Maria*, or someone else who is female, but is not *Maria*. This is not the case in Bora. In order for the embedded clause subject to be co-referent with the matrix clause subject, the 3rd person coreferent marker must be used (29). If the embedded clause subject is not coreferent with the matrix c-commanding subject, one of the personal pronouns from Table 2 (repeated below as Table 6) must be used instead, as in (31).

| | Singular | | Dual | | Plural |
|---|---|---|---|---|---|
| | Masculine | Feminine | Masculine | Feminine | |
| 1st | *(o=)*[55] | | *muhtsi* | *muhpɨ́* | *muúha / (me=)*[56] |
| 2nd | *(u=)* | | *ámuhtsi* | *ámuhpɨ́* | *ámuúha* |
| 3rd | *diíbye* | *diílle* | *diityétsi* | *diityépɨ́* | *diitye* |

Table 6: Bora Animate Personal Pronouns

(31) *wajpí$_i$ tsɨ́ɨ́méné$_j$ imíllé dibye$_{i/*j/k}$ májcho-ne*
man child.GEN want 3.M.SG eat-CL:IN
'The man's child wants him to eat.'

Note that in (31), there is possible coreference of the pronoun *dibye* with the non-c-commanding noun *wajpi*, or with another extrasentential referent.

Thiesen and Weber (2012: 358) also make the claim that the 3rd person coreferent morpheme is an anaphor, "in the sense that it is generally bound within the domain of the closest accessible

[54] The coreferent reading of this sentence, where *she* is bound by *Maria*, can be stated in English also as "Maria is afraid to die", with an OC clause complement. The Bora equivalent is still that in (29), as Bora does not have the finiteness distinction that differentiates the English examples.
[55] The full form of the 1st and 2nd person pronouns are *oó* and *uú*, but preverbal personal pronoun subjects in Bora undergo vowel shortening.
[56] The choice between the two 1st person plural subject expressions depends on an inclusive/exclusive distinction, described earlier in Chapter to and further in section 4.3.3.

subject." They give the following examples as evidence for this claim, showing that the 3COR marker in (32a) must be bound by the subject of the higher clause, while the pronoun in the embedded clause in (32b) cannot be bound by the higher clause subject.

(32) a. *tee-ne* *újcuu-be* *i=májcho-ki*
that-CL:IN get-CL:M.SG 3COR=eat-PUR
'He got that in order to eat it.'
b. *tee-ne* *újcuu-be* *dibye* *májcho-ki*
that-CL:IN get-CL:IN 3.M.SG eat-PUR
'He$_i$ got that in order for him$_j$ to eat it.' Thiesen & Weber (2012: 238)

Besides the evidence in (32), Thiesen & Weber provide evidence that the 3rd person possessive marker *i-* (see section 2.7) has anaphoric properties. Because I am focusing on the properties of *i=* as a preverbal subject proclitic, I do not consider this as evidence for my analysis, since I do not consider that the possessive marker *i-* and the 3rd person proclitic *i=* necessarily share the same anaphoric properties, as Thiesen & Weber do (see section 4.5 for a potential problem case regarding anaphoric properties of the possessive marker).

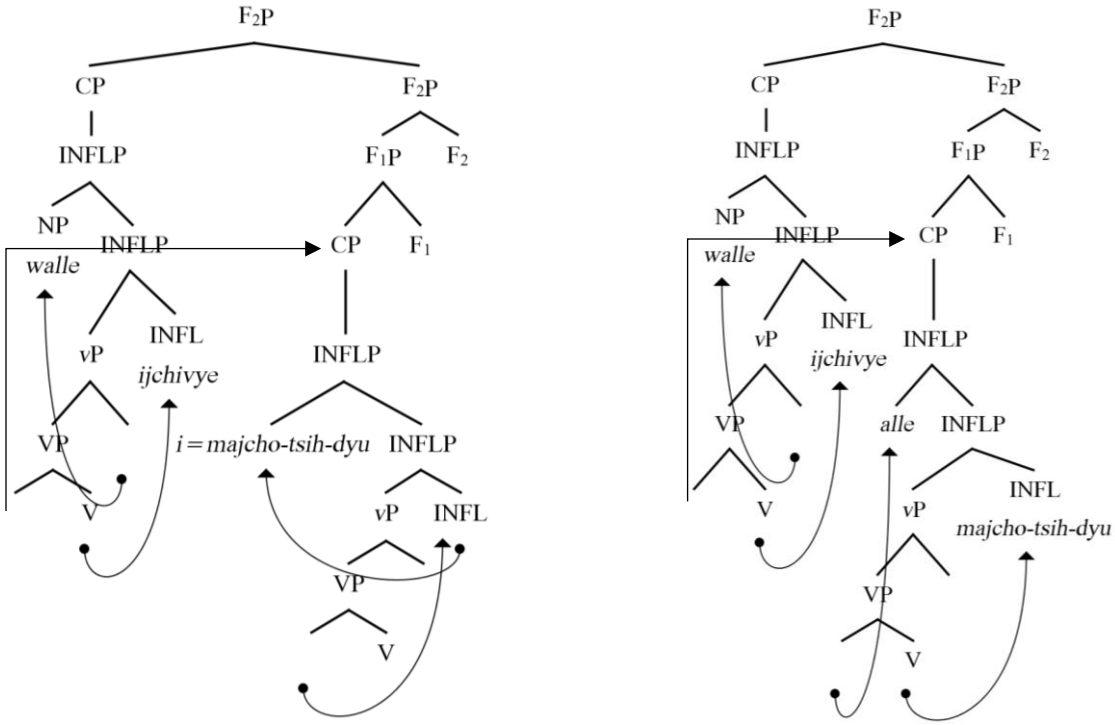
As this is the only evidence that Thiesen & Weber provide for their claim, I now expand on their analysis by considering whether the 3rd person coreferent marker can be analyzed as an anaphor using both canonical binding theory and Reinhart and Reuland's analysis of anaphora.

#### 4.3.2.1. Bora Proclitics as Anaphors under Canonical Binding Theory

If the 3rd person coreferent marker is to be analyzed as an anaphor in canonical binding theory, it must meet the properties of Condition A, namely being c-commanded by and co-indexed with its antecedent within a local binding domain. Consider the following examples in (33) and their proposed representations in (34).

(33) a. *walle* *ijchívye* *i=májcho-tsíh-dyu*
woman leave 3COR=eat-CL:place-ABL
'The woman left after eating.'
b. *walle* *ijchívye* *áálle* *májcho-tsíh-dyu*
woman leave that.one.F eat-CL:place-ABL
'The woman$_i$ left after she$_j$ ate.'

(34) a. b.




Recall from section 2.8 that, in order to achieve an SVO word order like the main clauses in (33), with the complement clause following the main clause verb, it was necessary to postulate two functional projections above the CP, in which the object (here, the complement clause) moved to the Spec, $F_1P$ position, and the remainder of the the original CP then underwent remnant movement to the Spec, $F_2P$ position.

What remains now is determining the binding domain of anaphors in Bora. To establish this, I first consider the Bora reflexive and reciprocal morphemes laid out in section 2.4. It is clear that the clause itself can act as a binding domain (35).

(35) a. *wajpi* *wállee-ma* *ɨ́ɨté-jcatsí*
man woman-SOC see-RECIP
'The man and the woman see each other.'

b. *tsɨ́ímene* *ɨ́ɨté-meí* *mɨ́ɨcúmɨ-ri*
child see-REFL mirror-INS
'The child sees themself in the mirror.'

The reflexive morpheme, as an anaphor, must be bound within its binding domain. In (35a), we see that the reciprocal morpheme is both c-commanded by and coreferent with the coordinated

subject *wajpi wálleema*, fulfilling the requirements for binding. The same is true for the reflexive morpheme in (35b), which is c-commanded by and coreferent with its antecedent, *tsɨ́mene*. All of this takes place within the reflexive and reciprocal morphemes' local CP domains, a domain which is widely considered to serve as a binding domain. The preverbal subject proclitics being investigated here, however, by nature of them being subjects, will not have an antecedent within their local CP that can serve as a binder for them, with the exception of the SAP. For this reason, if we would like to analyze these proclitics as anaphors, it is necessary to look outside of the local CP domain to determine what domain they can or must be bound in.

If the subject proclitic does in fact meet the requirements of c-command and coreference, it must also be determined what the local domain is that the anaphor must be bound in, and whether it shows any properties of long distance anaphora. Recall the properties pointed out in section 3.1.2. that tend to hold for long distance anaphors: (i) they tend to be monomorphemic, (ii) their antecedent tends to be a subject (or at least c-command the anaphor), (iii) their appearance tends to be limited to specific environments, such as infinitive and subjunctive clauses, and (iv) they tend to be subject to the Blocking Effect.

Property (i) is easily satisfied, as *i=* is monomorphemic, unable to be broken down into smaller morphemes. Property (ii) appears to hold also. Consider the case in (36).

(36) *wajpi*$_i$ *neé wállee-ke*$_j$ *i*$_{i/*j}$*=ímillé* *i*$_{i/*j}$*=májcho-ne*
man say woman-ACC 3COR=want 3COR=eat-CL:IN
'The man told the woman that he wants to eat.'

Note in (36) that, although both 'man' and 'woman' c-command the 3COR markers, 'woman' is not a possible antecedent. Only 'man', in the subject position, can serve as the antecedent. This seems to indicate that objects of higher clauses cannot serve as binders for embedded subject preverbal proclitics. Further data needs to be collected to assert this more concretely, as well as to show whether other oblique case-marked nouns in higher clauses can serve as binders in these cases. However, I have collected no evidence that shows anything other than a subject acting as a binder.

Property (iii) is perhaps the tendency that is hardest to show in Bora, since the types of environments that long distance anaphors have been shown to be restricted to in other languages,

namely subjunctives and infinitives, do not manifest in Bora. Rather, the clauses that the 3COR marker appears in are restricted to embedded clauses, which show certain properties that are unique to embedded clauses. Aside from the tonal differences described earlier, embedded clause predicates have a fixed clause-final position (compare the variation in word order that is allowed in main clauses, shown in (61-63) in Chapter 2). This is true of all types of embedded clauses, including relative clauses (37a), adjunct adverbial clauses (37b), complement clauses (37c), and negative clauses (37d), (though remarkably, negative clauses have tonal properties of embedded clauses, even if they are not embedded, as in (38)). Notably, although negative main clauses behave like embedded clauses, the 3COR marker only appears in clauses that are structurally embedded.

(37) a. *tá-ñah-be*$_i$ [*oohííbye-ke* *i*$_i$*=ájtyúmíí-be-dí-vú* ] *o-ke*
1SG.POSS-sibling.-CL:M.SG dog-ACC 3COR=find-CL:IN-AN-ALL 1.SG-ACC
*ajcú-hi*
give-PRED
'My brother gave me the dog he found.' (Thiesen and Weber 2012: 392)

b. [*mítya-ne* *i=wákímeí-ñe-ri* ] *tá-ñah-be* *pávyeenú-hi*
much-CL:IN 3COR=work-CL:IN-OBL.IN 1SG.POSS-sibling.-CL:M.SG tire-PRED
'By working a lot, my brother tired.' (Thiesen and Weber 2012: 377)

c. *diitye* *ɨjtsúcunú* [*péjcore* *i=wájtsɨɨ-ne* ]
3.PL think tomorrow 3COR=arrive-CL:IN
'They think they will arrive tomorrow.'

d. *wajpi* *ɨɨté* *níívúwa-ke* [*i=lííhánú-tuu-be* ]
man see deer-ACC 3COR-kill-NEG-CL:IN
'The man saw the deer without killing it.'

(38) *ípyejcó=ne* *tsáh* *o=cúwa-tu* *ímí*
last.night=REC NEG 1.SG=sleep-NEG well
'I did not sleep well last night.'

It seems then that the restricted environment for the 3COR marker is only limited to embedded clauses.[57] Another possible restriction on the environment for the 3COR marker noted by Thiesen and Weber (2012: 235-236) is what they term discontinuity. There is a tendency in Bora to use overt pronouns or noun phrases when there is a shift from a topical referent to another possible

[57] This may be too tenuous of an argument if one analyzes the 3COR morpheme to be identical to the 3rd person possessive pronoun *i-*. Thiesen and Weber (2012) analyze both of these as what they call the 'self' anaphor. If these are analyzed as the same element, then a restricted environment may not be identifiable for *i-/=*. This is not necessarily problematic for a claim that the 3COR morpheme *i=* is an anaphor however, since the tendencies pointed out for anaphors are not hard requirements. See section 4.5 for problematic cases related to this.

referent. This however might be more related to coreference generally (especially across sentences) in Bora and less related to anaphora. See Seifart (2010) for more on this topic.

The last tendency specific to long distance anaphors is that of a Blocking Effect, which, in Mandarin Chinese, prevents binding of an anaphor by a long distance antecedent if a subject of a different person intervenes. Bora already differs slightly from the languages that have been analyzed for long distance anaphora with Blocking Effects in that those languages are analyzing the anaphoric properties of objects, while the 3COR marker in Bora appears as a subject. Notably, if we consider the 3COR marker to be an anaphor, and we consider the clause level as the binding domain, there would be no way to locally bind the anaphor within the same INFLP. Recall that (36) above showed that the antecedent for the 3COR needs to be a subject of a higher clause (I argue, the structurally next highest subject, described in more detail below), which we have already pointed out is a property that long distance anaphors allow.

Yet another difference between the 3COR marker and other long distance anaphors surfaces in (39). Compare this to the Mandarin data in (40), repeated from section 3.1.2.

(39) a. *Nɨ́jɨho$_i$ ɨjtsúcunu Wajco$_j$ ímillé-ne i$_{*i/j}$=wájtsɨ-ne péjcore*
Nɨ́jɨho believe Wajco want-CL:IN 3COR=arrive-CL:IN tomorrow
'Nɨ́jɨho believes that Wajco wants to arrive tomorrow.'

b. *Nɨ́jɨho$_i$ ɨjtsámei Wajco$_j$ ímillé-ne dibye$_i$ péjcore wájtsɨ-ne*
Nɨ́jɨho think Wajco want-CL:IN 3.M.SG tomorrow arrive-CL:IN
'Nɨ́jɨho$_i$ thinks that Wajco$_j$ wants him$_{i/*j}$ to arrive tomorrow.'

(40) a. *Zhangsan$_i$ renwei Lisi$_j$ zhidao Wangwu$_k$ xihuan ziji$_{i/j/k}$*
Zhangsan think Lisi know Wangwu like self
'Zhangsan thinks Lisi knows Wangwu likes self.'

b. *Zhangsan$_i$ renwei wo$_j$ zhidao Wangwu$_k$ xihuan ziji$_{*i/*j/k}$*
Zhangsan think I know Wangwu like self
'Zhangsan thinks I know Wangwu likes self.' (Cole, Hermon, & Huang 2006: 44)

Notice that, as pointed out in Chapter 3, Mandarin allows for the long distance anaphor *ziji* to be bound by any 3rd person antecedent, so long as there is not another structurally intervening subject that is not 3rd person. This is not the case for the 3COR referent in Bora. While negative evidence of coreference of the type given in (40b) is lacking for Bora because of the elicitation process used (as described in Chapter 1), (39) shows that when the subject of the most embedded clause is coreferent with the structurally next highest subject, the 3COR marker is used, as in (39a). However, if the most embedded subject here is coreferent with a higher subject than the structurally closest

one, an overt pronoun is used as in (39b), rather than the 3COR marker. This does not preclude long distance binding of the anaphor, however, but rather restricts the domain for long distance binding in Bora. I return to the mechanics involved in long distance binding below (see (48) and discussion below).

I argue that this effect falls out naturally from the structure-building nature of the MP. We have established that anaphors are defective with respect to their phi-features. As the syntactic structure is built, the derivation is looking to check and value all of the features within it in order to converge at the interfaces. A derivation that has converged is then transferred as a phase to the interfaces. This means that, in Bora, once a possible subject antecedent for an anaphor is merged into the derivation, the phi-features on that anaphor are valued, which disallows any other possible antecedent that would be merged later to serve as the binder for the anaphor. It must follow from a requirement of the Bora proclitic anaphor that its phi-features should be valued as soon as possible in the derivation, though notably this does not have to be the case for other languages like Mandarin, which allows the anaphor to continue to raise at LF.

While in Mandarin, the fact that the intervening subject in (40a), *Lisi*, is 3rd person allows for *ziji* to take *Zhangsan* as an antecedent, the Bora case in (39b) shows that the embedded subject is expressed as an overt pronoun in order to take *Nɨjɨho* as an antecedent, despite the intervening subject *Wajco* also being 3rd person.

Additionally, if a subject in an embedded clause is bound by a long distance antecedent, the person of an intervening subject is not consequential for the form of the subject used in the embedded clause. In (41a), the intervening subject is 3rd person, while (41b) has an intervening 1st person subject. In both cases, the most embedded subject is expressed by an overt pronoun, because the structurally closest subject is not coreferent with that embedded subject. Compare this to the data in (39) above, in which the 3COR marker can only be used in the most embedded clause if it is bound by the structurally next highest subject, but not when the intended antecedent is any higher than that next highest subject.

(41) a. *Nɨjɨho$_i$ ɨjtsɨcunu Maríhmu$_j$ wajácu-ne dibye$_i$ wátsɨ-ne péjcore*
Nɨjɨho think Marihmu know-CL:IN 3.M.SG arrive-CL:IN tomorrow
'Nɨjɨho thinks that Marihmu knows that he is going to arrive tomorrow.'

b. *Nɨjɨho$_i$ ɨjtsɨcunu ó=wajácu-ne dibye$_i$ wátsɨ-ne péjcore*
Nɨjɨho think 1.SG=know-CL:IN 3.M.SG arrive-CL:IN tomorrow

‘Niɟiho thinks that I know that he is going to arrive tomorrow.’[58]

Therefore, it seems to be the case that the Bora 3COR marker is an anaphor that does not exhibit all of the properties of long distance anaphora, as it does not seem to be restricted to specific environments in the same way that long distance anaphors are in other languages, and does not seem to be subject to the Blocking Effect in the same way that Mandarin is. Rather, the binding domain of the 3COR marker can potentially be analyzed in terms of the definition given in section 3.3.1, repeated below as (42), with the corresponding defitions related to accessible subjects given in (43) and (44). These are then applied to the sentence in (45), the representation of which is given in (46).

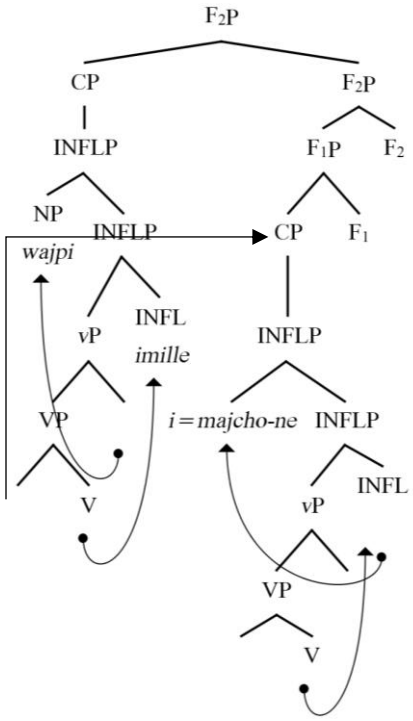
(42) Binding Domain: An anaphor NP must be bound within the smallest XP that contains the NP, the NP’s case assigner, and an accesible subject.
(43) Accessible Subject: α is accessible to β if and only if β is in the c-command domain of α, and assignment to β of the index of α would not violate the i-within-i condition.
(44) i-within-i Condition: [$_{\gamma}$…δ…] where γ and δ bear the same index.

(Reuland 2006a: 265)

(45) *wajpi$_i$ imíllé i$_i$=májcho-ne*
man want 3COR=eat-CL:IN
‘The man wants to eat.’

[58] It is not clear whether *dibye* in (41a-b) can refer to an entity outside the sentence. There seems to be a preference for *dibye* to take an intrasentential antecedent, while *aádi* ‘other (masc.)’ would be used for an extrasentential antecedent. However, this observation may be due to the elicitation process.

(46)




Looking at the representation in (46), the embedded subject *i*= occurs in an embedded CP. Because all clauses in Bora are finite, that embedded CP contains a case assigner for the anaphor, as well as the anaphor itself. This meets all the requirements for a binding domain except for the accessible subject. While, in a technical sense, the embedded CP in (46) contains a subject, that subject is the clitic itself, which does not fit the definition of an accessible subject. According to the definition in (43), the accessible subject α must c-command β, which cannot hold if α and β are the same entity.[59] With the embedded CP ruled out as a binding domain, the search for an accessible subject yields *wajpi*, the only other subject in the sentence.

It seems to be the case then that, under canonical binding theory, *i*= behaves as an anaphor whose binding domain extends to the next highest clause where it can find an accessible subject. I discuss in section 4.4 what makes this possible for Bora, but not for a language like English. This property of Bora explains why the 3COR anaphor is not beholden to a Blocking Effect in the same way as Mandarin Chinese, since whatever subject occurs in the immediately higher clause will serve as

[59] This also rules out "He said that himself would leave" in English, since the embedded anaphor, while a subject, is not an accessible subject to itself, for the same reasons listed above for Bora.

the accessible subject that constitutes the binding domain. Consider the Bora cases repeated below against the traditional Blocking Effect.

(47) a. *Nɨ́jɨ́ho$_i$* *ɨjtsúcunu* $_A$[ *Wajco$_j$* *ímillé-ne* $_B$[ *i=wájtsɨ-ne* *péjcore* ] ]
Nɨ́jɨho believe Wajco want-CL:IN 3COR=arrive-CL:IN tomorrow
'Nɨ́jɨho believes that Wajco wants to arrive tomorrow.'

b. *Nɨ́jɨ́ho$_i$* *ɨjtsámei* $_A$[ *Wajco$_j$* *ímillé-ne* $_B$[ *dibye$_i$* *péjcore* *wájtsɨ-ne* ] ]
Nɨ́jɨho think Wajco want-CL:IN 3.M.SG tomorrow arrive-CL:IN
'Nɨ́jɨho thinks that Wajco wants him to arrive tomorrow.'

In the examples in (47), the subject of the most embedded clause $_B$[ ] has a binding domain $_A$[ ] that extends to the next highest clause, with *Wajco* as the accessible subject. This means that, in (47a), *Nɨ́jɨho* cannot serve as the binder for the subject of the most embedded clause, and in (47b), *Nɨ́jɨho* can bind the most embedded clause's subject, but not within a domain that allows for use of the 3COR marker in the embedded clause. This means that the anaphor *i=* is bound within its binding domain $_A$[ ], in accordance with Binding Condition A, and *dibye* is free within its binding domain $_A$[ ], in accordance with Binding Condition B.

The same logic that predicts the binding conditions in (47) also allows for successive cyclic binding of the type seen in (48) (repeated from section 2.5.3).

(48) $_A$[ *wajpíi-mu* *imíllé* $_B$[ *i=ímivyé-ne* $_C$[ *jaá* *i=myéénu-ne* ] ] ]
man-PL want 3COR=complete-CL:IN house 3COR=make-CL:IN
'The men want to finish building this house.'

Because Bora is a language with overt subjects (though see section 2.8.1 for some exceptions), each successive instance of a subject allows for the reference of an anaphor to connect through a chain of anaphors. In (48), this allows for the anaphor in the most embedded clause ($_C$[ ]) to be bound by the subject in domain $_B$[ ], itself an anaphor which is bound by the subject in domain $_A$[ ]. This is similar to Lebeaux's (1983) (and others') work suggesting that long distance anaphora is comprised of LF movement of an anaphor in a series of local dependencies. Although the most embedded anaphor in $_C$[ ] in (48) may appear like a long distance anaphor on the surface, I argue that the series of local dependencies allows for an analysis in which the 3COR morpheme can be analyzed purely by the Binding Conditions, once we establish what constitutes the anaphor's binding domain (see section 4.4). I will argue that the 3COR marker does constitute a long distance

anaphor based on its adherence to the properties outlined in section 3.1.2, but that binding of the 3COR marker is based on local dependency.

This raises an interesting difference in the behavior of anaphoric binding between Bora and languages like English. Consider the sentence in (49) and its English gloss, and compare it to the ungrammatical English equivalent in (50).

(49) *wajpi*$_i$ *neé* *wállee-ke*$_j$ *i*$_{i/*j}$*=ímillé* *i*$_{i/*j}$*=májcho-ne*
man say woman-ACC 3COR=want 3COR=eat-CL:IN
'The man told the woman that he wants to eat.'

(50) **The man told the woman that himself wants to eat.'*

While in Bora, the embedded subject anaphor can only refer to the main clause subject, this is not possible in English. The sentence in both languages contains a finite embedded clause, and should, in theory, be subject to the same binding conditions with the same binding domains. I return to this topic in section 4.4, in order to use data from both 3rd person, and 1st and 2nd person non-singular subjects as evidence for the analysis.

#### 4.3.2.2. Bora Proclitics as Anaphors under RR Binding Conditions

Having established that the 3COR marker behaves as an anaphor given canonical binding theoretic principles, it should be the case that Reinhart and Reuland's (RR) analysis of anaphora (given in section 3.4) also holds. Recall the Conditions of their analysis, repeated below.

(51) a. A predicate is *reflexive* iff two of its arguments are coindexed.
b. A predicate (formed of P) is *reflexive-marked* iff either P is lexically reflexive, or one of P's arguments is a SELF anaphor.

(52) a. RR Condition A: A reflexive-marked syntactic predicate is reflexive.
b. RR Condition B: A reflexive semantic predicate is reflexive-marked.
(Reinhart and Reuland 1993: 662-663)

An important property of these definitions is that (51) and (52) work together. For example, the criteria in (51a) state that, in order to be semantically reflexive, two arguments of a predicate must be coindexed. Additionally, in order to be syntactically reflexive-marked, the predicate must either be lexically reflexive, or have an argument that is a complex (SELF) anaphor.

Of importance to this analysis is the difference between SE anaphors and complex anaphors. RR note the defining characteristic of a complex anaphor as being in itself an N, which combines with an SE anaphor or a pronoun to form a compound. SE anaphors and complex anaphors are equally referentially dependent on an antecedent. However, SE anaphors do not have the reflexivizing function.[60] RR explain that complex anaphors, as reflexivizers, impose identity on two arguments of a single predicate, one of which is the N in the complex anaphor compound. SE anaphors do pattern with pronominals, however, in that they are each analyzed as filling a determiner position in the syntax. For these reasons, Evaraert (1986) classifies SE anaphors as pronominal anaphors, and as a result, RR note that they are subject to RR Condition B. As a single morpheme, the 3COR marker in Bora classifies as an SE anaphor, whereas the reflexive marker *–mei* carries the reflexive function in Bora.

As a pronominal anaphor, the SE anaphor patterns with pronominals with respect to their reflexivizing function (±R).[61] However, SE anaphors also pattern with anaphors in that they are referentially dependent on an antecedent. Since the difference in the reflexiving function is what determines whether a pronoun is subject to RR Condition A or B, there must be some other difference between SE anaphors and pronominals. RR make the claim that this has to do with the pronominal properties of structural case and fully specified phi-features. This is an important distinction for languages like Dutch, as described in Chapter 3, which have a three-way distinction between complex anaphors, SE anaphors, and pronominals. The relevant example is repeated below.

(53) a. *He$_i$ accidentally assigned himself$_i$/*him$_i$ to himself$_i$.*
b. *Henk$_i$ wees zich$_i$/*hem$_i$ aan zichzelf$_i$ toe*
Henk assigned SE/3SG to himself to
'Henk assigned himself to himself.' DUTCH (Reinhart & Reuland 1993: 691)

Relevantly, the SE anaphor *zich* in (53b) is an argument of the main clause predicate, and, as an anaphor, must be licensed by RR Condition B in (52b). In order to be licensed as reflexive then, it

[60] Recall the difference between referential deficiency and phi-feature deficiency. While both SE and complex anaphors are deficient with respect to their referent, complex anaphors have specified phi-features through the element they are compounded to, while SE anaphors are deficient with respect to phi-features.
[61] (±R) in this case is intended to indicate reflexivity, not referential independence, as it was used in Chapter 3.

must either be part of a lexically reflexive predicate (which *wees* is not), or be a complex anaphor (which *zich* is not).

RR argue that SE anaphors (pronominal anaphors) are actually the entity that is subject to RR Condition B, and that pronominals are beholden to the Chain Condition, as described in section 3.4. A chain for them must be headed by a referentially independent element, meaning that the element is fully specified for phi-features, as well as for structural case.

For the sentence in (53b), the matrix clause predicate is reflexive based on one of its arguments being semantically reflexive, since it has two arguments that are coindexed via the property in (51b). This means that RR Condition B could allow either an SE anaphor or a pronoun. However, in RR's analysis, pronouns are classified as referentially independent, whereas SE anaphors are classified as referentially dependent. As a result, (53b) disallows a pronoun, since it would be a referentially independent link in a nonhead position in a chain. Since a chain must be headed by a referentially independent element, a referentially independent pronoun would form a separate chain, rather than form a chain with the intendended antecedent *Henk* in (53b).

Turning now to the data in Bora, the Bora reflexive marker *–mei* (54) (see section 2.4 for more on the reflexive marker), by nature of it being a reflexive marker, will always indicate coindexation with the subject of its clause. This meets the criterion given in (51a), and as such, *–mei* is subject to RR Condition B.

(54) *dibye* *imíllé* *i̵=i̵i̵te-meí-ñe* *mi̵i̵cúmi̵-ri*
3.M.SG want 3COR=see-REFL-CL:IN mirror-INS
'He wants to see himself in the mirror.'

The Bora preverbal subject proclitics do not indicate coindexation with an argument of the same predicate like the *–mei* morpheme[62], nor are the proclitics complex anaphors[63], meaning that the

[62] In fact, it seems that verbs in Bora that are lexically reflexive are derived, either synchronically or diachronically, from the reflexive morpheme. E.g. *nijtyu* "wash" vs. *nijtyaméi* "wash oneself".

[63] Though it is possible to have a reflexive morpheme on a predicate which takes the 3COR marker as a subject, as in the example below.

(i) *wajpi* *imíllé* *i̵=ni̵i̵tsá-meí-ñe*
man want 3COR-shave-REFL-CL:IN
'The man wants to shave (himself).'

proclitics do not meet the criteria given in (51b) for pertaining to a predicate that is reflexive-marked, and subsequently not subject to RR Condition A.

RR Condition B states the reflexive semantic predicates are reflexive-marked. Notably, the criteria in (51) refer to properties of a single predicate (P), while we have seen that the subject SE anaphor *i=* in Bora takes an argument of a different predicate than its antecedent. Consider the example in (55).

(55) *wajpi imíllé i=májcho-ne*
man want 3COR=eat-CL:IN
'The man wants to eat.'

If the Bora proclitics are not subject to RR Condition A, as stated above, they should be subject to RR Condition B. As such, the proclitic should be part of a reflexive semantic predicate in order to be reflexive-marked. The criteria in (51a) state that, in order to be semantically reflexive, two arguments of a predicate must be coindexed. However, looking at (55), the 3COR proclitic is an argument of the embedded clause, which has no second argument for the 3COR marker to be coindexed with. The intended antecedent, *wajpi*, is an argument of the main clause predicate, which the 3COR marker is not an argument of.

As a result, if the 3COR marker in Bora is indeed an SE anaphor, as I have claimed it is, then it is necessary to resort to the Chain Condition in order to license the Bora proclitics. While the Chain Condition was used in the Dutch example in (53b) in order to rule out the possibility that *zich* is a pronoun, it must be used here to rule *in* the SE anaphor nature of the 3COR marker. Because the 3COR marker is referentially dependent, any analysis that it is a pronominal is ruled out. A chain is then formed with the referentially independent antecedent of the 3COR marker as the head, allowing for the 3COR marker to be licensed.

I now turn to looking at the 3COR marker as either a possible marker of 3rd person agreement in the clause, or as a possible anaphor.

#### 4.3.2.3. Evidence against an Agreement Analysis of *i=*

Another possible analysis of the 3COR marker, besides that of anaphora, is that *i=* could be a subject agreement marker for 3rd person coreferent subjects. The topic of agreement marking versus

pronominal clitics has received some attention in the literature, e.g. by Bresnan & Mchombo (1987), Deen (2006), Baker & Kramer (2018), and others.

Baker & Kramer (2018) point out the importance of this distinction in their determination that doubled clitics in Amharic are pronouns rather than agreement markers. Although the weak crossover effects described by Baker and Kramer have not yet been tested for Bora, Baker & Kramer also point out that an important property distinguishing Amharic doubled clitics is that they are beholden to Binding Conditions ((56-57) below, repeated from section 4.1).

(56) *Lämma gäddäl-ä-w*
Lemma.M kill.PFV-3M.SG.SUB-3M.SG.OBJ
'Lemma killed him.' (him ≠ Lemma)

(57) a. **Lämma ras-u-n gäddäl-ä-w*
Lemma.M self.M-his-ACC kill.PFV-3M.SG.SUB-3M.SG.OBJ
'Lemma killed himself.'

b. *Lämma ras-u-n gäddäl-ä*
Lemma.M self.M-his-ACC kill.PFV-3M.SG.SUB
'Lemma killed himself.' AMHARIC (Baker & Kramer 2018: 1067-1068)

The clitic in question in (56-57) is the *–w* enclitic marking the object of the clause. Baker & Kramer argue that this clitic is a pronominal, and as such is subject to Condition B of conventional Binding Theory. This accounts for (56) above, since in order for the clitic to be free in its binding domain, it cannot be bound by *Lemma*. Since a c-command relationship already holds, the two cannot be coreferent, forcing *–w* to refer to another individual. In (57), because the object of the sentence is a reflexive already referring back to *Lemma*, the *–w* clitic actually forces the sentence to be ungrammatical, since the reflexive object *rasun* does not allow the *–w* clitic to refer to another individual.

Alternatively, Deen (2006) argues that the subject marker in Nairobi Swahili is an agreement marker rather than a pronoun. Deen's analysis hinges on the idea that in an analysis in which the subject marker *a-* is analyzed as a pronoun, an overt subject besides the subject marker can be pronounced with a TOPIC intonation. Deen points out that it is generally known that relativizers and answers to a question cannot be a topic. When an answer (59) to a question (58) has TOPIC intonation (indicated by the comma in the example), the resulting sentence has questionable grammaticality.

(58) *nani* *a-li-fi-ka* *mapema*
who SUB.AGR-PST-arrive-IND early
'Who arrived early?'

(59) ??*Juma,* *a-li-fi-ka* *mapema*
Juma SUB.AGR-PST-arrive-IND early
'Juma, he arrived early.' NAIROBI SWAHILI (Deen 2006: 230)

While this is an argument against Nairobi Swahili subject markers being pronouns, Deen's evidence that they are agreement markers comes from wh- questions. Deen brings up that, while relative pronouns bear a TOPIC function, wh- pronouns bear a FOCUS function, and a single argument cannot bear both a TOPIC and FOCUS function in the same clause. This means that, unlike the TOPIC subject in (59) above, the subject marker in Nairobi Swahili can co-occur with a subject wh- pronoun.

(60) *nani*$_i$ *a*$_i$*-me-end-a*
who SUB.AGR-PRES.PRF-go-IND
'Who has gone?' NAIROBI SWAHILI (Deen 2006: 229)

Based on these criteria, Bora subject clitics line up with being analyzed as pronouns (in the sense pointed out in Chapter 3 that the category of pronouns includes both pronominals and anaphors). Firstly, they have in common with the Amharic doubled clitics that they are beholden to Binding Conditions, shown in the previous two sections.

The Bora clitics also notably differ from the agreement markers in Nairobi Swahili in wh-questions. Wh- movement in Bora has direct consequences on how coreferent (and disjointly referent) subjects are expressed in the language. Recall from section 2.8 that Bora involves movement of the subject to the Spec of the INFL projection. Notably, A-movement of the subject to that position (indicated in (61) with the bolded items) does not trigger the 3COR anaphor to attach to the verb in either main or embedded clauses.

(61) [$_{INFL}$ *wajpi*$_i$ *neé* [$_{vP}$ [$_{VP}$ [$_{CP}$ [$_{INFLP}$ ***walle***$_j$ [$_{INFL}$ ~~***walle***~~ *ímillé-ne* [$_{vP}$ [$_{VP}$
man say woman want-CL:IN
[$_{CP}$ *i=májcho-ne* ]]]]]]]]
3COR$_{*i/j/*k}$=eat-CL:IN
'The man said that the woman wants to eat.'

Having established that the 3COR marker is an instance of an SE anaphor, this is not surprising. Not only is A-movement ruled out as an environment for the 3COR anaphor, but appearance of the 3COR anaphor in instances of local wh- (A'-)movement is also ruled out, since wh- movement of a subject to the C-domain does not trigger the 3COR marker.[64] Otherwise in (62) below, we would expect the subject, which has been moved out of INFL, to manifest as a 3COR marker on the verb.

(62) [$_{CP}$ *múha=né* [$_{INFL}$ [$_{vP}$ *ahdó* ]] *baajúri* ]
who=REC buy yuca
'Who bought the yuca?'

However, when a wh- element is moved out of the subject position of the embedded clause into the C domain of the higher clause, the 3COR anaphor still appears in the subject position of the embedded clause(s). So while local A' movement does not trigger the 3COR marker, as in (46) above, long distance A' movement does (63). In (63), the wh- element is serving as the antecedent for the 3COR markers on both embedded verbs.

(63) *múha* *ú=imíllé* *i=májchoo* *i=ádoo*
who 2SG=want 3COR=eat.CONJ 3COR=drink.CONJ
'Who do you want to eat and drink?'

The inability to find a unified analysis of triggering the appearance of the 3COR marker by movement indicates that the 3COR marker is not an agreement marker. We would expect the appearance of an agreement marker not to be dependent on whether there has been movement, or in the case of movement, dependent on which kind of movement (A- or A'-movement). In fact, the data in (62) and (63) show opposite behavior of whether the clitic in question can appear with a wh- pronoun. If it were an agreement marker, we would expect the 3COR marker to appear in each case of a 3rd person subject, including in (62). The data show, however, that the 3COR marker only appears in the embedded clause case in (63).

[64] There is a possible exception to this in topic and focus constructions in Bora. In the example below, Um̨nur̨i is marked by the 3COR marker on the verb in addition to being expressed overtly. More research on the left periphery in Bora is necessary to say more about this.

(i) *mú-hwúu-ké-jucó* *i=tsívá-ne* *Úmínur̨i* *įi-ná*
who-DIM-ACC-FOC 3COR=bring-CL:IN Um̨nur̨i this-CL:IN
'Whom is this Um̨nur̨i bringing…' (Thiesen and Weber 2012: 509)

Additionally, the fact that the expression of embedded clause subjects is dependent on coreference or disjoint reference is also indicative that the 3COR marker is not an agreement marker. Recall the examples below, repeated from Chapter 2.

(64) *imíllé-mé* *i=májcho-ne*
want-CL:AN.PL 3COR=eat-CL:IN
'They want to eat.'

(65) *imíllé-mé* *dibye* *májcho-ne*
want- CL:AN.PL 3.SG.M eat-CL:IN
'They want him to eat.'

If the 3COR marker were an agreement marker, we would expect it to appear in each clause with a 3rd person subject. However, it does not appear in clauses with classifier subjects (64, 65), nor does it appear in clauses with overt non-coreferent NPs as subjects, as in the embedded clause in (65).

The example in (61) raises interesting points about the binding of the embedded anaphors with respect to the subject orientation and Blocking Effect properties of anaphors, since the embedded 3COR marker can only refer to the structurally next highest subject, and not any higher subjects, even if the intervening subjects match in person and number. This makes the Bora data unlike the Mandarin Chinese data. These points are addressed in section 4.5 with respect to what constitutes a binding domain in Bora.

#### 4.3.2.4. Evidence against an Analysis of Valency-Changing Morphology

One further possible analysis is one in which the 3COR marker has a valency reducing property, of the kind outlined in section 2.4. That is, in the same way the *–mei* reduces the valency of the verb in reflexives and passives by eliminating the patient and conflating the agent and patient roles, could it be possible the *i=* does the same by eliminating the agent?

It is unlikely that this is the case. Valency changing morphology is highly productive in Bora, and each instance where a valency changing suffix appears has a counterpart where it does not appear.

(66) a. *wajpi* *tsájtyé-hi* *wajácu-hámi*
man carry-PRED know-CL:leaf
'The man carried the book.'

b. *wajpi* *tsájtyé-meí-hi*
man carry-REFL-PRED
'The man carried himself.'

(67) a. *ó=dsɨ̈né-hi*
1SG=run-PRED
'I ran.'
b. *wajpi o-ke dsɨ́́ne-tsó-hi*
man 1SG-ACC run-CAUS-PRED
'The man made me run.' (Seifart 2015b: 1479-1480)

The data for the 3COR marker, however, do not illustrate the same dichotomy. Consider the data in (68), repeated from (33) above.

(68) a. *walle ijchívye i=májcho-tsíh-dyu*
woman leave 3COR=eat-CL:place-ABL
'The woman left after eating.'
b. *walle ijchívye áálle májcho-tsíh-dyu*
woman leave that.one.F eat-CL:place-ABL
'The woman$_i$ left after she$_j$ ate.'

In a sense, the REFL and the 3COR marker share some properties in common. They both indicate coreference, and the argument of the predicate that they represent is not overtly expressed otherwise. Additionally, if the argument represented by the 3COR marker is taken to be suppressed, then the argument is taken to be coreferent with its antecedent, similarly to the REFL marker.

The semantics in (66) and (67) above show clearly that one of the arguments has been omitted, in the case of (66), or added, in the case of (67). Rather, in (68a), although the subject of the embedded clause verb is pronounced in Bora but is not pronounced in the English equivalent, the embedded subject still receives a theta role, though in theory this could also be said of the REFL marker.[65]

It would seem strange, however, for only the 3rd person coreferent clitic to act as a valency changing morpheme, and for the other proclitics being investigated not to act in this way. We have already seen in section 4.3.1 that the 1st and 2nd person singular proclitics behave as pronominals, and are therefore overt arguments of the predicates they appear in. Additionally, as seen in earlier Chapters, and will be seen in section 4.3.3, the SAP marker often appears with an overt expression of a noun phrase in the same clause. With the argument being overt in these cases, it is clear that

[65] This is given the PRO theory of control in English which abides by the theta criterion.

the argument is not being suppressed. It is apparent then that there is no argument being eliminated in cases of the 3COR marker in Bora and that Bora proclitics are not valency-reducing morphemes.

#### 4.3.2.5. Revisiting the Subject Antecedent Property of Anaphors

Assuming an analysis in which the 3COR marker acts as an anaphor, recall the phenomenon in (61) above in which the antecedent of the 3COR marker can only be the structurally next highest subject, and not any subject beyond that. This calls into question the long distance anaphor property of taking an antecedent that is a subject (as described in section 3.1.2). Notably, the subject of the matrix clause of the sentence in (61) is not the antecedent of the embedded 3COR marker. This raises the question of what subject properties are necessary for an NP to serve as the antecedent for an anaphor.

Cole, Hermon, & Huang (2006) revisit the subject antecedent property of long distance anaphors through the lens of long distance anaphora in Asian languages. They note that there are apparent exceptions in Mandarin, where subject orientation seems to hold in some cases (69a), but not in others (69b).

(69) a. *Wangwu*$_i$ *shuo* *Zhangsan*$_j$ *zengsong* *gei* *Lisi*$_k$ *yipian* *guanyu* *ziji*$_{i/j/*k}$
Wangwu says Zhangsan give to Lisi one about
*de* *wenzang*
DE article
'Wangwu says that Zhangsan gave an article about him/himself to Lisi.'

b. *Zhangsan*$_i$ *yiwei* *Lisi*$_j$ *hui* *ba* *Xiaoming*$_k$ *dai* *hui* *ziji*$_{i/j/k}$
Zhangsan thought Lisi will BA Xiaoming take self self's
*de* *jia*
DE home
'Zhangsan thought Lisi would take Xiao Ming back to his home.'

(Cole, Hermon, & Huang 2006: 40)

Cole, Hermon, and Huang argue that the difference illustrated in (69a-b) is reducible to c-command, in that the potential antecedent must also c-command the anaphor. Since *Lisi*, as part of a prepositional phrase, does not c-command *ziji* in (69a), *Lisi* is ruled out as a potential antecedent. In contrast, although *Xiaoming* is not a subject in (69b), *ziji* is still c-commanded by *Xiaoming* and can still take *Xiaoming* as an antecedent.

The *i*= anaphor in Bora does not seem to be reducible solely to c-command, however. Consider (22) above, repeated here as (70). Although *wálleeke* c-commands the anaphoric subjects given the analysis of Bora syntactic structure given in section 2.8, both *i*=, in the embedded clauses, it cannot serve as their antecedent; only the subject of the main clause, *wajpi*, can bind the embedded anaphoric subjects.

(70) *wajpi*$_i$ *neé wállee-ke*$_j$ *i*$_{i/*j}$*=ímillé* *i*$_{i/*j}$*=májcho-ne*
man say woman-ACC 3COR=want 3COR=eat-CL:IN
'The man told the woman that he wants to eat.'

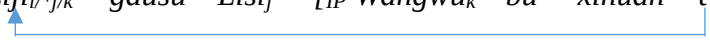
(71) *wajpí*$_i$*=ñe táúmeí ovátsa*$_j$*-ke dibye*$_{*i/j}$ *i-hjya ímibájacho*
man=REC ask.for youth-ACC 3.M.SG 3.POSS-house repair
'The man asked the youth to fix his house.'[66]

Notably, (71) shows that, when the object of the main clause serves as the antecedent for the embedded clause subject, the embedded clause subject is expressed by an overt pronoun rather than a preverbal subject proclitic. This seems to indicate that pure c-command is not sufficient for an NP serve as the antecedent for an embedded subject proclitic; rather subjecthood seems to be a relevant property as well.

For the cases in Mandarin which do not allow objects to serve as antecedents, as in (72) below, Cole, Hermon, and Huang propose movement of *ziji* at LF to a functional projection in the higher clause, in line with the analysis of Huang & Tang (1991) presented in Chapter 3. The LF representation of (72) is given in (73).

(72) *Zhangsan*$_i$ *gausu Lisi*$_j$ *Wangwu*$_k$ *bu xihuan ziji*$_{i/*j/k}$
Zhangsan told Lisi Wangwu not like self
'Zhangsan told Lisi that Wangwu does not like him/himself.'

(73) [$_{IP}$ *Zhangsan*$_i$ *ziji*$_{i/*j/k}$ *gausu Lisi*$_j$ [$_{IP}$ *Wangwu*$_k$ *bu xihuan* t ]]



This analysis accomplishes two things: (i) it explains why an object cannot serve as the long distance antecedent for an embedded anaphor, since the anaphor that raises at LF raises to a functional projection higher than the object, assuming an analysis like that for Mandarin in Chapter 3, in which the anaphor adjoins to he INFL head[67], and (ii) it explains how the wh- element in (62)

[66] Interestingly, the 3.POSS marker in (71) is ambiguous between referring to the man or the youth.
[67] Cole, Hermon, & Huang (2006) explain that the object *Xiaoming* in (69b) is an acceptable antecedent because it is marked with *ba*, which marks preverbal objects. In Mandarin, *ba* projects its own functional projection (BaP in Cole,

above is able to serve as the antecedent for the embedded 3COR anaphors despite the wh- element not being a subject. I have also argued in this section that the 3COR marker is an SE anaphor, and I have provided evidence for this using both canonical binding theory and RR's Reflexivity Conditions. Additionally, although the 3COR marker shows aspects of long distance anaphora (such as the requirement of being bound by a subject), I follow Lebeaux's (1983) work in analyzing such anaphora as a series of local dependencies. I have also noted that the 3COR marker is not subject to the Blocking Effect in the same was as has been shown for Mandarin Chinese. I return to my proposal for the reason for this in section 4.4.

### 4.3.3. 1st and 2nd Person Non-Singular

The 1st and 2nd person non-singular subjects in Bora are represented in any clause by the SAP marker *me=*, which must be cliticized onto the verb for any 1st or 2nd person dual or plural overt subject, whether that subject has been expressed by an overt pronoun or not (as in the case of 1st person plural inclusive subjects, see section 2.3.2).[68] The relevant examples are repeated below.

(74) *muhá=ne-cu* *me=tsaa* *iquito-vu* *wákimyéi-vu*
1PL=REC-STR SAP=come Iquitos-ALL work-ALL
'We came to Iquitos to work.'

(75) *ímí* *me=cúwá-hajchíí* *tsá* *me=chémé-i-tyu-ró-hi*
good SAP=sleep-COND NEG SAP=be.ill-FUT-NEG-FRS-PRED
a. 'If we (incl.) sleep well, we are not likely to get sick.'
b. 'Whoever sleeps well is not likely to get sick.' (Thiesen and Weber 2012: 127)

In this section, I focus on instances of the SAP marker which indicate coreference across clauses, rather than the other uses pointed out in section 2.5.2. Following that, I point out differences between the SAP and 3COR markers with respect to whether A and/or A' movement trigger the appearance of the proclitic. I then show that, like the 3COR marker, the SAP is not an agreement marker, but rather an anaphor that is subject to Binding Conditions.

I focus primarily on the use of the SAP that indicates coreference with the subject of the main clause ((76), repeated from Chapter 2).

(76) *muúha* *me=imíllé* *me=íjchi-ñe*

Hermon, & Huang's analysis), and the Spec and head of this functional projection c-command *ziji* even after *ziji* has undergone movement at LF.

[68] Recall that there are instances of the SAP marker that are used for impersonal subjects and weather verbs. I do not consider these uses to fall under the same analysis presented in the rest of this section.

1.PL SAP=want SAP=swim-CL:IN
'We want to swim.'

At first glance, it appears possible that the embedded SAP marker in (76) is not necessarily indicative of coreference, but it could also be that the SAP in the embedded clause indicates a 1st person plural inclusive subject. However, the same pattern in (76) appears across the paradigm for 1st and 2nd person non-singular embedded subjects that are coreferent with the subject of the higher clause, all of which use the SAP in the embedded clause to indicate that coreference.

(77) a. *muhtsi* *mé=imíllé* ***me****=íjchi-ñe*
1M.DU SAP=want SAP=swim-CL:IN
'We both (masc.) want to swim.'
b. *ámuhpɨ* *mé=imíllé* ***me****=íjchi-ñe*
2F.DU SAP=want SAP=swim-CL:IN
'You two (fem.) want to swim.'
c. *ámuúha* *mé=imíllé* ***me****=íjchi-ñe*
2PL SAP=want SAP=swim-CL:IN
'You all want to swim.'

Given the fact that in the examples in (77) do not give an interpretation of 1st person plural inclusive embedded subjects[69], it is clear that there is a use of the SAP marker that marks coreference with a higher clause subject in the absence of an embedded personal subject pronoun.[70]

However, unlike the 3COR marker described in the previous section, every clause with a 1st or 2nd person non-singular subject in Bora must be marked with the SAP, whereas the 3COR anaphor

[69] All of the examples in (77) seem to indicate exhaustive control. Partial control was difficult to elicit in Bora, since the typical examples using "together" do not quite carry the same meaning in Bora. The word that speakers used to translate "together" was *tsaíjyu* "at the same time". Other examples, like (i) below, seem to indicate that there may not be partial control in the language, in that the embedded clause 3COR subject likely refers only to the youth, though a partial control analysis would predict the embedded clause subject to refer to the group that built the house. Future research may be able to expand on this.

(i) *ovátsa* *píaabó* *jáá* *i=myeénu*
youth help house 3COR=make
'The youth helped build the house.'

[70] As with the examples of 1st and 2nd person singular subject pronouns in section 4.2.1, the differences in tone on the SAP are a result of the type of clause (main or embedded) that the clitic appears in.

appears only in embedded clauses. The SAP can, however, continue to mark coreference across several clauses, similarly to the 3COR anaphor in (48) above.

(78) *múha mé=imíllé me=wájacú-ne muhdú jaa me=méénu-ne*
1PL SAP=want SAP=know-CL:IN how house SAP=make-CL:IN
'We want to know how to build a house.'

Similarly to the 3COR anaphor in section 4.3.2, the SAP appears without an overt subject in embedded clauses with subjects that are coreferent with the subject of the main clause.

(79) *múha mé=imíllé me=májcho-ne*
1PL SAP=want SAP=eat-CL:IN
'We want to eat.'
(80) *wajpi imíllé i=májcho-ne*
man want 3COR=eat-CL:IN
'The man wants to eat.'

The SAP marker differs from the 3COR marker, however, with regard to other properties of movement of the subject. Recall from section 4.3.2 that A movement and local A' movement of the subject in a clause do not trigger appearance of the 3COR marker. This is not true of the SAP marker, however. Assuming movement of the subject to the Spec, INFLP position, the SAP marker still attaches to the verb, which is not the case for the 3COR anaphor. Compare (81), which shows local A movement of the 1st person plural subject (evidenced by the ability of the adverb to appear between the overt subject and the SAP marker), to (82), which shows the same for a 3rd person subject. Movement of the subject to the Spec, INFLP was established in section 2.8.

(81) *muuhá=pe iíjyu mé=majchó-h*
1.PL=REM yesterday SAP=eat-PRED
'We ate yesterday.'
(82) *wajpi majchó-hi*
man eat-PRED
'The man is eating.'

Local A' movement of an SAP subject is harder to show, since wh- question movement is restricted to 3rd person, and work on topic and focus markers in Bora requires further investigation. Another place one could expect to find A' movement would be in relative clauses. However, recall from

section 2.9.3 that Bora does not raise a relative pronoun out of relative clauses in the same way English does. Rather, the entire relative clause is marked with a classifier, which refers to the syntactic element being relativized. Notice in (83) that, given this, it is not the case that *muha* has raised out of the relative clause. This does not allow for an analysis of relative clauses as A' movement in a straightforward way, though further research on properties of islands in Bora would help inform an analysis.

(83) *[tsɨ́ɨ́me-ke muha me=úwaabo]-me tsá tsí-ñé-hjɨ́-ri*
children-ACC 1.PL SAP=teach-CL:AN.PL NEG other-EVENT-PL-INS
*me=íjcya-tu-ne*
SAP=be-NEG-CL:IN
'We who teach our children do not get involved in other things.'
(Thiesen & Weber 2012: 391)

Second position clitics in Bora do provide a clue regarding A' movement. Following the description provided in section 2.8, I assume for the sake of this analysis that the host of the second position clitic is an element in the C domain.

As far as whether the clitic host has undergone local A or A' movement, Thiesen & Weber's claim that second position clitics attach to the first constituent rather than the first word of a clause indicates that the clitic host has undergone A' movement to Spec, CP. They give the following example with the reportative *=va* clitic (though they do not offer a semantic distinction between (84a) and (84b)), in which the clitic does not follow the first word of the clause.

(84) a. *tsá o=va o=péé-i-tyú-ne*
NEG 1SG=RPT 1SG=go-FUT-NEG-EVENT
'Someone said that I am not going.'
b. *tsáha=va o=péé-i-tyú-ne*
NEG=RPT 1SG=go-FUT-NEG-EVENT
'Someone said that I am not going.' (Thiesen & Weber 2012: 308)

Even in cases when the subject A' moves to a higher position in order to host a second position clitic, the verb still takes the SAP as a proclitic.

(85) *muuhá=pe iíjyu mé=majchó-h*
1PL=REM yesterday SAP=eat-PRED
'We ate yesterday.'

As with the 3COR marker, there is a possibility that the SAP could be an agreement marker or an anaphor. Recall that one of the primary arguments that the 3COR marker was not an agreement marker was that local A and A' movement in general did not trigger appearance of the 3COR marker. Another argument was that the 3COR marker is not triggered only by phi-features, as we would expect for an agreement marker, but rather also by coreference with the antecedent. That is, the 3COR marker does not appear in every case of an embedded 3rd person subject. Rather, in cases of disjoint reference with a 3rd person NP as the structurally next highest subject, the 3COR marker cannot appear, indicating that the appearance of the 3COR marker is not agreement. Neither of these arguments hold for the SAP marker. As shown above, the SAP marker appears independently of both local A and A' movement, and the SAP marker will always appear with a 1st or 2nd person non-singular subject, regardless of whether that subject is coreferent with the subject of a higher clause.

This might initially hint that the SAP marker is in fact an agreement marker, and not a pronoun, though further investigation reveals that this is not the case. Recall from (77a-c) above that the SAP in the embedded clause actually picks out the referent for the embedded clause subject from the main clause subject in cases of coreference. An agreement marker would only provide phi-feature agreement for the subject, but would not pick out a referent. In fact, recall from Chapter 2 that it is ungrammatical for an embedded SAP subject that is coreferent with the matrix clause subject to appear together with an overt subject pronoun (86), whereas an overt pronoun *is* used in the embedded clause if the embedded subject is disjointly referent with the higher clause subject, as in (87).

(86) ***múha**$_i$ *mé=imíllé* *muha*$_i$ *me=májcho-ne*
1.PL SAP=want 1.PL SAP=eat-CL:IN
'We want to eat.'

(87) *múha* *mé=wajácú* *ámuha* *jáá-hañe* *me=méénu-ne*
1.PL SAP=know 2.PL house-VAR SAP=make-CL:IN
'We know that you all build houses.'

The ungrammaticality of (86) shows that there must be restrictions on the appearance of overt subjects in embedded clauses with 1st and 2nd person non-singular subjects. This is reminiscent of

the restriction that 3rd person embedded coreferent subjects are beholden to. If the appearance of an overt subject in the embedded clause is dependent on the referent of the subject rather than on the subject's phi-features, it is likely that what is at play are binding conditions, which are dependent on coreference.

If the SAP marker, like the 3COR marker, is anaphoric, then their different distributions still need to be explained. That is, why does the SAP marker occur in each clause with a 1st or 2nd person non-singular subject, while the 3COR marker only appears in embedded clauses with subjects that are coreferent with a higher clause subject? I argue that this is because the SAP can be both a local and long distance anaphor (recall the Inuit example in Chapter 3, repeated below), while the 3COR marker is only a long distance anaphor, as argued in section 4.3.2, in the sense that the 3COR marker is not bound within its local clausal domain (though, as was pointed out in section 4.3.2, the 3COR marker does have restrictions on how 'long distance' the antecedent of the anaphor can be).

(88) *Kaali$_i$* *uqar-p-u-q* *Pavia$_k$* *immi-nit$_{i/k}$* *angi-nir-u-sinnaa-nngi-tsu-q*
Kaali say-IND-3SG Pavia self-ABL big-CMP-be-can-NEG-PRT-3SG
'Kaali said that Pavia couldn't be taller than self/him.'
INUIT (Huang 2000: 94, citing Bittner 1994: 147)

In (88), the embedded anaphor *immi-nit* can be either locally bound by *Pavia*, or long distance bound by *Kaali*. While similar in this sense, the ambiguity that is present in (88) does not manifest in Bora. That is, in Bora, the coreferent use of *me*=, when there is no local NP to bind the SAP, must be bound at a distance (though recall from section 2.3.2 that the SAP can be ambiguous between the coreferent and impersonal readings). Conversely, when coreference is not being indicated with a higher subject, the SAP must be locally bound by an antecedent. The sentence in (89) shows two instances of the SAP marker that are both locally bound by the subjects of their respective clauses, while the sentence in (90) shows an embedded SAP marker that is bound at a distance by the main clause subject with which it is coreferent, as well as a main clause SAP that is locally bound by *múha*.

(89) *múha* *mé=wajácú* *ámuha* *jáá-hañe* *me=méénu-ne*
1.PL SAP=know 2.PL house-VAR SAP=make-CL:IN
'We know that you all build houses.'

(90) *múha* *mé=imíllé* *me=májcho-ne*
1PL SAP=want SAP=eat-CL:IN

'We want to eat.'

Recall also that there is an ambiguity regarding the SAP marker, in that it can also be used to indicate a 1st person plural inclusive subject, as in (91), repeated from Chapter 2.

(91) *ímí* *me=cúwá-hajchíí* *tsá* *me=chémé-í-tyu-ró-hi*
good SAP=sleep-COND NEG SAP=be.ill-FUT-NEG-FRS-PRED
'If we (incl.) sleep well, we are not likely to get sick.' (Thiesen and Weber 2012: 127)

In this example, the SAP marker in the main clause does not find any overt antecedent. As such, it does not appear to be bound in the same way as the SAP seen so far in this Chapter. In fact, the 1st person plural inclusive reading appears to be getting its referent from the discourse in the same way as the 1st and 2nd person singular subjects outlined in section 4.3.1. In this way, the 1st person plural inclusive use of the SAP marker seems to have properties of a pronominal, in that it is free within its binding domain.

Despite being different types of anaphors, both the 3COR and SAP markers are still subject to the same binding conditions. The domain in which they are bound, however, seems to differ from other languages (recall (49) and (50) from above).

This section has analyzed Bora preverbal subject proclitics in terms of binding theory, as well as RR's theory of reflexivity. Whereas the data did not conform to theories of control, the proclitics do seem to conform to theories of binding and reflexivity, with 1st and 2nd person singular proclitics behaving as pronominals, and the SAP and 3COR markers behaving for the most part as anaphors. I have also argued against the proclitics being analyzed as agreement markers or a type of valency-changing morphology. I return now to a discussion of binding domains specific to Bora, and how they account for the data analyzed so far in this Chapter.

### 4.4. A Return to Binding Domains

Recall that the Bora 3COR anaphor seems to be subject to a different binding domain than a similar anaphor in a language like English ((49-50) above, repeated here), or at least the domains that it is bound in have different properties.

(92) *wajpi*$_i$ *neé* *wállee-ke*$_j$ *i*$_{i/*j}$*=ímillé* *i*$_{i/*j}$*=májcho-ne*
man say woman-ACC 3COR=want 3COR=eat-CL:IN

'The man told the woman that he wants to eat.'

(93) **The man told the woman that himself wants to eat.'*

To account for this, I look more closely in this section at the properties of embedded clauses in Bora, and how they compare to English. I first consider whether Bora embedded clauses may be deficient domains, similar to English ECM constructions. After showing that this is likely not the case, I consider that embedded clauses in Bora may not necessarily be completed phases, in that there may remain unvalued features on the subject of the embedded clause.

First, recall from Chapter 2 the claim that all clauses in Bora are finite, indicating a tensed INFL projection. Additionally, since (with few exceptions) every clause in Bora must have an overt subject, that subject must receive Case. It is a widely held view (following Chomsky 2008) that uninterpretable features on INFL are inherited from C. The fact that each embedded clause contains an overt subject means that the subject must be receiving Case from somewhere. While subjects generally receive their Case from INFL via feature inheritance, there are examples of subjects in English that receive their Case by other means. Recall from Chapter 3 the so-called exceptional case marking (ECM) clauses allow the embedded clause subject to receive its Case from the *v*P of the higher clause, although the theta role of the pronoun comes from the embedded clause. Compare the ECM construction in (94a) with the equivalent full clause in (94b).

(94) a. *The judge believes **him** to be guilty*.
b. *The judge believes that **he** is guilty*.

ECM constructions in English have traditionally been analyzed with the embedded clause extending up to an INFLP. This means that INFL does not receive Case via inheritance to assign to the embedded clause subject, allowing the subject to receive its Case from the higher clause instead. This does not seem to be the case in Bora. Although *v*P assigns accusative case to animate objects, as in (92) above, embedded animate subjects in Bora do not receive that accusative Case.

(95) a. *mé=wajácú* ***diityé**=pe* *amóóbe-ke* *dóó-ne*
SAP=know 3.PL=REM fish-ACC eat.meat-CL:IN
'We know that they ate fish.'
b. *aatye* *wajácú* *iijyú=pe* *amóóbe-ke* ***me**=dóó-ne*
those know yesterday=REM fish-ACC SAP=eat.meat-CL:IN
'Those ones know that we (incl.) ate fish yesterday.'

Importantly, as was pointed out in Chapter 3, ECM constructions allow for the exceptionally Case-marked noun to essentially be part of two different binding domains. This sort of deficiency in the functional clausal structure (i.e. the lack of a CP projection in ECM cases) allows for the binding domain to extend beyond the clause.

The idea of deficient functional projections allowing for the binding domain to extend beyond an embedded clause is not limited to ECM constructions, however. In fact, in Chapter 3 it was pointed out that Cole, Hermon, & Huang (2006) described the tendency that long distance anaphora tends to be restricted to environments in which the anaphor and its antecedent occur in specific domains, "i.e., specific types of IPs such as infinitival or subjunctive" (Cole, Hermon, & Huang 2006: 23). In order to show that Bora embedded clauses are not defective domains, I show the properties that embedded clauses in Bora have, and how those differ from what would be expected for a defective domain.

The nature of the deficiency of the domain may determine to what extent the domain allows for the binding domain to extend into the higher clause. Pires (2006) for example, shows that there is a class of obligatory control infinitives in Brazilian Portuguese that allow a higher clause to act as the binding domain for the embedded clause. Another class of embedded infinitives, inflected infinitives, can instead form their own independent binding domain.

I have already shown that Case on embedded subjects in Bora comes from the embedded clause verb itself, meaning that an embedded clause subject receives nominative Case, and not accusative Case which would indicate that Case may be coming from the matrix clause verb. It is also the case that embedded clauses carry their own Tense information as well (see section 2.6). In Bora, the only tenses that are marked overtly are recent past, remote past, and future, with the two past tenses occurring as second position clitics, and the future tense marker as a verbal suffix. These tenses appear in embedded clauses in Bora, both in cases of coreference (96), and disjoint reference (97).

(96) *mé=wajácú diityé=pe amóóbe-ke dóó-ne*
SAP=know 3.PL=REM fish-ACC eat.meat-CL:IN
'We know that they ate fish.'

(97) *wajpi tsá ílli-tyú-né ɨ=dsɨjɨvé-i-yó-ne*

man NEG fear-NEG-CL:IN 3COR=die-FUT-FRS-CL:IN
'The man is not afraid to die.'

It is also necessary to make it clear that embedded clause subjects in Bora do not raise any higher than INFLP; that is, they do not have to sit in the C projection at the edge of the embedded clause (unless they are hosting 2nd position clitics or are being topicalized), which would in principle allow them to access the higher clause as a binding domain. This is evident from indirect questions in Bora, in which the wh- element sits in the Spec, CP position at the edge of the clause (92a-b), eliminating the possibility of the subject filling that position. Recall from example (61) in Chapter 2 that the Spec, CP position is also the position in embedded clauses that hosts second position clitics, which are also shown in (98b) with the 'nonwitnessed' evidential clitic.

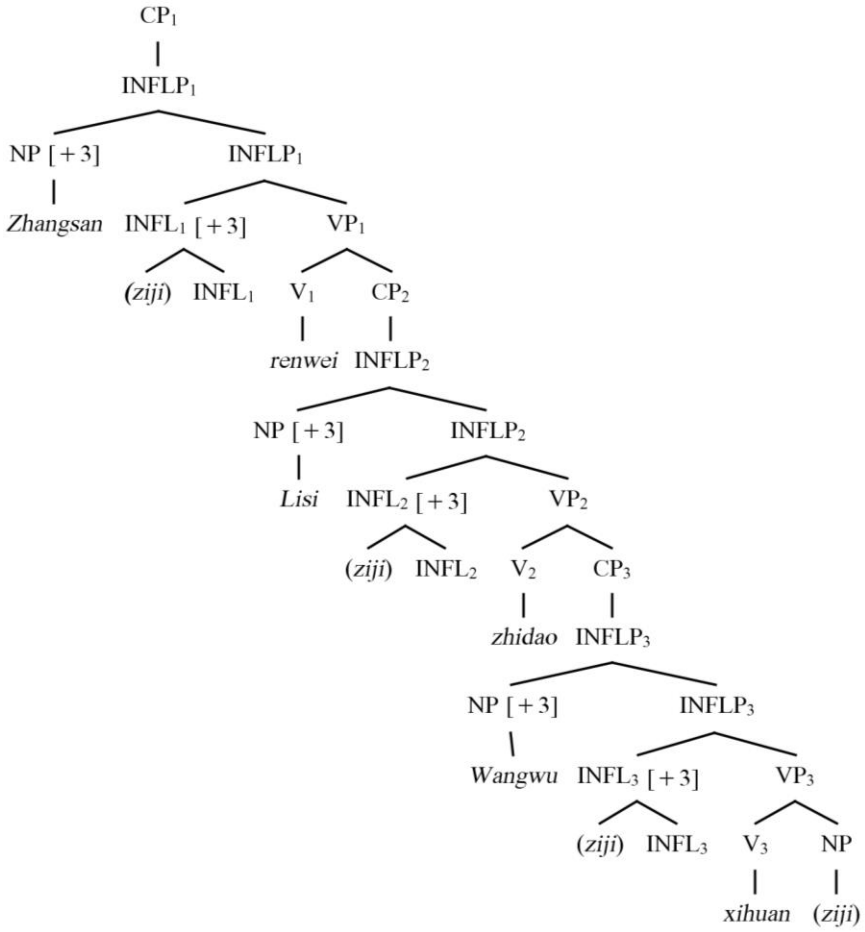
(98) a. *tsá* *o=wájácu-tu* *múha* *dsɨ́jɨve-ne*
NEG 1SG=know-NEG who die-CL:IN
'I don't know who died.'

b. *tsá* *o=wájácu-tu* *muhdú=hja=né* *wajpi* *méénu-ne* *mɨ́ɨne*
NEG 1SG=know-NEG how=NWIT=REC man make-CL:IN canoe
'I don't know how the man made the canoe.'

So far, I have established that embedded clauses in Bora, being tensed finite clauses (as also shown in section 2.6), assign Case to their own subjects, and are full CP's. None of these factors indicate any sort of defective embedded clause. The only property that would seem to suggest that these clauses are defective is that the proclitic subjects of embedded clauses in Bora can, and in many cases, must be bound outside of their own CP, which traditionally acts as a binding domain in other languages. This differentiates the Bora 3COR and SAP proclitics from pronominals that can be bound as long as they are not bound locally, though crucially pronominals do not need to be bound. I argue that the property of Bora proclitics needing to be bound outside of their local CP can be accounted for by the same means as other long distance anaphors, through raising of the anaphor at LF. Recall the Mandarin data from section 3.3.1 from Huang & Tang (1991).

(99) a. *Zhangsan*$_i$ *renwei* *Lisi*$_j$ *zhidao* *Wangwu*$_k$ *xihuan* *ziji*$_{i/j/k}$
Zhangsan think Lisi know Wangwu like self
'Zhangsan thinks Lisi knows Wangwu likes self.'

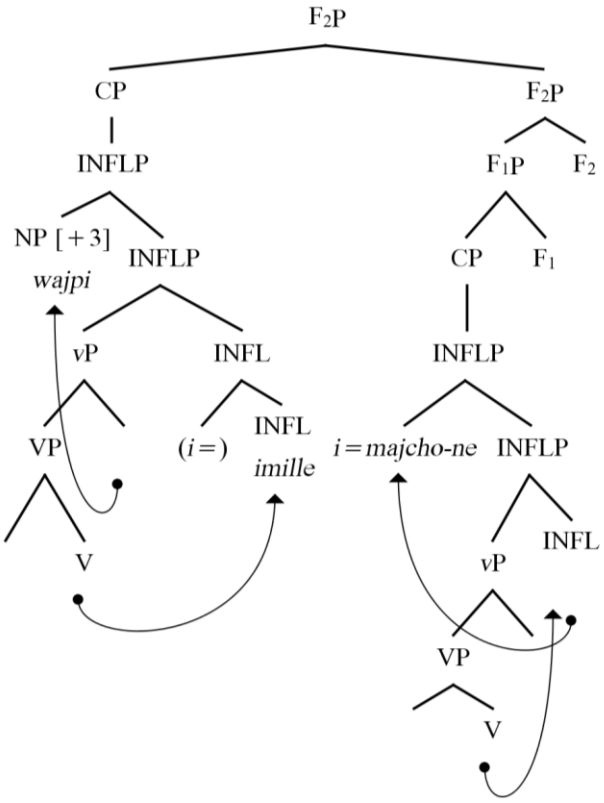
(100)




In the Mandarin data, the anaphor *ziji* covertly raises to INFL projections through the derivation, allowing for each of the subjects located in the INFL projections to act as a potential binder for the anaphor, so long as no subject of a different number intervenes. For the Bora 3COR marker, I propose that this covert movement *must* happen in order for the anaphor to be bound. Differently from Mandarin, however, once the raised anaphor encounters a higher subject, it cannot raise any further. This can be reformulated in a similar way to Mandarin's Blocking Effect. Whereas in Mandarin, any subject of a different person will block the anaphor from raising further, in Bora any subject, regardless of person, will block the anaphor from raising further. A representation of the Bora case is given below.

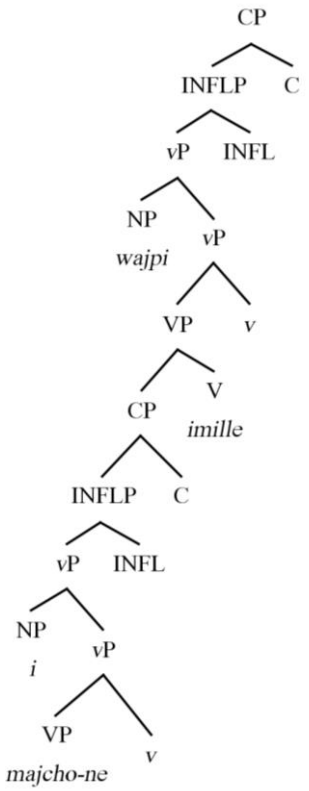
(101) *wajpi imíllé i=májcho-ne*
man want 3COR=eat-CL:IN
'The man wants to eat.'

(102)




Because of the remnant movement in Bora syntax, it does not appear, given the syntactic representation in (96), that the antecedent of the 3COR anaphor actually c-commands the anaphor, allowing binding to take place. (103) below shows the syntactic representation before any movement (Internal Merge) applications having taken place, so that c-command relations can be demonstrated.

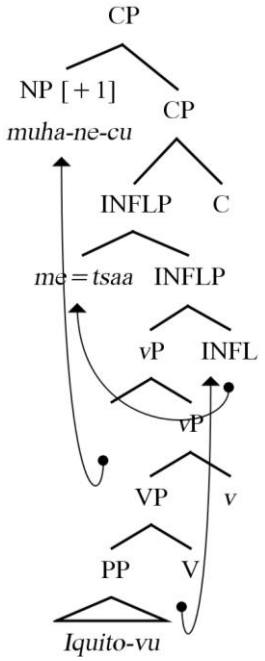
(103)




Recall that the difference between the 3COR marker and the SAP marker is that the SAP, in disjoint reference cases and in main clauses, must be bound in its own clause, which is not the case for the 3COR marker. While coreferent cases will function in the same way as (102), the disjoint reference and main clause cases of the SAP will be bound by subjects within the same clause.

(104) *muhá=ne-cu* *me=tsaa* *iquito-vu*
1PL=REC-STR SAP=come Iquitos-ALL
'We came to Iquitos to work.'

(105)




Following this analysis, the Bora 3COR marker and the SAP marker follow the same type of analysis as other languages with long distance anaphora. This analysis for Bora is then based on the anaphors themselves, rather than the clause-type. I showed that while defective embedded clauses was a possible analysis for the necessary binding of Bora proclitics outside of their local clausal domains, such an analysis could not capture all and only the cases of cross-clausal coreference indicated by the embedded clause subject proclitics, while also accounting for cases of the SAP marker in main clauses and disjoint referent embedded clauses, where it must be locally bound. That is, a theory of defective embedded clauses could not account for some proclitics being necessarily bound outside their local clausal domain while also accounting for some proclitics that must be bound in their local clausal domain.

This section has provided a structural analysis of long distance anaphora in Bora as it relates to the SAP and 3COR markers. Following Huang & Tang's (1991) analysis, provided in Chapter 3, I show that these proclitics in Bora are analyzable through covert raising at LF to the INFL projection of the higher clause. This is consistent with that way the long distance anaphora is analyzed in Mandarin Chinese, although the Blocking Effect shown by the Bora proclitics manifests differently. I now turn to cases where the current analysis may still leave open questions, leaving topics to be answered by future research.

### 4.5. Problematic Cases and Directions for Further Research

Recall that Thiesen & Weber (2012) analyze both the 3COR marker and the 3rd person possessive marker as the same anaphoric morpheme, what they call the 'self' marker (I continue to refer to the possessive use as the 3POSS marker and the anaphoric subject use as the 3COR marker).

(106) *caámé-ne* *i-úniba* *caáme-vu* *pɨjɨhtúcuu-be*
upward-CL:IN 3POSS-lip upward-ALL turn.inside.out-CL:IN
'He inverted his lip upward…'

They also point out a case that is problematic for an analysis of this morpheme as an anaphor, specifically in some instances when it acts as a possessive. In (106), the clitic's referent does not c-command the clitic itself, and thus would not meet the requirements for binding in canonical binding theory.

(107) *í-oohíí-bye* *Jóáá-ke* *ɨhdó-hi*
3COR-dog-CL:M.SG John-ACC bite-PRED
'His$_i$ dog bit John$_i$.'

Notably, this is not problematic according to Reinhart and Reuland's (1993) analysis of anaphora, since their conditions (see section 3.4) do not make any reference to c-command or syntactic hierarchy.[71] Rather, this would create a problem for a canonical binding theory analysis, since, in order to be bound, the 3COR marker must be c-commanded by its antecedent, which is not the case in (107). Further investigation of the 3COR marker as a possible possessive anaphor could shed more light on this, possibly leading to a unified analysis.

Additionally, recall from section 2.9.1 that there are some instances of conditional clauses in which the protasis precedes the apodosis and the two have coreferent 3rd person subjects, but the apodosis does not have an overtly expressed subject. In (108b), *wajpi* is the only available subject in the sentence, and serves as either the subject of the main or embedded clause, but not both.

(108) a. *wajpi* *májcho* *ímí* *i=táává-hajchíí*
man eat well 3COR=hunt-COND
'The man will eat well if he hunts.'

[71] RR state that they believe that "all hierarchical effects follow independently from the Chain Condition" (Reinhart & Reuland 1993: 712).

b. *wajpi* *táává-hajchíí* *májcho* *ímí*
man hunt-COND eat well
'If the man hunts, he will eat well.'

Hierarchically, we should expect the structure of both sentences in (108) to be equivalent. Given the syntactic analysis given in section 4.2.2 for the 3COR marker, linear sentential order should not play a role. However, we see other sentences in which other types of embedded clauses precede the main clause with an overtly expressed 3COR subject. For instance, in (109), although the 3rd person plural subject is overtly expressed by the classifier in the main clause, the embedded clause uses an overt personal pronoun, where we would have expected a 3COR marker in the case of coreference. It may be the case that the two subjects were intended to show disjoint reference, but this was not listed in Thiesen & Weber (2012).

(109) [[*ditye* *tsá-cooca* ] *pee-í-myé* *ih-jyá-vu* ]
3PL come-when go-FUT-AN.PL 3POSS-house-ALL
'When they come, they will go to their house.' (Thiesen & Weber 2012: 373)

Another type of conditional clause in Bora are counterfactual conditional clauses, introduced by the suffix *–ca*. Thiesen and Weber (2012) described this marker as (i) having the truth value of the main clause being dependent on the truth of the conditional clause, and (ii) the conditional clause is false, which consequently forces the main clause to be false. Similarly to the "normal" conditional clauses seen in (108) above, the main clause does not have an overt subject when the condition precedes the main clause. Further research is necessary to determine if this is the case for the SAP marker as well.

(110) *a-ca* *i=íímú-tu-ca* *me-ke* *pɨ́áábó-i-yó-ne*
Y/N-doubt 3COR=be.savory-NEG-CF SAP-ACC help-FUT-FRS-CL:IN
'If it were not savory, would it help us?' (Thiesen & Weber 2012: 372)

In general, conditional clauses in Bora require further investigation in order to make claims about overt subjects in their respective main clauses.

Another unusual property of Bora surfaces in the distribution of overt subject pronouns. Consider again the examples below.

(111) a. *Níjɨho$_i$ ɨjtsúcunu Wajco$_j$ ímillé-ne i$_{*i/j}$=wájtsɨ-ne péjcore*
Níjɨho believe Wajco want-CL:IN 3COR=arrive-CL:IN tomorrow
'Níjɨho believes that Wajco wants to arrive tomorrow.'

b. *Níjɨho$_i$ ɨjtsámei Wajco$_j$ ímillé-ne dibye$_i$ péjcore wájtsɨ-ne*
Níjɨho think Wajco want-CL:IN 3.M.SG tomorrow arrive-CL:IN
'Níjɨho$_i$ thinks that Wajco$_j$ wants him$_{i/*j}$ to arrive tomorrow.'

I have analyzed the embedded subject proclitic in (111a) as a long distance anaphor, which takes *Wajco* in the immediately higher clause as its antecedent. Notably, in (111b), the embedded pronominal *dibye* cannot take *Wajco* as an antecedent. Taking the CP domain of its own clause as the relevant binding domain in this case, *dibye* conforms to Condition B, as it is free within its binding domain. However, Condition B does not predict that *Wajco* is not a possible antecedent for the pronominal.

One possible reason for this could be that the binding domain extends beyond the most embedded CP to include the subject of the immediately higher clause. This would predict the impossibility of the immediately higher clause subject acting as the antecedent for embedded pronominals, and reduce the analysis of subject proclitics from long distance binding to binding within a binding domain. Such an analysis, however, would also not account for the fact that embedded subject proclitics are bound by subjects, a property of long distance anaphora, which is accounted for by the analysis presented in section 4.3.2.5, in which the anaphor covertly raises to the INFL projection at LF. Further research into structures similar to (111), as well as research into pronominals in Bora is necessary to explain the restrictions on antecedents for pronominals.

This Chapter has provided a formal theoretical analysis of Bora preverbal subject proclitics. After comparing Bora with other languages with seemingly similar phenomena, I established that, while Bora shared some properties in common with them, the phenomenon being described for Bora was ultimately unique in some respects. I considered potential analyses of Bora embedded clauses with proclitic subjects as control clauses, but ultimately established that Bora embedded clauses did not constitute anything that could be called obligatory control. I also established that the proclitics did not fit into an analysis which classifies them as agreement markers. Ultimately, I provided arguments that the 1$^{st}$ and 2$^{nd}$ person singular pronouns are nothing more than pronominals, with their referents determined by discourse factors.

Meanwhile, I analyzed both the 3COR and SAP markers as instances of anaphora. The 3COR marker always displays properties of long distance anaphora, since it cannot be bound within its own clause. The SAP marker, however, can either be bound in its own clause, in cases of main clauses and disjoint reference embedded clause subjects, or bound at a distance in cases of cross-clausal coreference. In the end, I analyze both the 3COR and SAP markers in a way similar to the analysis provided for Mandarin by Huang & Tang (1991), with the anaphors that are not bound in their own clause covertly raising at LF to the INFL of the clause with the structurally next highest c-commanding subject. In that position, they can be bound by a subject that matches in person.

# CHAPTER V

## Conclusion

This dissertation has presented extensive research and analysis of subjects and coreference in Bora, with a special focus on subject clitics in the language, an aspect of Bora syntax which has important consequences for linguistic theories of control and binding. I have laid out different environments in Bora which involve preverbal subject proclitics that, in many cases, indicate co-reference across clauses. The unusual typological nature of these clitics is empirically interesting when considering the theories of binding and control I have discussed throughout.

In Chapter 1, I introduced the Bora language, focusing on ethnographic information about the language and the methodology used for data collection. In this Chapter, I also indicated the importance of both: (i) performing research on the Bora language, especially given its status as an endangered language, and (ii) using data from understudied languages to inform formal linguistic theory. This dissertation has accomplished both of these goals, and has laid out opportunities for future research in the same veins.

The use of firsthand fieldwork has benefitted this project greatly. The interaction between data collection and theoretical analysis is remarkably symbiotic. Considering current formal theories drives theoretical questions that can be empirically tested in the field, while fieldwork data provides empirical evidence that can directly provide evidence for or against different linguistic theories. This dissertation project has allowed for a unique union between these endeavors.

Chapter 2 of this dissertation established the base for a syntactic analysis of Bora. Firstly, it outlined several relevant aspects of previous literature on Bora, especially phonological rules, the orthographic system, and the tonal system. I then went on to establish the ways in which subjects are expressed in Bora, specifically as overt nouns or pronouns, classifiers, or proclitics. Focusing especially on proclitic instantiations of the subject, I showed the environments in which proclitic

subjects appear depending on the person and number of the subject. Additionally, before establishing what these clitics are in later chapters, I established what they are not. That is, I showed how coreferent subject proclitics are different from object reflexives and reciprocals and how subject proclitics differ from possessive markers. Finally, I established the basic syntax of Bora clause structure, presenting syntactic representations of Bora clauses, showing that all the clauses being investigated here are finite clauses, and illustrating that, barring the exceptions outlined in section 2.9, subjects are always overtly expressed in Bora clauses.

The foundational literature on formal syntax that was used for this analysis was laid out in Chapter 3. Specifically, I reviewed some prominent theories of binding and control. Regarding theories of binding, I focused primarily on canonical theories of binding, focusing primarily on binding of anaphors and pronominals (Binding Conditions A and B in Principles and Parameters). Within this discussion, I consider the important distinction between local and long distance anaphors, and the properties of each. For long distance anaphors, I discuss some cross-linguistic variation in how they manifest and properties that they share across languages. As an example of long distance anaphora, I review the analysis of *ziji* in Mandarin Chinese based on Huang & Tang (1991), and how these authors have analyzed *ziji*'s anaphoric properties by appealing to covert movement of the anaphor at LF, thus explaining the binding of *ziji* by certain 3rd person antecedents both locally and in higher clauses. Chapter 3 also looks at an alternative approach to canonical Binding Theory by Reinhart & Reuland (1993) that relies of reflexivity and reflexive-marking as a property of anaphors.

Apart from theories of anaphora, Chapter 3 also provides an overview of prominent theories of control. Firstly, I focused on a theory of control which involves the phonologically null element PRO, which is licensed by clauses that are not specified for both Agreement and Tense, a discussion based primarily on Landau (2004, and subsequent work). I also reviewed some work by authors who have proposed analyses with overt realizations of PRO, such as Madigan (2008) and Szabolsci (2009). Finally, I provided a description of theories of control which have dispensed with PRO in favor of a movement approach to control, especially O'Neill (1997) and Hornstein (1999), considering some of the reasons for the authors' departure from the PRO analysis.

Chapter 4 provides my analysis of Bora subject clitics based on the data from Chapter 2 and the theories presented in Chapter 3. I begin by presenting phenomena in other languages that may

appear similar to Bora on the surface, but I go on to point out the typologically unique nature of the Bora proclitics. I analyze the Bora data considering the theories presented for control clauses, especially those theories with overt instantiations of control clause subjects. As a result, I establish that the Bora subject proclitics do not constitute subjects of control clauses.

I then turn to analyze the Bora subject proclitics as anaphors, dividing the analysis based on the person of the subject, as I did in Chapter 2. While I propose that 1st and 2nd person singular subjects are not anaphors, being determined solely by discourse factors, I do propose that the Speech Act Participant and 3rd person coreferent markers are anaphors, subject to the Binding Conditions or to conditions on reflexivity, depending on the theories as described in Chapter 3. I also point out the difference between the two anaphors, in that the Speech Act Participant can be bound both locally and at a distance, whereas the 3rd person coreferent marker can only be long distance bound. I point out that these anaphors can only be bound by the structurally next highest subject, and conclude that the long distance binding in these cases can be analyzed as a series of instances of binding by the subject in the next highest clause. I show that this analysis can be reached by the same mechanism as was described for Mandarin *ziji* in Chapter 3.

Chapter 4 concludes with some opportunities for further research that are directly relevant for an analysis of Bora subject proclitics. These include investigating in more detail the few types of clauses in Bora that do not require an overt subject, as well as determining whether the 3rd person possessive marker should be analyzed as having the same anaphoric properties as the 3rd person coreferent marker.

Overall this project raises interesting new questions about both binding theory and control. Primarily, if Bora subject clitics are indeed anaphors as I have proposed, the nature of what it means to be an anaphor in subject position should receive further attention in the literature on binding and reflexivity. Additionally, for theories of control, the overt presence of an embedded syntactic subject indicating or agreeing with a possible controller provides new evidence to examine when considering different theories of control. Although I did not adopt a theory of control for Bora embedded complements generally, the topic remains relevant since the proclitics I have analyzed in this dissertation are the only means of establishing cross-clausal coreference in Bora, a process that involves control and raising in other languages.

Finally, this dissertation has highlighted the importance of considering understudied languages when it comes to developing formal theories of language and human linguistic knowledge. The subject proclitics that I have analyzed in this dissertation do not, to my knowledge, behave like other unrelated languages with respect to their distribution and coreference capabilities. In establishing the anaphoric and pronominal properties of these proclitics, I have not only expanded the existing literature on the Bora language, but also expanded the data we have available to propose and evaluate theories of human linguistic knowledge.

# References


Aschmann, Richard P. (1993). *Proto Witotoan* (Publications in Linguistics 114). Dallas: The Summer Institute of Linguistics and the University of Texas at Arlington.

Aikhenvald, Alexandra Y. (2012). *Languages of the Amazon*. Oxford University Press.

Baker, Mark, & Kramer, Ruth (2018). Doubled clitics are pronouns. *Natural Language & Linguistic Theory*, *36*(4), 1035-1088.

Barss, Andrew. (1986). *Chains and anaphoric dependence: On reconstruction and its implications* (Doctoral dissertation, Massachusetts Institute of Technology).

Berger, Marcus & Frank Seifart. Forthcoming. The Boran languages. In *Amazonian Languages. An International Handbook*, eds. Patience Epps and Lev Michael, Berlin: de Gruyter Mouton.

Bittner, Maria. (1994). Case, scope, and binding. In *Case, Scope, and Binding* (pp. 1-48). Springer, Dordrecht.

Boeckx, Cedric, Hornstein, Norbert, & Nunes, Jairo (2010). *Control as movement* (Vol. 126). Cambridge University Press.

Bouchard, Denis (1984). *On the Content of Empty Categories*. Dordrecht: Foris.

Bresnan, Joan. (1982). *The mental representation of grammatical relations* (Vol. 1). The MIT Press.

Bresnan, Joan, & Mchombo, Sam A. (1987). Topic, pronoun, and agreement in Chicheŵa. *Language*, 741-782.

Büring, Daniel. (2005). *Binding theory*. Cambridge University Press.

Chelliah, Shobhana L. (2001). The role of text collection and elicitation in linguistic fieldwork. *Linguistic fieldwork*, Newman, Paul & Ratliff, Martha eds. New York, Cambridge University Press. 152-165.

Chomsky, Noam. (1973). Conditions on transformations. *A festschrift for Morris Halle*.

Chomsky, Noam. (1981). Lectures on government and binding. Dordrecht: Foris. 1982. Some concepts and consequences of the theory of government and binding. *Linguistic inquiry monograph*, (6).

Chomsky, Noam. (1986a). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Chomsky, Noam. (1993). *Lectures on government and binding: The Pisa lectures* (No. 9). Walter de Gruyter.

Chomsky, Noam. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. (2008). On phases. *Current Studies in Linguistics Series*, *45*, 133.

Chomsky, Noam, & Lasnik, Howard (1993). The theory of principles and parameters. Reprinted in: The minimalist program (1995), ed. *N. Chomsky. MIT Press.[RE]*.

Clements, George N. (1975). *The logophoric pronoun in Ewe: Its role in discourse*.

Cole, Peter, Hermon, Gabriella, & Huang, C. T. James (2006). Long-Distance Binding in Asian Languages. *The Blackwell companion to syntax*, 21-84.

Cole, Peter, Hermon, Gabriella, & Sung, Li-May (1990). Principles and parameters of long-distance reflexives. *Linguistic Inquiry*, 1-22.

Cole, Peter, & Wang, Chengchi (1996). Antecedents and blockers of long-distance reflexives: The case of Chinese ziji. *Linguistic inquiry*, 357-390.

Déchaine, Rose-Marie and Martina Wiltschko (2017). A Formal Typology of Reflexives. *Studia Linguistica* vol. 10 (1-2), 60-106.

Deen, Kamil Ud (2006). Subject agreement in Nairobi Swahili. In *Selected proceedings of the 35th annual conference on African linguistics* (pp. 225-233).

Eberhard, David M., Gary F. Simons, and Charles D. Fennig (eds.). 2019. *Ethnologue: Languages of the World.* Twenty-second edition. Dallas, TX: SIL International. Online version: http://www.ethnologue.com.

Echeverri, Juan Álvaro and Frank Seifart. 2015. A re-evaluation of the Witotoan family and its genealogical relation to the Boran family (North West Amazon). Submitted to *International Journal of American Linguistics*.

Everaert, Martin (1986). *The Syntax of Reflexivization*. Dordrecht: Foris.

Fox, Daniel. (1993). Chain and binding – a modification of Reinhart and Reuland's 'Reflexivity'. *Ms., MIT, Cambridge, Mass*.

Giorgi, Alessandra. (1984). Toward a theory of long distance anaphors: A GB approach. *The Linguistic Review*, *3*(4), 307-362.

Haddad, Youssef A. (2009). Copy control in Telugu. Journal of Linguistics, 45(1), 69-109.

Hagège, Claude (1974). Les pronoms logophoriques. *Bulletin de la Société de Linguistique de Paris*, *69*(1), 287-310.

Hestvik, Arild (1991). Subjectless binding domains. *Natural Language & Linguistic Theory*, *9*(3), 455-496.

Holmberg, Anders, & Sheehan, Michelle (2010). Control into finite clauses in partial null-subject languages. *Parametric variation: Null subjects in minimalist theory*, 125-152.

Hornstein, Norbert (1999). Movement and Control. *Linguistic Inquiry* 30: 69-96.

Hornstein, Norbert, Nunes, Jairo, & Grohmann, Kleanthes K. (2005). *Understanding minimalism*. Cambridge University Press.

Hornstein, Norbert, & Polinsky, Maria (2010). Control as movement. *Movement theory of control. Amsterdam: John Benjamins*, 1-41.

Huang, C. T. James., & Tang, C. C. Jane. (1991). 13 The local nature of the long-distance reflexive in Chinese. *Long distance anaphora*, 263.

Huang, Yan. (2000). *Anaphora: A cross-linguistic approach*. Oxford University Press on Demand.

Kapetangianni, Konstantia. (2010). The Minimalist Syntax of Control in Greek. Doctoral Dissertation, University of Michigan.

Landau, Idan. (2003). Movement out of control. *Linguistic inquiry*, *34*(3), 471-498.

Landau, Idan. (2004). The scale of finiteness and the calculus of control. *Natural Language & Linguistic Theory*, *22*(4), 811-877.

Landau, Idan. (2006). Severing the distribution of PRO from case. *Syntax*, *9*(2), 153-170.

Landau, Idan. (2013). *Control in generative grammar: A research companion*. Cambridge University Press.

Landau, Idan (2015). *A Two-Tiered Theory of Control*. Cambridge, MA: MIT Press.

Lebeaux, David (1983). A distributional difference between reciprocals and reflexives. *Linguistic inquiry*, 723-730.

Madigan, Sean William (2008). *Control constructions in Korean*. University of Delaware.

Martin, Roger. (2001). Null case and the distribution of PRO. *Linguistic inquiry*, *32*(1), 141-166.

Martius, Carl Friedrich Philipp von. 1867. *Beiträge zur Ethnographie und Sprachenkunde Amerika's, zumal Brasiliens. Vol. II: Zur Sprachenkunde. Wörtersammlung Brasilianischer Sprachen. Glossaria linguarum Brasiliensium. Glossarios de diversas lingoas e dialectas, que fallao os Indios no imperio de Brazil*. Leipzig: Friedrich Fleischer.

Mayer, Mercer. (1969). *Frog, where are you?*. New York: Dial Press.

Mayer, Mercer. (1978). *A boy, a dog and a frog*. Turtleback.

Mayer, Marianna, & Mercer Mayer. (1992). *A boy, a dog, a frog and a friend*. Houghton Mifflin.

Nediger, Will (2017). Overcoming Empirical Challenges for an Extended Approach to Condition C. *Linguistic Inquiry*. 48 (1): 99-127.

O'Neill, John (1997). *Means of Control: Deriving the Properties of PRO in the Minimalist Program*. PhD Dissertation, Harvard University.

Ostrove, Jason (2017). Severing PRO from its Silence: Control Constructions in San Martín Peras Mixtec. *Handout from the 22nd Workshop on the Structure and Constituency of the Languages of the Americas (WSCLA), University of British Columbia, Vancouver.*

Ostrove, Jason. (2018). *When phi-Agreement Targets Topics: The view from San Martín Peras Mixtec* (Doctoral dissertation, UC Santa Cruz).

Pearson, Hazel (2015). The interpretation of the logophoric pronoun in Ewe. *Natural Language Semantics* 23 (2): 77-118.

Pires, Acrisio (2001). The syntax of gerunds and infinitives: Subjects, case and control. PhD dissertation. University of Maryland, College Park.

Pires, Acrisio (2006). *The minimalist syntax of defective domains: Gerunds and infinitives* (Vol. 98). John Benjamins Publishing.

Polinsky, Maria, & Potsdam, Eric (2002). Backward control. *Linguistic Inquiry*, *33*(2), 245-282.

Polinsky, Maria, & Potsdam, Eric (2006). Expanding the scope of control and raising. Syntax, 9(2), 171-192.

Reinhart, Tanya, & Reuland, Eric (1991). 14 Anaphors and logophors: an argument structure perspective. *Long distance anaphora*, 283.

Reinhart, Tanya and Eric Reuland (1993). Reflexivity. *Linguistic Inquiry* 24: 657-720.

Reuland, Eric (2001). Primitives of Binding. *Linguistic Inquiry* 32.2: 439-492.

Reuland, Eric (2006a). Binding Theory: Terms and Concepts. *In The Blackwell Companion to Syntax* (*vol. 1*), M. Everaert and H. Van Riemsdijk eds. Malden, MA, Blackwell, 260-283.

Reuland, Eric (2006b). Logophoricity. *In The Blackwell Companion to Syntax* (*vol. 3*), M. Everaert and H. Van Riemsdijk eds. Malden, MA, Blackwell, 1-20.

Reuland, Eric (2006c). Long-Distance Binding in the Germanic Languages. *In The Blackwell Companion to Syntax* (*vol. 3*), M. Everaert and H. Van Riemsdijk eds. Malden, MA, Blackwell, 85-108.

Reuland, Eric. (2006d). Icelandic logophoric anaphora. *The Blackwell companion to syntax (vol. 2),* M. Everaert and H. Van Riemsdijk eds. Malden, MA, Blackwell, 544-557.

Reuland, Eric (2011). *Anaphora and Language Design*. Cambridge, MA: MIT Press.

Roe, Amy. 2014. *The Phonetics and Phonology of Bora Tone*. Grand Forks: University of North Dakota MA Thesis.

Ross, John Robert (1970). On declarative sentences. *Readings in English transformational grammar*, *222*, 272.

Safir, Ken (2014). One True Anaphor. *Linguistic Inquiry* 45 (1): 91-124.

Seifart, Frank. 2005. *The structure and use of shape-based noun classes in Miraña (North West Amazon)*. Radboud Universiteit Nijmegen Ph.D. dissertation.

Seifart, Frank. 2010. The Bora connector pronoun and tail-head linkage: a study in language-specific grammaticalization. *Linguistics* 48: 893-918.

Seifart, Frank. 2015a. Syntactic Sketch: Bora. In Tibor Kiss and Artemis Alexiadou (eds.), *Syntax: Theory and Analysis: An International Handbook. Vol. 3* (Handbook of Linguistics and Communication Science (HSK), 42/3), 1764-1791. Berlin: de Gruyter Mouton.

Seifart, Frank. (2015b). Valency classes in Bora. In *Valency Classes in the World's Languages, Volume 2: Case Studies from Austronesia, the Pacific, the Americas, and Theoretical Outlook*, eds. Andrej Malchukov & Bernard Comrie. 1481–1514. (Comparative Handbooks of Linguistics 1.2). Berlin, München, Boston: de Gruyter.

Seifart, Frank and Juan Álvaro Echeverri. 2015. Proto Bora-Muinane. *Linguas Indígenas Americanas* 15 (2). 279-311.

Seifart, Frank, Meyer, Julien, Grawunder, Sven, & Dentel, Laure (2018). Reducing language to rhythm: Amazonian Bora drummed language exploits speech rhythm for long-distance communication. *Royal Society open science*, *5*(4), 170354.

Sells, Peter. (1987). Aspects of logophoricity. *Linguistic inquiry*, *18*(3), 445-479.

Szabolcsi, Anna. (2009). Overt nominative subjects in infinitival complements in Hungarian. In *Approaches to Hungarian: Volume 11: Papers from the 2007 New York Conference* (Vol. 11, p. 251). John Benjamins Publishing.

Terzi, Arhonto. (1993). PRO in finite clauses: A study of the inflectional heads of the Balkan languages.

Terzi, Arhonto. (1997). PRO and null case in finite clauses. *The Linguistic Review*, *14*(4), 335-360.

Thiesen, Wesley. 1996. *Gramática del Idioma Bora.* (Serie Lingüística Peruana 38.) Pucallpa: Ministerio de Educación and Instituto Lingüístico de Verano.

Thiesen, Wesley, and Eva Thiesen. (1998). *Diccionario Bora - Castellano, Castellano - Bora*. (Serie Lingüística Peruana 46.) Pucallpa: Ministerio de Educación and Instituto Lingüístico de Verano.

Thiesen, Wesley, & Weber, David (2012). *A grammar of Bora: With special attention to tone*. SIL International.

Thráinsson, Höskuldur (1991). Long-distance reflexives and the typology of NPs. *Long-distance anaphora*, 49-75.

Vengoechea, Consuelo. (2015). Categorías léxicas y clasificación nominal en muinane. Noroeste Amazónico. *Revista Brasileira de Linguistica Aplicada* 6, 1: 63-86.

Walton, James W., Janice P. Walton, and Clementina Pakky de Buenaventura. 1997. *Diccionario bilingüe muinane-español, español-muinane*. Santafé de Bogotá: Editorial Alberto Lleras Camargo.

Zribi-Hertz, Anne. (1989). A-type binding and narrative point of view. *Language*, *65*, 695-727.