UNIVERSITY OF CALIFORNIA
SANTA CRUZ

WHEN φ-AGREEMENT TARGETS TOPICS: THE VIEW FROM SAN MARTÍN PERAS MIXTEC

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

LINGUISTICS

by

Jason Ostrove

June 2018

The Dissertation of Jason Ostrove
is approved:

Professor James McCloskey, Chair

Professor Emerita Sandra Chung

Professor Ryan Bennett

Professor Ruth Kramer

Dean Tyrus Miller
Dean of Graduate Studies

Table of Contents

Abstract

When φ-agreement targets topics: The view from San Martín Peras Mixtec

by

Jason Ostrove

This dissertation presents what is, to my knowledge, the first in-depth theoretical syntactic investigation of a Mixtec language: San Martín Peras Mixtec. The syntax of this language is explored through the lens of a process called pronoun doubling, which duplicates the φ-features of a nominal with a preverbal morpheme that is morphophonologically identical to a regular pronoun.

(1) a. **Rà**$_i$ xá'antsya rà Juan$_i$ chìkí.
he cut.PRES he J. tuna
'Juan is cutting *tunas* (the fruit of the prickly pear cactus).'

b. **Rí**$_i$ xá'antsya rà Juan chìkí$_i$.
it.AML cut.PRES he J. tuna
'Juan is cutting *tunas*.'

The narrow question of this dissertation is what the synchronic analysis of pronoun doubling is. Ultimately, I argue that, despite initial appearances, the "pronoun" in pronoun doubling constructions is not a true pronominal, but rather the morphophonological exponent of a φ-agreement process that targets topics.

But as this process replicates φ-features, I consider two possibilities. First, I investigate if pronoun doubling could be a form of clitic doubling. I argue that clitic doubling is best derived by phrasal movement of the doubled nominal, followed by reduction of the higher copy to a clitic. Therefore, in order to diagnose clitic doubling within this "Move-and-Reduce" framework, we must look for evidence of movement in pronoun doubling constructions.

I first consider A-movement. I describe a novel A-movement process which I term "quantifier fronting." I then show that none of the A-movement properties of quantifier fronting can be found in pronoun doubling. I also show that A-movement in SMP Mixtec is typologically unusual in being much less restricted than A'-movement.

A'-movement in SMP Mixtec shows several interesting properties. First, it demonstrates an A'-extraction restriction reminiscent of syntactic ergativity: external arguments may not undergo A'-movement. I investigate this restriction through the properties of two A'-constructions. I first investigate restrictive relative clauses, which require a resumption strategy in order to avoid the ban against extraction of external arguments. I then investigate content wh-questions, which avoid this ban by employing a movement construction that demonstrably lacks the properties of both A-movement and A'-movement. Crucially, no evidence for this extraction restriction is replicated in pronoun doubling constructions.

As clitic doubling within this Move-and-Reduce framework does not seem like a viable option, I consider the possibility that pronoun doubling is a form of φ-agreement. Despite appearances, this turns out to be the best fit for several reasons. First, pronoun doubling obeys the PIC yet allows Closest Conjunct Agreement. Second, pronoun doubling fails to affect variable or anaphora binding. That said, pronoun doubling is not vanilla φ-agreement for two reasons. First, as shown in (1), both subjects and objects are equally accessible. Second, the choice between the two sentences in (1) depends on which argument is the topic.

(2) a. Nă xá'antsya rà Juan?
what cut.PRES he J.
'What is Juan cutting.'

b. **Rà**$_i$ xá'antsya [rà Juan$_i$]$_{Top}$ [chìkí]$_{Foc}$.
he cut.PRES he J. tuna
'Juan is cutting *tunas*.' *Answer to (2a).*

c. #? **Rí**$_j$ xá'antsya [rà Juan$_i$]$_{Top}$ [chìkí$_j$]$_{Foc}$.
it.AML cut.PRES he J. tuna
Intended as an answer to (2a).

(3) a. Yó nà xá'antsya chìkí?
who they cut.PRES tuna
'Who is cutting the tunas?'

b. **Rí**$_j$ xá'antsya [rà Juan]$_{Foc}$ [rí chìkí$_j$]$_{Top}$.
it.AML cut.PRES he J. it.AML tuna
'Juan is cutting *tunas*.' *Answer to (3a).*

c. #? **Rà**$_i$ xá'antsya [rà Juan$_i$]$_{Foc}$ [rí chìkí]$_{Top}$.
he cut.PRES he J. it.AML tuna
Intended as an answer to (3a).

To account for this system, I develop a theory in which φ-agreement can target topics. This functions similarly to a traditional minimalist Case theory, in which nouns must stand in an AGREE relation with some functional head in order to receive Case. I adapt this general approach to topic agreement by proposing that nominals which are to be interpreted as topics must stand in an AGREE relation witha head Top, and Top also acts as a φ-probe.

The dissertation is organized as follows. Chapter 1 provides the necessary background on this language, including a brief syntactic and phonological overview. Chapter 1 also provides a detailed discussion of the prosodic conditions which restrict the appearance of topic agreement. Chapter 2 provides an extensive literature review of both φ-agreement and clitic doubling, including an exhaustive evaluation of the diagnostics which have been put forward to distinguish the two. In this chapter I also argue for why clitic doubling is best understood within a Move-and-Reduce framework. This approach to clitic doubling allows for intuitive argumentation for either clitic doubling or φ-agreement: clitic doubling constructions show the characteristics of movement chains, while φ-agreement does not.

With this background, Chapter 3 begins the investigation of movement constructions in San Martín Peras Mixtec by investigating A-movement. It begins with a detailed discussion of the syntax of embedded clauses. This is because failure to extract from finite embedded clauses is a common feature of A-movement cross-linguistically. In this language, though, it is difficult to tell which embedded clauses are finite because there is no non-finite verbal morphology and all embedded clauses have an overt subject. I argue that, despite appearances, this language distinguishes between Obligatory Control and simple pronominal coreference, despite this language's equivalent of PRO always being overt. This relation is mediated through a CP layer, which faithfully tracks when Oblogatory Control occurs.

Once these aspects of the clause structure have been established, I demonstrate the A-movement properties of quantifier fronting, such as failure to reconstruct for scope or binding, and the ability to expand the variable and anaphora binding possibilities within a clause. I also show that A-movement in this language is much less restricted than A'-movement, despite being clause bounded. While A'-movement forbids extraction of possessors, objects of prepositions, and, as shown in Chapter 4, external arguments, A-movement shows no such restriction. Additionally, A'-movement, particular content wh-questions, require pied-piping with inversion, while A-movement is incompatible with it. Importantly for our purposes, pronoun doubling fails to show any of these A-chain properties: it fails to affect variable or anaphora binding, and it may not target objects of prepositions or possessors.

As pronoun doubling fails to show the properties of an A-chain, Chapter 4 considers the possibility that the movement chain involves is an A'-chain. This requires an extensive investigation of A'-movement in this language. I demonstrate that San Martín Peras Mixtec shows a strict A'-extraction restriction, in addition to the restictions shown in Chapter 3: external arguments may not undergo A'-extraction. This restriction, while similar in many ways, is fundamentally different from syntactic erga-

tivity for two reasons. First, San Martín Peras Mixtec is not morphologically ergative. Second, intransitive predicates show varying behavior, with unergative subjects patterns with transitive subjects while unaccusative subjects pattern with transitive objects.

After this extraction restiction is established, I demonstrate its properties in more detail. First I investigate restrictive relative clauses, which require resumption of external arguments. I show that this resumption strategy fails to show any properties of movement. This allows the language to straightforwardly avoid the A'-extraction restriction. Second, I investigate the syntax of content wh-questions. These also show the effects of the A'-extraction restriction, albeit the repair is different. Content wh-questions of external arguments do involve movement, but I demonstrate that this movement process fails to show the hallmark characteristics of either A-movement or A'-movement. This also allows the language to satisfy the restriction against A'-extraction of external arguments. I conclude by demonstrating that this extraction restriction that is so crucial to understanding A'-constructions in this language is not replicable in pronoun doubling constructions because external arguments may be freely targeted, as shown in (1a).

As pronoun doubling fails to show the diagnostic movement chain properties of clitic doubling, Chapter 5 considers the possibility that this process is a φ-agreement. After cataloging the φ-agreement properties of pronoun doubling, such as observance of the PIC, Closest Conjunct Agreement, and the failure to affect variable and anaphora binding, I develop a theory of topic agreement. First, I demonstrate that topic agreement is not an especially rare phenomenon cross-linguistically, having been reported in a range of languages from Northern Khanty (Uralic, Russia), Oromo (Cushitic, Ethiopia), and many others. While the semantic investigation is cursory, I demonstrate that topichood is the best notion to determine which nominals will be targeted in pronoun doubling constructions. Finally, I present the formal analysis of topic agreement, which I described above.

In memory of my grandfather, Istvan Jarai. I miss you every day, and wish you could have made it just a few more to see me finish this dissertation.

Acknowledgments

Primero, agradezco profundamente a las fuertes e inteligentes mujeres Oaxaquñeas, Rocelia Cruz, Roselia Duran y Natalia Gracida Cruz, por pasar tanto tiempo conmigo enseñándome su idioma. Tengo tanto respeto por ustedes, y espero que ustedes estén orgullosas de lo que hemos logrado. No lo podría haber hecho sin su apoyo, su ánimo, o su fe en mi. Lo único que lamento es no haber podido escribir esta tesis en Español para que ustedes también puedan disfrutarla.

A Rocelia, le agradezco su paciencia, fuerza y transparencia con la que me explicó el Mixteco. A Roselia, le agradezco su humor, su gracia, su fe y su calidez. Y finalmente, pero no menos importante, a Natalia le agradezco por su aguda inteligencia y por su amplia sabiduría sobre el Mixteco, su generosidad y su bondad. Les agradezco a todas por su amistad. Ha sido un gran honor conocerles y trabajar con ustedes. Cada una de ustedes me inspira y espero seguir trabajando con ustedes en el futuro. Taxa'abi ntskû ndó.

Second, I thank those who helped and supported this work on the theoretical end. Chief among them is Jim McCloskey. Not only was Jim the chair of this dissertation, but he was also my advisor basically from my first day at UCSC on. It has been a true joy working with you. Thank you for your attention, your dedication, your trust in me. Thank you also for your red pen, and the hours your spent taking it to my work. I appreciate each second. I could not, nor would not, have done this work under any other advisor, and I thank you sincerely for that. It has been a tremendous honor and privilege to be your student.

I also thank Sandy Chung, who served on the committee for this project and as the chair for my qualifying exam. This dissertation benefited in many ways from your comments, your support, and the connections you made that I did not see. I am proud and am honored to consider myself one of your students.

Ryan Bennett, I thank you for so much as well. As the committee member with the most knowledge of Oto-Manguean, your support and perspective were invaluable. In addition, although I don't remember discussing it explicitly with you afterwards, the seeds of this dissertation were sown in a conversation we had about this material a year before the committee or the project were even established. You encouraged me to look beyond the prosody and investigate the morphosyntax of this process, as this was where the real mysteries were. I thank you for this, as well as for your creativity, your theoretical flexibility, and your support.

Last of the committee members, but certainly not least, I thank you, Ruth. Your fingerprints are all over this dissertation, and your thinking, particularly your attention to detail, have been deeply influential to this project. I thank you for your enthusiasm for this project - it has been invaluable.

This dissertation has been profoundly affected by many other people at UC Santa Cruz in addition to the committee. First, I thank Judith Aissen. I consider you just as influential as the official committee. Your support for every aspect of this project since my second qualifying paper has been tremendously important to me. Indeed, when I wavered in the analysis, your clear perception and encouragement set me back on course. I can say, without a shadow of a doubt, that this analysis of topic agreement would never have happened without your guidance, and for this I am immensely grateful.

I also thank Junko Itô and Armin Mester. You supported and guided, along with Judith, my first attempts at understanding this process in my second qualifying paper. This dissertation would certainly not exist without your early guidance and encouragement. I thank you tremendously for this.

Thank you as well to Maziar Toosarvandani and Pranav Anand. Much of the fieldwork necessary for this dissertation was funded, financially, intellectually, and spiritually, by the Workshop on the Languages of Mesoamerica (WLMA). I thank you for providing me the space to be a part of it, and for your warm support at every stage

of this project.

I thank and acknowledge Ivy Sichel as well, as the content on Obligatory Control in Chapter 3 was done under your guidance. I thank you for your careful discussions, your thoughtful guidance, and your interest in this work.

My thanks as well to Donka Farkas. Thank you for coming in at such a late stage in the project with some critical advice on the semantics. I regret not having come to you sooner. I thank you for encouraging me to act like a good tourist to the semantics: I hope I didn't break or damage anything, and I hope I gave what I saw the respect it deserves.

I also thank Erik Zyman. Your enthusiastic, convivial approach to syntax are the beating heart of UCSC syntax. I thank you for your time, your joy, your humor, and your friendship.

Last of the UCSC community, but certainly and emphatically not least, I thank the department staff, Ashley Hardisty, Matt MacLeod, Maria Zimmer, and Logan Roberts. In a much less theoretical way, this dissertation could never have happened without you. I thank you for your time, your help and support, and your friendship. I feel sorry for people who don't work in the same community as your four.

In addition to people in the immediate UCSC community, this project benefited in myriad ways from many conversations with different people. I thank and acknowledge, in no particular order, Christian DiCanio, for raising my perspective to the cross-Mixtec level, Gillian Ramchand, for making sure I crawled before I ran, Peter Svenonius, for making sure I emphasized the positive evidence in addition to the negative, and Martina Wiltschko, for encouraging me to think outside the box.

Finally, as a human, I would not have finished this without my friends, my family, my boyfriend, or my therapist. Thank you especially to Julio, for keeping me loved and sane, my parents, David and Lilly, and my siblings, Danielle and Ryan, and my grandmother, Marie, for making sure I never forgot what was really important.

List of abbreviations

ABS	Absolutive
ACC	Accusative
ADJ	Adjectival predicate
ADV	Adverbial
AGR	Agreement
AML	Animal gender
ANTIPASS	Antipassive
AUG	Augment vowel
C	Complementizer
CL	Clitic
CLASS	Classifier
CS	Causative
COMP	Completive aspect
COND	Conditional
CONT	Continuous aspect
COP	Copula

DAT	Dative
DECL	Declarative
DEF	Definite
DET	Determiner
DIM	Diminutive
ERG	Ergative
EMPH	Emphatic
ENC	Enclitic
EPEN	Epenthetic vowel
EXCL	Exclusive
F(EM)	Feminine
FOC	Focus
FUT	Future
FV	Final vowel
GEN	Genitive
HAB	Habitual
IMPERF	Imperfective aspect
INCL	Inclusive
IND	Indicative
INF	Infinitive
INT	Interrogative

INTR	Intransitive marker
INST	Instrumental
INTS	Intensifier
IRR(EAL)	Irrealis
LIQ	Liquid gender
LOC	Locative
M(ASC)	Masculine
NEG	Negative
NOM	Nominal predicate or nominative
O(BJ)	Object
PERF	Perfective aspect marker
PG	Parasitic gap
PL	Plural
POSS	Possessive
POT	Potential aspect
PR(E)S	Present tense
PROG	Progressive
Q	Polar question
RECIP	Reciprocal
REL	Relative clause
REM.PAST	Remote past

SA	Subject agreement
SG	Singular
SJV	Subjunctive
SQ	Sequential
SUBJ	Subject or subjunctive
TH	Thematic suffix
TRANS	Transitive
WOOD	Wooden gender
⊠	Absence of relevant morphology

Chapter 1

Introduction

When one encounters a process in a language in which the φ-features of a nominal are copied, there are two main hypotheses one entertains as possible derivations.

(1) a. φ-agreement

b. Clitic pronoun doubling

Distinguishing between these two possibilities is easier said than done. As Baker and Kramer (2018) say, "It is an awkward fact that generative linguistics has had a hard time distinguishing reliably between pure agreement and clitic doubling."

Broadly, two sorts of diagnostics are employed to tease clitic doubling and φ-agreement apart. The first of these are morphophonological diagnostics, such as "Does the morpheme in question look like a doubled clitic?" The second are syntactic/interpretive diagnostics, like "Does the morpheme in question demonstrate the hallmark interpretive and syntactic properties of a pronoun?" Of course, when one approaches these problems when considering actual data, the results are rarely clean because the two sets of diagnostics frequently point to opposite conclusions. This leaves the analyst in the uncomfortable theoretical position of making a choice that is not completely grounded in the data.

In this dissertation, I consider such a process in San Martín Peras Mixtec, an Oto-Manguean language of Oaxaca, Mexico. I refer to this pattern as 'pronoun doubling,' as it involves a morpheme, morphophonologically identical to a regular pronoun, in clause-initial position that cross-references the φ-features of an argument within the clause.

(2) a. **Rà**$_i$ íxutsya míí mástro$_i$.
he swim.PRES the teacher
'The teacher is swimming.'

b. **Rà**$_i$ íxutsya rà$_i$.
he swim.PRES he
'He's swimming.'

c. **Rí**$_i$ xa nùhmi̯ rà lo'o míí tsi̯na$_i$.
it.AML PERF hug.PAST he little the dpg
'The boy already hugged the dog.'

d. **Rí**$_i$ xa nùhmi̯ rà rí$_i$.
it.AML PERF hug.PAST he it.AML
'He already hugged it (an animal).'

In each of the sentences in (2), a morpheme that looks like a regular pronoun in initial position duplicates the φ-features of one of the arguments in the sentence. This can be either the subject, as in (2a-b), or the object, as in (2c-d).

Pronoun doubling, as the name implies, passes the morphophonological diagnostics for clitic pronoun doubling because the morphemes in question, namely those bolded in (2), are identical in form to regular pronouns. This is made particularly clear in (2b) and (2d), in which we see that the doubling morphemes are identical to the pronominal argues they cross reference.

Despite this morphophonological identity, the primary goal of this dissertation is to demonstrate that the bolded morphemes in (2) are not true pronouns. In other words, these morphemes systematically fail to demonstrate the key interpretative and

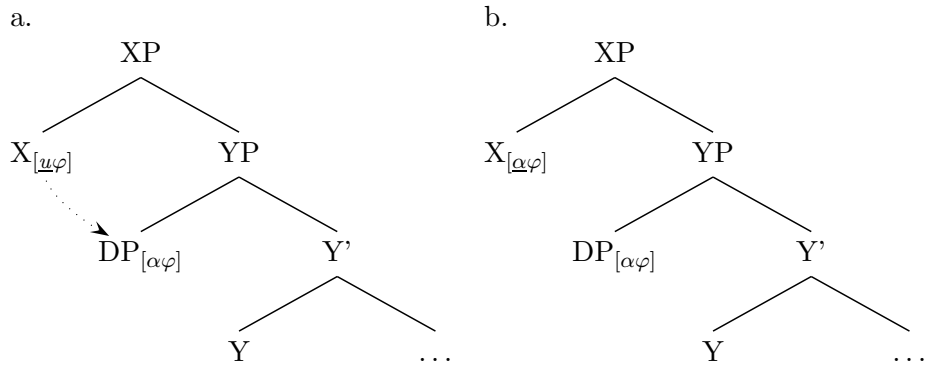
syntactic properties of elements of the category D, both internal to this language and cross-linguistically. Rather, I contend that pronoun doubling is instead a form of *topic agreement*, defined in (3).

(3) **Topic agreement**: A φ-agreement system that targets sentence topics.

In this dissertation, topic agreement refers to a φ-agreement system in which the targets for agreement are sentence topics, rather than more familiar cases of "subject" or "object" agreement. This sensitivity to information structural categories, rather than syntactic notions like subject or object, allows this agreement process to target both subjects and objects, as we saw in (2).

The argument for topic agreement is presented in three ways. First, in Ch. 2, I present a comprehensive theoretical overview of both φ-agreement and clitic doubling systems. For φ-agreement, I adopt the AGREE formalism of Chomsky (2000, 2001), presented briefly in (4).

(4)

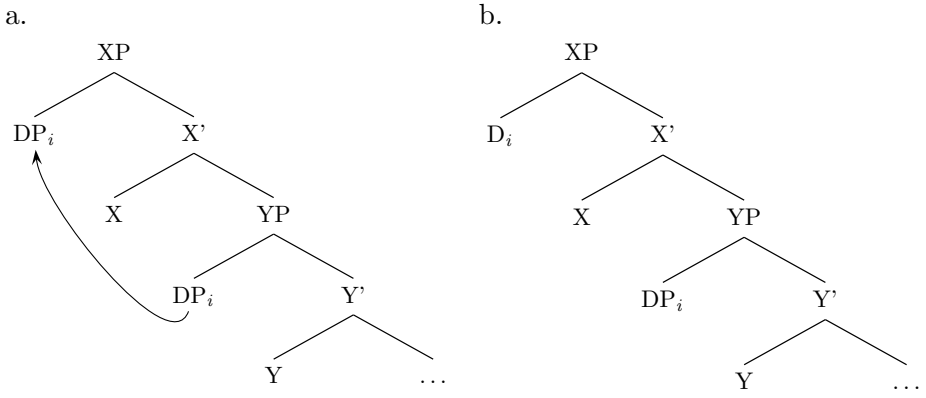


In (4a), we see a probe X that lacks valued φ-features. I notate this as $[\underline{u}\varphi]$. X searches its c-command domain for a set of φ-features, which it copies to provide a value for its unvalued feature.

For clitic doubling, I adopt a formalism which I term "Move-and-Reduce" (see

Alexiadou and Anagnostopoulou 1997, Suñer 2000, Anagnostopoulou 2007, Preminger 2009, Harizanov 2014b,a, and Kramer 2014).[1] This is shown briefly in (5).

(5)



In a Move-and-Reduce system of clitic doubling, XP movement yields two syntactic occurrences, or copies (Chomsky 1995) of a DP. The higher copy of DP is reduced to a simple D^{o}, while the lower copy receives full Spell-Out. This higher, reduced copy of DP is the doubled clitic pronoun that is the hallmark of clitic doubling constructions.

In adopting this syntax, we are provided with clear methods to distinguish between φ-agreement and clitic doubling. If a process is derived through clitic doubling, we should be able to diagnose the presence of the two core properties of the system in (5): XP movement, and the higher D^{o}. In contrast, if a process is derived by φ-agreement, we should not see either of these. Rather, the process should demonstrate the key signatures of Chomskyian Agree, discussed in Ch. 2.

In Chapters 3 through 5, I show that pronoun doubling in SMP Mixtec systematically fails to demonstrate any of the expected properties of clitic doubling. In Chapter 3, I demonstrate that no A-movement can be diagnosed, and likewise that doubled pronouns do not show the interpretive properties of true pronouns. In Chapter 4,

[1] I adopted this terminology from an earlier, unpublished draft of Baker and Kramer (2018). While this analysis did not appear in the published version, it was still influential on my thinking.

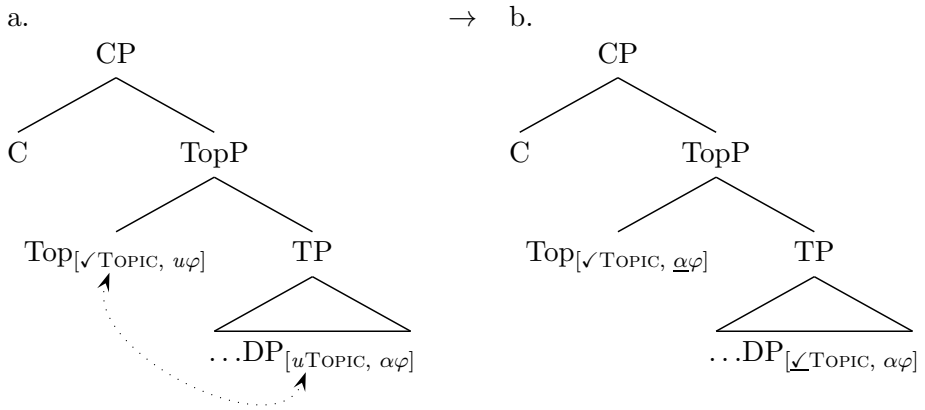
I show that, likewise, A'-movement cannot provide the requisite movement for a clitic doubling analysis of pronoun doubling. This is because, as demonstrated exhaustively in Chapter 4, SMP Mixtec has an A'-extraction restriction similar to syntactic ergativity (Dixon 1979, Deal 2016, Polinsky 2016). This extraction restriction is defined in (6).

(6) A'-extraction restriction in SMP Mixtec: External arguments may not undergo A'-extraction.

In Chapter 4, I demonstrate that this restriction cannot be replicated in pronoun doubling constructions. This means that A'-movement cannot be used to derive pronoun doubling.

In Chapter 5, I develop a formal theory of pronoun doubling as topic agreement, as defined in (3). I begin by showing that pronoun doubling exhibits a set of characteristics that are much more compatible with φ-agreement than clitic doubling. Second, I show that φ-agreement in this language indeed targets sentence topics. Once we have established that pronoun doubling is topic agreement, I develop and motivate the formalism in (7).

(7)



Formally, the topic agreement system in (7) functions identically to a traditional minimalist Case system of the sort developed in Chomsky and Lasnik (1993) and Chomsky

(1995). The only formal difference is that the features shared in (7) between the probe and the goal besides φ-features are the information structural feature [✓Topic], rather than Case features. This is precisely the sort of system predicted by Miyagawa (2010, 2017): discourse configurational/information structural features, or δ-features to use Miyagawa's terminology, serve a function comparable to φ-features or Case feature. Indeed, in this language we see a mixed probe in the sense of Coon and Bale (2014) that combines aspects of the traditional A-system, ϕ-features, with traditional A'-features, or δ-features. This supports the general view put forward by Miyagawa (2010, 2017) and van Urk (2015) the A/A'-system is an emergent property of the features implicated in particular Agree relations.

Before diving into the particulars of the analysis, section 1 of this chapter introduces San Martín Peras Mixtec. This is necessary because, to my knowledge, this language has never been reported before in any literature, apart from relatively passing mentions of its existence. I include a basic phonological sketch in §1.1, as well as a cursory syntactic description in §1.2 give readers unfamiliar with Mixtec languages a grounding. In §2 of this chapter, I discuss several prosodic confounds that prevent the appearance of topic agreement in otherwise licit syntacticosemantic contexts.

1.1 Introducing San Martín Peras Mixtec

San Martín Peras Mixtec belongs to the Mixtec branch of Oto-Manguean. Oto-Manguean is one of the core language families of Mesoamerica, spoken from the central Mexican state of San Luis Potosí south to Oaxaca, although the family once extended as far south as Costa Rica. (Campbell et al. 1986). Oto-Manguean speakers have occupied their current homelands since around 2000 BCE (Kaufman and Justeson 2009), and their ancenstors were likely one of the core groups which gave rise to civilization in Mesoamerica. Although they have been in contact with other groups for millennia,

they have formed a coherent and culturally distinct sub-group within Mesoamerica since their appearance in the archaeological record. Early Oto-Manguean speakers are strongly associated with the domestication of maize and other staple crops, as well as major archaeological sites such as Monte Albán. See the papers in Josserand et al. (1984). Well over a million people still speak Oto-Manguean languages today.

According to the 2005 census, Mixtec is the thirdmost spoken indigenous language of Mexico with just shy of 500,000 speakers, after only Nahuatl and Yucatec Mayan. That said, there is considerable and complex diversity within the Mixtec family, with very little consensus as to how many distinct languages there are. The Summer Institute of Linguistics has assigned unique SIL codes to 33 Mixtec languages, though mutual intelligibility surveys put the number closer to 27 (Egland and Bartholomew 1983). Either way, many observers have commented on the rich and complicated variation within the Mixtec languages (see especially Suárez 1983), and others have argued that the Mixtec languages have the same historical diversity and time-depth as Romance (Hock and Joseph 1996). Given this pervasive diversity, there is little consensus about the higher level groupings within the Mixtec languages. SMP Mixtec speakers identify their language as a divergent branch within the informal Mixteca Baja grouping, which is supported by historical and mutual intelligibility studies (see Josserand 1983 and Egland and Bartholomew (1983)).

San Martín Peras Mixtec (natively *tù'un ntá'bi* 'language of the poor') is spoken in the town of San Martín Peras (SMP Mixtec *Ñû Kă'anu* 'the big town') and in the surrounding municipality to the north as far as Ahuajutía. San Martín Peras is located within the district of Juxtlahuaca along the far western boarder of the state of Oaxaca, about 5 miles from the neighboring state of Guerrero. Population estimates for the region generally refer to the municipality as a whole. According to the 2005 national census, about 12,000 lived in the region.

San Martín Peras is one of the poorest regions of Oaxaca. The road to San

Martín Peras was only completed in the mid-2000's, and many people continue to use mules for transportation. Small-scale subsistence agriculture is the primary economic activity in the region, along with some small mining operations. Given the poor economic conditions in the area, many people began migrating to the United States in the early to mid-2000s. The majority settled in California, particularly in Oxnard and Watsonville. As San Martín Peras has been historically isolated from the rest of Mexico by poverty, many of these new Americans come speaking only Mixtec. Most learn Spanish and some English once they are in the U.S. This lead to something of a public health crisis in the areas where Mixtecos settled because there was no linguistic infrastructure in place to allow them to access basic health care or educational resources for their children. As of 2018, most hospitals and schools in Oxnard and Watsonville employ full-time Mixtec language liaisons.

San Martín Peras Mixtec is assigned the same ISO code, JMX, as its sister language, Coicoyán de las Flores Mixtec, and they are often referred to collectively as 'Western Juxtlahuaca Mixtec.' But there are several pronounced differences between the two that complicate mutual intelligibility. First, San Martín Peras Mixtec has undergone a phonological process of affricating coronals before [i]. This is shown in (8). All Coicoyán de las Flores Mixtec examples are from Beatham and Beatham (2014).

(8) Affication of Coronals in San Martín Peras Mixtec

	San Martín Peras Mixtec	*Coicoyán de las Flores Mixtec*	*Translation*
a.	[kitsi̋]	[kitíˊ]	'animal'
b.	[ⁿdʒuʃíˊ]	[ⁿtiuxíˊ]	'hen'
c.	[tsi̋na]	[ti̋na]	'dog'
d.	[tsi̋ː]	[ti̋ː]	'mouse'
e.	[ⁿtsikaⁿtsiː]	[ⁿtìkàⁿti̋ː]	'sun'

Second, like other Mixteca Baja languages, neither San Martín Peras Mixtec nor Coicoyán de las Flores Mixtec retain the proto-Mixtec vowel [ɨ], but the two lan-

guages often differ in their synchronic reflexes of this vowel.[2] Where San Martín Peras Mixtec usually has [u], Coicoyán de las Flores Mixtec usually has [i], although this general pattern is frequently reversed.

(9) Western Juxtlahuacan Reflexes of [ɨ]

	San Martín Peras Mixtec	*Coicoyán de las Flores Mixtec*	*Translation*
a.	[t͡ʃútũ]	[t͡ʃ́itũ]	'cat'
b.	[usu]	[isu]	'deer'
c.	[t͡sɨ̀kiβa]	[tɨ̀kuβa]	'butterfly'
d.	[iyɨ̌ʔβi]	[iyŭʔβi]	'fear (v.)'
e.	[ⁿt͡sɨ̀βáʔji]	[ⁿtɨ̀βáʔju]	'coyote'

In addition to these phonological differences the two languages have different pronoun inventories. The Coicoyán de las Flores forms are taken from Beatham and Beatham (2014) and Hollenbach (2015).

(10) San Martín Peras

		Singular		Plural
First		[ɨ̀]	INCL.	[jé]
			EXCL.	[ⁿdú]
Second		[ṹ]		[ⁿdó]
Third				
	MASC.	[ɾà]		-
	FEM.	[ɲá]		[ná]
	ANIMAL	[ɾɨ́]		-
	WOODEN	[tɨ̃́]		-
	LIQUID	[ɾá]		-
	NEUTRAL	[ɲà]		[nà]

(11) Coicoyán de las Flores

		Singular		Plural
First		[ɨ̀]	INCL.	[jo]
			EXCL.	[ⁿdi]
Second		[ũ]		[ⁿdo]
Third				
	MASC.	[tià]		-
	FEM.	[ɲá]		?
	ANIMAL	[ɾi]		-
	WOODEN	?		-
	LIQUID	?		-
	NEUTRAL	[ɲà]		[nà]

Finally, there are also interesting morphosyntactic differences between the two languages in how intransitive predicates are marked. In Coicoyán de las Flores Mixtec, intransitives are marked with the pan-Mixtec 'equative verb,' or the copula *kúú*.

(12) Coicoyán de las Flores predicates (Beatham and Beatham, 2014)

[2] My thanks to Christian DiCanio for making this point to me.

a. Tia [xáku **kúú**] -ra.
he laugh.PRES COP he
'He is laughing.'

b. Ñá [ntiáa ntikachi **kúú**] -ñá.
she watch.PRES sheep COP she
'She is sheep-watching.'

c. Na [xáxí ntiká **kúú**] na válí yó'o.
they eat.PRES banana COP they little.PL this
'These children are banana-eating.'

In (12), we see Coicoyán de las Flores Mixtec intransitives are marked with the copula *kúú*. The syntax of these constructions is not clear, but the apparently obligatory presence of the copula is different from San Martín Peras Mixtec.

In SMP Mixtec, the copula *kúú* in these constructions has been reanalyzed as an intransitive affix *-k*, which occurs at the right edge of the verbal complex.[3] Due to this phonological reduction, intransitive *-k* may only occur before vowel-initial subjects. Otherwise, *-k* would be parsed as an illegal coda or as part of a forbidden onset cluster. Only three pronouns, *ì* 'I,' *ú* 'you,' and *é* 'we (INCL)' are truly vowel initial; most other vowel initial words actually begin with a glottal stop. A brief overview of SMP Mixtec phonology will be provided in §1.3.

This means that intransitive predicates are usually unmarked, although they are consistently and obligatorily marked with *-k* so long as the independent phonological conditions are met. To see how this works, consider (13).

(13) San Martín Peras Mixtec Predicates

a. Rà xákù rà.
he laugh.PRES he
'He is laughing.'

[3]One of the San Martín Peras speakers I have worked with optionally allows forms like the Coicoyán de las Flores forms in (12), though she much more frequently produces the usual San Martín Peras forms in (13).

b. Yù'u xákù **-k** ì.
I laugh.PRES INTR I
'I am laughing.'

c. Yù'u ndzyáa lánchi **-k** ì.
I watch.PRES sheep INTR I
'I am sheep-watching.'

d. Ndzyáa -ì lánchi.
watch.PRES I sheep
'I am watching sheep.'

e. Yù'u tsyáa **-k** ì.
I write.PRES INTR I
'I am writing.'

f. Tsyáa -ì tutu.
write.PRES I paper
'I am writing a paper.'

In (13a) we see that the intransitive predicate *xakù* 'laugh' is not marked as intransitive with the affix *-k*. As stated above, this is for purely phonological reasons. If *-k* were parsed with the predicate *xakù -k* [ʃa.kùk], an illegal coda would surface. Alternatively, if *-k* parsed with the subject *rà* 'he,' yielding [krà], an illegal onset cluster would occur. To avoid these illicit phonological structures, the language simply deletes *-k*.

In contrast, in (13b), *-k* may parse with the vowel initial subject *ì* 'I,' making the legal syllable [kì]. Along with these phonological alternations, (13c-d) and (13e-f) provide the critical minimal pairs which show that *-k* marks intransitivity. In (13c), we see that when the predicate is intransitivized through noun incorporation. As such, the predicate is marked with *-k*. In contrast, in (13d) no noun incorporation takes place, and the underlying transitive structure of this predicate surfaces without any explicit marking. Finally, (13e-f) show both the transitive and the intransitive uses of the predicate *tsyaa* 'write.' In the intransitive use, the intransitive marker *-k* must be used, while in the transitive (13f), no marking takes place.

Given these differences between San Martín Peras Mixtec and Coicoyán de las Flores Mixtec, it is clear that we are either dealing with two closely related languages or two rather divergent dialects. Setting aside this question, this dissertation focuses exclusively on San Martín Peras Mixtec.

1.1.1 Methodology

All uncited San Martín Peras Mixtec data in this dissertation come from field-notes that I gathered between the spring of 2014 up through the spring of 2018. I worked with three speakers who all lived in Watsonville, California. All three speakers are women, and range from their early 20s to their mid 50s. Each speaker identified as having been raised in a Mixtec dominant household, and all of the women spoke Mixtec on a weekly, if not daily, basis. One speaker in particular works as a language liaison for Mixtec-speaking families in the Pajaro School District, in Watsonville. Two speakers are from the town of San Martín Peras itself, while the third speaker is from Ahuajutía, a smaller town about a 5 hours walk from San Martín Peras proper. All three speakers knew each other personally and were familiar with each others' Mixtec, and none identified any noticeable dialect differences amongst them. All three speakers were bilingual in Spanish, and one speaker also spoke English. For the two speakers who did not speak English, Spanish was the sole contact language of elicitation sessions, and the language through which the speakers translated the Mixtec sentences they offered. I translated their Spanish translations into English for the this dissertation, though where the exact Spanish translation was important, particularly in chapter 5, both the original Spanish and my translated English are provided.

All of the data in this dissertation come from elicited examples. Texts and narratives were not consulted because pronoun doubling simply does not occur in narratives. To see this, first consider (14), which a speaker provided after I asked directly about pronoun doubling.

(14) "*Usamos mucho el ñà. Se usa en muchas y en diferentes frases de hablar, como (14a-b). Siempre es ñá, ñá, ñá. (14c). Sí, también (14d-e), referiendo a colores. Tambien (14f-g). Hay muchas! Implica a cosas, a personas, no sé. En muchos, muchos, muchos entra.*"

"We use the *ñà* [it, referring to pronoun doubling constructions] a lot. It's used in many different sorts of sentences, like (14a-b). Always it's ñá, ñá, ñá [referring to pronoun doubling]. (14c). Yes, also (14d-e), referring to colors. Also (14f-g). There are so many! It involves things, people, I don't know. It goes in many, many, many places."

a. Ñá$_i$ íyo ñá$_i$.
she be.PRES she
'Ella sí está.' 'She is indeed here.'

b. Ñá$_i$ kǒó ñá$_i$.
she be.NEG.PRES she
'Ella no está.' 'She is not here.'

c. Ñà$_i$ xa xìxi yá$_i$.
it PERF eat.PAST it
'Ya comió.' 'It already ate.'

d. Ñà$_i$ tùhùn yá$_i$.
it black it
'Es negro.' 'It's black.'

e. Ñà$_i$ kwǎ'àn yá$_i$.
it red it
'Es rojo.' 'It's red.'

f. Ñà$_i$ kǎ'nu yá$_i$.
it big it
'Es grande.' 'It's big.'

g. Ñà$_i$ lo'o yá$_i$.
it little it
'Es pequeño.' 'It's little.'

As is clear from (14), pronoun doubling is a common process in SMP Mixtec, and speakers have no trouble generating natural sentences with it.

Despite being such a natural process, pronoun doubling never occurs in narratives, and is judged as impossible in the context of a narrative. Consider the narration that speakers provided for the Pear Film (see Chafe 1980).

(15) a. Iin na'à, iin tsyahá nàkâ kichiñu.
one morning one man be.PAST work.IRR
'One morning, a man was working.'

b. * (Rà$_i$) nàkâ rà$_i$ xa'antsya rà pêra.
he be.PAST he cut.IRR he pear
'He was cutting pears.'

c. * (Rí$_i$) nàkààba rí pêra.
it.AML fall.PAST it.AML pear
'The pear fell.'

In the narrative context in (15), pronoun doubling is never offered and judged as severely degraded if present. Speakers report that this effect is because "[pronoun doubling] *va sólo en una conversación... Sólo se usa cuando dos personas están hablando en una conversación*" ([pronoun doubling] goes only in a conversation... it is used only when two people are talking in a conversation). When asked to consider sentences like (15b-c) with pronoun doubling, speakers provide sample conversations such as (16) and (17).

(16) a. Nă kìxa rà?
what do.PAST he
'What was he doing?'

b. (***Rà**$_i$) nàkâ rà$_i$ xa'antsya rà pêra.
he be.PATS he cut.IRR he pear
'He was cutting pears.'

(17) a. Nă kúu?
what COP.PRES
'What happened?'

b. (***Rí**$_i$*) nàkààba rí pêra$_i$.
it.AML fall.PAST it.AML pear
'The pear fell.'

In (16-17), we see that pronoun doubling is considered to occur most naturally in answers to questions. The analysis in chapter 5 will account for this conversational requirement[4], but this distribution has the narrow consequence that narratives cannot be used to explore pronoun doubling.

Within the elicitation paradigm, I followed fairly standard practices (Matthewson 2004, Henderson 2012, Deal 2015b, Tonhauser and Matthewson 2015). This approach involves judgment of felicity in a context. A typical exchange of the sort used throughout this dissertation is presented in (18).

(18) a. **A:** *Imagínese que tengo unos perros. Un día, estamos en mi casa y oímos algo pasando afuera. Reconozco que el ruido es uno de los perros corriendo, pero no los veo, y por eso, no puedo saber qual perro está corriendo. Usted me pregunta "Qué está pasando?" Cómo respondería yo: "Uno de los perros está corriendo, pero no sé cual es."*

Imagine that I have some dogs. One day, we are in my house and we hear something happening outside. I recognize the sound as one of the dogs running, but I don't see them, and for this reason, I can't know which dog is running. You ask me "What's happening?" How would I respond "One of the dogs is running, but I don't know which one it is."

b. **B:** Iin tsi̍na xíka xínu rí, so xi̍n ì ntsyâ
one dog walk.PRES.SG run.PRES it.AML but know.PRES.NEG I which
rí kúú rí.
it.AML COP.PRES it.AML

(18a) presents a context against which a target sentence, in this case "*Uno de los perros*

[4]Specifically, we shall see that pronoun doubling is only licit if the nominal it targets is identifiable in the Common Ground. As narratives have no Common Ground as they lack conversational participants, this effect is a predicted consequence of the analysis.

está corriendo, pero no sé cual es ∼ One of the dogs is running, but I don't know which one it is," is meant to be translated. While this explicit context may seem trivial, it is necessary to fully understand what the target translation in (18b) actually means.

Once a context for the target utterance was established, I would present the speaker with a question of the form in (19).

(19) *En este contexto, cómo le parece decir:*

In this context, how does it seem to you to say:

a. Xíka iin tsǐna xínu rí, so xǐn ì ntsyâ rí
walk.PRES.SG one dog run it.AML but know.PRES.NEG I which it.AML
kúú rí.
COP.PRES it.AML

The sentence in (19) differs minimally from the speaker's original translation in (18b); in (18b), the quantification expression *iin tsǐna* 'a dog' occurs preverbally, while in the sentence in (19), *iin tsǐna* 'a dog' occurs postverbally. The speaker is asked to judge (19) relative to the same context as for (18b), and speakers were not given any restrictions on how they could react to questions like (19).

Generally, speakers react in one of two ways to a question like (19). One possibility is acceptance of the artificial sentence as a plausible and natural alternative to their first translation in (18b), which is indeed the case for (19). If speakers responded in this way, then I consider the elicited sentence to be grammatical and able to serve as data.

Another possibility is that the speaker rejects the elicited sentence. There are two ways that this rejection could manifest itself. The speaker could reject the elicited sentence categorically, judging it as not a part of their language. This judgment of ungrammaticality can serve as data, just as a judgement of grammaticality can.

Alternatively, the speaker could reject an elicited sentence as a plausible alternative relative to the given context, while still acknowledging that it could serve as a grammatical sentence in a different context. If this occurs, then one may probe further

by asking the speaker if they can think of any contexts in which the elicited sentence would sound more natural. In my experience, speakers are generally good at this task, and usually provide alternative contexts readily and confidently. It is then up to the analyst to understand what aspect of the alternative context is relevant for the analysis, and to then generate alternative hypotheses that may be tested using the same task.

This translation task relative to a provided context served as the primary method of gathering the data in this dissertation. The only time where this method was avoided was if the target sentence, no matter the context, could be construed as ambiguous. This was most commonly the case when investigating quantifier scope. To accomplish this, I relied on the richer, visual contexts of Bruening (2008) in order to make sure that the speaker was judging the target sentences relative to the intended scopal relations.

Finally, all elicitation sessions were recorded in a quiet room, either in the linguistics department at UC Santa Cruz or in a study room at the Watsonville Public Library. Rarely, elicitation sessions are recordings took place at a speaker's home, if that was most convenient for them. All speakers were compensated at a rate of $25 an hour for their time.

1.1.2 A very basic phonological sketch of San Martín Peras Mixtec

While the focus of this dissertation is firmly in the syntax, semantics, and morphology of SMP Mixtec, some cursory phonological remarks are in order. In particular, SMP Mixtec is an unwritten language, so the orthography used here is largely ad hoc and non-standardized.

(20) provides the consonant inventory, as well as the letters used to represent those sounds.

(20) SMP Mixtec Consonant Inventory

	Labial	*Coronal*	*Velar*	*Glottal*
Stops	b /β/	t /t/, nt /^{n}t/, nd /^{n}d/	k /k/, kw /k^{w}/	’ /ʔ/
Nasals	m /m/	n /n/, ñ /ɲ/		
Tap		r /ɾ/		
Fricative		s /s/, x /ʃ/	g /ɣ/	h /h/
Affricates		ts /t͡s/, nts /^{n}t͡s/, ch /t͡ʃ/, nch /^{n}t͡ʃ/, ndz /^{n}d͡ʒ/		
Laterals		l /l/		
Approximant			y /j/	

SMP Mixtec is particularly rich in coronal obstruents, while labials are rare. Of the velars, only /k/ and /k^{w}/ are common, with /ɣ/ being attested only in a handful of morphemes, notably *ga* ‘more.’ /l/ is also infrequent. Loan words from Spanish are rarely adapted at the segmental level. For instance, Spanish *pato* ‘duck’ is borrowed as *páto* [páto], even though SMP Mixtec lacks /p/ as a native phoneme.

The vowel inventory consisting of /a/, /e/, /i/, /o/, and /u/. This five vowel system is common to all Mixteca Baja varieties. As discussed briefly in §1.1, the proto-Mixtec sixth vowel /ɨ/ is usually reassigned to /u/, or occasionally /i/.

Like every other Oto-Manguean language (see DiCanio and Bennett To appear), SMP Mixtec is highly tonal. By my best estimate without having done any in-depth statistical or acoustic analysis, there are eight, perhaps nine, tones, although this number varies depending on analysis. These include a low tone, marked with a grave accent, and a high tone, marked with an acute accent. Additionally, there are approximately six rising and falling tones: low-to-mid rise, mid-to-high rise, low-to-high rise, with the same three in reverse to yield the three falling tones. Throughout, all rising tones are marked as V̌, while all falling tones are marked with V̂. All vowels in a word may be marked with any one of these eight tones, but vowels need not be tonally specified (Daly and Hyman 2007). Throughout, I have transcribed tone to the best of my abilities, but my tonal transcriptions and analyses should be taken as preliminary at best.

Tone is indispensable in understanding SMP Mixtec morphology because many morphemes are exponed via tone. Of particular note, verbs are conjugated for tense/aspect/mood[5]

[5]Traditionally, this system is described as an aspect system in Mixtec languages. After conducting the TAM survey in Dahl (1985), it seems that SMP Mixtec speakers have begun to reanalyze this as a

and even negation with tonal morphemes. Consider the paradigms in (21).

(21)

		PAST	PRES	IRR	IRR.NEG
a.	'give'	tàxi	táxi	taxi	tăxi
b.	'like/love/want'	kòni	kôni	koni	kŏni
c.	'believe'	kà'a	ká'a	ka'a	kă'a
d.	'talk'	nì ka'an	ká'an	ka'an	kă'an

As can be seen in (21), most TAM categories are marked with changes to the tone of the first syllable. The present tense/progressive aspect is marked with a high tone or a high-to-low fall on the first syllable, the negative irrealis with a low-to-high rise on the first syllable, and the declarative irrealis is marked by not tonally marking the first syllable. The past tense/completive aspect is usually marked with a low tone on the initial syllable, although some predicates take the prefix *nì*. In SMP Mixtec, *nì* behaves like an idiosyncratic allomorph of the past tense, although it seems to be derived from a productive completive aspect marker.

In addition to tone, vowels may be marked with nasalization. That said, the nasalization system in SMP Mixtec is a shadow of that found in other Mixtec languages where there is a rather elaborate system of nasalization in which nasalization associates with the rightmost syllable in the word and then spreads, leading to nasal harmony throughout the word (Marlett 1992). Nasalization in SMP Mixtec is still restricted to the rightmost syllable of the word, though, and minimal pairs do exist, such as *ka'a* [kaʔa] ∼ *ka'an* [kaʔã] 'believe ∼ talk,' *xá'a* [ʃáʔa] ∼ *xá'an* [ʃáʔã] 'goes ∼ smells,' and *kwá'a* [kʷáʔa] ∼ *kwă'àn* [kʷăʔà̃] 'red.' I mark nasalization throughout with a post-vocalic 'n.'

Finally, vowels in SMP Mixtec are not necessarily modal. The most common kind of non-modal vowel is 'glottalized' or 'creaky,' as in many other Mixtec languages

tense system. That says, the system is leveraged to convey aspect in traditional narratives. Ultimately, the precise and optimal semantic characterization of this is ultimately peripheral to our concerns, and I refer to this system as a tense system throughout.

(Macaulay and Salmons 1995). In SMP Mixtec, the phonetic realization of non-modal vowels depends largely on its environment in the word. If a non-modal vowel precedes a syllable with an onset, it is usually pronounced with a glottal stop and an optional echo vowel. For instance *ndá'bi* 'poor' is articulated as either [^{n}daʔβi] or [^{n}daʔ^{a}βi].

In addition, the pronunciation of glottalized vowels is also sensitive to the broader prosodic context of the word. First, SMP Mixtec, like every other Mixtec language, has a strict word minimality requirement. Famously referred to as the 'couplet' going back to Pike (1948), all maximal prosodic words in SMP Mixtec must be at least bimoraic. Now, consider a root like *tù'un* 'language,' which is underlying [tu̯]. If this were to occur in isolation, then it would violate the language's word minimality requirement. As such, it is pronounced [tùʔũ]. But if *tù'un* 'language' occurs in a compound, its subminimal underlying form may surface. This can be seen in the name of SMP Mixtec in the language, which is *tù'un ndá'bi* [tu̯ ^{n}daʔβi] 'language of the poor.'

Lastly, a distinctive feature of SMP Mixtec is the presence of breathy vowels. To my knowledge, these are unattested elsewhere in the Mixtec languages, but they are found in the broader Mixtecan family, such as in San Martín Itunyoso Trique (DiCanio 2008). Breathy vowels are usually articulated with an initial modal vowel which transitions into a voiceless vowel, like *náhma* [na̤ma] ∼ [nahma] 'soap.' In some roots, breathy vowels alternate with geminates, like *kanahta* [kana̤ta] ∼ [kanatːa] 'go out (SING).'

With this cursory phonological sketch, let us begin to consider the basic syntax of this language.

1.1.3 The basics of SMP Mixtec syntax and pronoun doubling

As presented briefly above, San Martín Peras Mixtec demonstrates a pattern unattested in other Mixtec languages[6] that involves an initial pronoun that cross-

[6]Coicoyán de las Flores Mixtec seems to have an identical pattern, although I only have a few sources in that language available to me. See (12).

references the φ-features of a nominal in the clause. Recall that ☒ indicates an absence of a doubled pronoun.

(22) A Random Sampling of Other Mixtecs

a. ☒ Vāchi yuhu.
come.CONT I
'I am coming.' (Jamiltepec Mixtec, Johnson 1988)

b. ☒ Ndávā sāa.
jump.CONT bird
'The bird is jumping.' (Ocotepec Mixtec, Alexander 1988)

c. ☒ Kíxi ndó.
sleep.CONT you.PL
'You (PL) are sleeping.' (Silacayoapan Mixtec, Shields 1988)

d. ☒ Kūshū yó.
eat.POT we.INCL
'We (INCL) will eat.' (Alacatlatzala Mixtec, Zylstra 1991)

e. ☒ N-shíhí vilú.
COM-die cat
'The cat died.' (Diuxi-Tilantogo Mixtec, Kuiper and Oram 1991)

f. ☒ Kúhu ī.
be.sick.CONT GEN
'The child is sick.' (Yosondúa Mixtec, Farris 1992)

(23) San Martín Peras Mixtec

a. *(**Yù'u**) báxí -k ì.
I come.PRES -INTR I
'I am coming.'

b. *(**Rí**) ndába sâ.
it.AML jump.PRES bird
'The bird is jumping.'

c. *(**Ndó**) kíxi ndó.
you.PL sleep.PRES you.PL
'You (PL) are sleeping.'

d. *(**Yé**) kuxi -k yé.
we.INCL eat.IRREAL -INTR we.INCL

'We (INCL) will eat.'

e. *(**Rí**) ní xi'i chútun.
it.AML PAST die cat
'The cat died.'

f. *(**Ñà**) kú'u se'e.
it be.sick.PRES child
'The child is sick.'

(22) presents a random sampling of other Mixtec languages. Nothing like pronoun doubling occurs in any of these, or in any other Mixtec language to my knowledge.

From this perspective, consider (23). Here we see a variety of intransitive predicates in SMP Mixtec where the subject seems to be marked in two places. The first is with a preverbal morpheme that is morphophonologically identical to a regular pronoun, while the second is a post-verbal DP. In (23), we see that no other Mixtec language demonstrates a similar pattern.[7]

This leads to a question: which of the two apparent DPs in sentences like (23) is the true subject? This is not a trivial question because, as I have stated various times, the preverbal bolded morphemes in (23) are identical to regular pronouns in the language. Reconsider the complete pronoun inventory of this language, reproduced and expanded from (10).

(24) Local Persons

	SG	PL	
1	yù'u ∼ ì [jṵ̀ ∼ ḭ̀]	INCL	yé ∼ é [jé∼ é]
		EXCL	ndú [ⁿdú]
2	yô'o ∼ ú [jô̰ ∼ ú]		ndó [ⁿdó]

(25) 3rd Persons

	Neutral	Feminine	Masculine	Liquid	Wooden	Animal
SG	ñà ∼ yá [ɲà ∼ já]	ñá [ɲá]	rà [ɾà]	rá [ɾá]	tún [tṹ]	rí [ɾí]
PL	nà [nà]	ná [ná]				

[7]While comparable processes are unattested in other Mixtec language, agreement systems abound in other branches of Oto-Manguean, such as Chinantec (Foris 2000), Tlapanec (Wichmann 2007) and Otomí. (Lastra 1992, 1997, Palancar 2009).

In (24) we see that the local persons show a clusivity distinction, while the non-local in (25) demonstrate a range of grammatical genders. Feminine and masculine humans receive distinct pronouns, as do liquids, wooden things, and animals. If a nominal does not fit into one of these categories, it is marked using 'neutral' pronouns. Only the neutral and feminine pronouns have distinct plural forms. For other genders, the same pronoun is used for both singular and plural referents. Where two pronouns are provided, the choice between them is prosodically determined, with the former forms occurring before nuclear stress within a phrase.

This system is best understood as grammatical gender. For instance, cars and other vehicles, as well as long, skinny objects like *itsyă* 'candle' obligatorily take 'wooden' pronouns despite not being made of wood. Most celestial bodies, especially *yohó* 'moon,' all fruit, and some other words like *nahmà* 'soap' take 'animal' pronouns. Human babies, no matter their sex, take neutral pronouns. In spite of the obvious shortcomings of this terminology, I will continue to use the terms 'wooden,' 'animal,' and 'neutral' pronouns, following the traditional terms in the Mixtec literature (see Hollenbach 2015). Speakers find these labels for these categories to be the most intuitive as well.

In addition to the use of such pronouns in pronoun doubling constructions, the pronouns in (24-25) serve in functions comparable to pronouns in other languages, serving, for example, as subjects, objects, possessors, and the objects of prepositions.[8]

(26) Pronoun as Subject

a. Bítsí táxa'a **ñá**.
now dance.PRES she
'She is dancing now.' (Intransitive subject)

[8]Pronouns in SMP Mixtec, like other Mixtecan languages, do not show morphological case distinctions. I gloss pronouns into English using the appropriate case forms for expository clarity.

Additionally, it is unclear if SMP Mixtec, or any other Mixtec language, truly has prepositions. This is becausethese candidate prepositions are morphophonologically identical to body parts. For instance, *nda'ă* 'to' in (26c) also means 'hand,' *xí'in* 'with' in (29a) also means 'side,' and *nùhŭ* 'to' in (29b) also means 'face.' Throughout, I set this question aside because it is unclear to me what evidence might be used to argue for one position or another (but see Lillehaugen 2003). I gloss these body part prepositions according to the most natural preposition in English.

b. Xíxi **rí** uhtu iin rà tsyahá.
eat.PRES it.AML corn one he man
'It (an animal) ate a man's corn.' (Transitive subject)

c. Tàxi **rà** nuhnĭ nda'ă míí kînì.
give.PAST he corn.in.husk to the pig
'He gave corn to the pig.' (Ditransitive subject)

(27) Pronoun as Object

a. Kôni ñá **rá**.
want.PRES she it.LIQ
'She wants it (a liquid).'

b. Kani míí burru **tún**.
hit.IRR the donkey it.WOOD
'The donkey will hit it (a wooden thing).'

(28) Pronoun as Possessor

a. be'e **rà**
house his
'his house'

b. málí **ñá**
conmadre her
'her *conmadre*'

c. ntó'ó **rí**
tail it.AML
'its tail (an animal)'

(29) Pronoun as Object of Preposition

a. Să káchi rà să [xí'in **rà**]...
then say.PAST he then with him
'Then he said to him...'

b. Ixiko tún [nuh **ì**]!
sell.IRR it.WOOD to me
'Sell it (a wooden object) to me!'

The pronouns in (26-29) are not particularly noteworthy from a cross-linguistic perspective. But from this rather vanilla distribution, an interesting dilemma emerges when we reconsider pronoun doubling. (30) provides further examples of this phenomenon.

(30) a. **Yô'o** táxa'a -k ú.
you dance.PRES -INTR you
'You are dancing.'

b. **Ñà** kùnaha yá.
it get.dark.PAST it
'It got dark.'

c. **Ñá** náka'ama ísi' ì.
she sew.PRES wife my
'My wife is sewing.'

d. **Tún** nàkaaba ítŭn.
it.WOOD fall.PAST tree
'The tree fell.'

e. **Rá** bìxì ntsyahi ó'óba.
it.LIQ cold soup thick
'The *mole* (thick soup) is cold.'

In (30), as well as in (23), we have two elements that could conceivably be the true argument of the verb. The first is the DP in post-predicative position, such as *ísi'ì* 'my wife' in (30c) or *ntsyahi ó'óba* 'mole' in (30e). But as we have just seen, these pronouns can also be arguments, so we could equally imagine that the preverbal pronouns are the true argument while the post-predicative DPs are some kind of adjunct. Notably, this sort of configuration is argued for by Jelinek (1984) for Warlpiri. This is reasonable at first blush for that language because second position pronouns are obligatory, where full DPs are not.

(31) Warlpiri (Jelinek, 1984)

a. Ngarrka -ngku ka -∅ -∅ wawirri panti -rni.
man -ERG PRES -3SG.ERG -3SG.ACC kangaroo spear -NONPAST

'The man is spearing the kangaroo.'

b. Ngarrka -ngku ka -∅ -∅ ☒ panti -rni.
man -ERG PRES -3SG.ERG -3SG.ACC spear -NONPAST
'The man is spearing him/her/it.'

c. Wawirri ka -∅ -∅ ☒ panti -rni.
kangaroo PRES -3SG.ERG -3SG.ACC spear -NONPAST
'He/she is spearing the kangaroo.'

d. Panti -rni ka -∅ -∅ ☒ ☒.
spear -NONPAST PRES -3SG.ERG -3SG.ACC
'He/she is spearing him/her/it.'

Jelinek (1984) argues that the pattern in (31) makes sense if the second position pronominals are the true arguments that satisfy the subcategorization requirements of the predicate. This would make the full DPs not obligatory syntactically or semantically, deriving their optional occurrence.

Returning to SMP Mixtec, there are several reasons to consider the post-predicative DP to be the true subject. First, in contrast with Warlpiri, doubled pronouns do not occur in many cases, whereas post-predicative DPs are always obligatory. We will thoroughly examine the conditions which both license and block otherwise licit doubled pronouns in Section 2 of this chapter, but for now the core observation in (32-34) suffices.

(32) Verbal predicate

a. ☒ Kamá ntáchi̱ sâ.
fast fly.PRES bird
'The bird is flying quickly.'

b. *__Rí__ kamá ntáchi̱ ☒.

c. *Kamá __rí__ ntáchi̱ ☒.
fast it.AML fly.PRES
Intended: 'It is flying fast.'

(33) Nominal predicate

a. ☒ Doktór ndibi rà Juân.
doctor handsome he J.
'Juan is a handsome doctor.'

b. *__Rà__ doktór ndibi ☒.
he doctor handsome
Intended: 'He is a handsome doctor.'

(34) Adjectival predicate

a. ☒ Kă'anu kwê ntsìka kân.
big very banana that
'That banana is very big.'

b. *__Rí__ kă'anu kwée ☒.
it.AML big very
Intended: 'It is very big.'

In (32-34), we see that preverbal pronouns may not occur in a variety of instances, but throughout postpredicative DPs are obligatory. This provides initial evidence that it is the post-predicative DP that satisfies the subcategorization requirements of the predicate, and are therefore the true argument, unlike in Warlpiri.

There is further evidence that post-predicative nominals act as the syntactic subject. First, post-predicative DPs may bind PRO in the absence of a doubled pronoun. Note that PRO in SMP Mixtec is obligatorily overt, as we will discuss in Chapter 3.

(35) a. ☒ Kìxă míí leso$_i$ [taxa'a rí$_i$].
start.PAST the rabbit dance.IRR it.AML
'The rabbit started to dance.'

b. ☒ Kôni rà lo'o$_i$ [tsii rà$_i$ míí sá'aba].
want.PRES he little catch.IRR he the frog
'The boy wants to catch the frog.'

c. ☒ Nàntŏso ñá kân$_i$ [nakatsya ñá$_i$ míí tsyàà].
forget.PAST she that wash.IRR she the clothes
'That woman forgot to wash the clothes.'

The ability to bind PRO is a common diagnostic of structural arguments. Therefore, we have a second piece of evidence that post-predicative DPs are the true arguments of the predicate.

Second, post-predicate DPs can bind reflexives in object position, again in the absence of a doubled pronoun. Reflexives in SMP Mixtec are composed of a morpheme *míí*, usually translated by speakers as 'the,' and a pronoun.

(36) a. Nuhŭn yùtátá xìto'oni ñá Maria$_i$ míí ñá$_i$.
in mirror see.PAST M. the she
'Maria saw herself in the mirror.'

b. *Nuhŭn yùtátá xìto'oni míí ñá$_i$ ñá Maria.
in mirror see.PAST the she M.
Intended: 'Maria saw herself in the mirror.'

In both examples in (36), we see that no doubled pronoun occurs. Despite this, in (36a) a post-predicative DP may bind a reflexive. This indicates that the post-predicative DP, in this case *ñá Maria* 'Maria,' is structurally higher than the reflexive 6textitmíí ñá 'herself.' Furthermore, (30b) demonstrates that the first DP in a string [Pred DP DP] must be more prominent than the second. This has a natural explanation if we analyze the first DP as the external argument and the second as the internal argument.

From (32-36) a unified picture emerges: in SMP Mixtec, arguments follow their selecting predicates in a strict order: first, the subject, then the object. In other words, SMP Mixtec is a VSO language, like other Mixtec languages (see Macaulay 2005). In this way, the pre-predicative doubled pronouns in pronoun doubling constructions cannot be considered arguments.

Before continuing our investigation into what these preverbal pronouns in pronoun doubling constructions are, I would like to point out a few aspects of the syntax of this language that I hope will aid readers moving forward. First, as we have seen already in the examples so far, nominals that refer to humans are frequently preceded

by a pronominal that matches the noun in gender and number. These include proper names like *rà Juân* 'lit. he Juan' and *ñá Maria* 'lit. she Maria,' names of professions such as *ñá doktóra* 'lit. she doctor,' and terms like *rà lo'o* 'lit. he little' to mean 'the boy.' Speakers vary in whether these pronouns are obligatory. For some, proper names and professions do not require a pronoun or are compatible with *míí* 'the,' while for others the pronoun is obligatory. For all speakers, the pronoun is obligatory for phrases like *rà lo'o* 'lit. he little.'

Second, adjuncts may never appear between V and S, or S and O. Adjuncts can be divided into two broad classes which I refer to as high adjuncts and low adjuncts.

(37) **Low adjuncts**: *bàko'é* 'often,' *na'à* 'early,' *xkwàhà* 'late,' *ntsìbe'e* 'recently,' *ba'á* 'well,' *kamá* 'quickly,' *sábáso'o* 'suddenly,' *yatsíín* 'almost,' *intoso'o* 'briefly,' *inkatúkú* 'again,' *tásа'aba* 'usually,' *táto'oba* 'possibly/probably,' *bà'akwê'e* 'thankfully/fortunately'

(38) **High adjuncts**: *koni* 'yesterday,' *bítsí* 'now,' *itsyààn* 'tomorrow,' *nuhŭ iin tòò lo'o* 'soon (in a little while),' *yó'o* 'here,' *kân* 'there'

Note that the distinction between high and low adjuncts in (37-38) is not the same as the notion of 'high' and 'low' in the sense of Cinque (1999). As can be seen in (37), the low adjuncts here include typologically low adjuncts like *bàko'é* 'often' and *kamá* 'quickly' as well as typologically high adjuncts like *bà'akwê'e* 'thankfully/fortunately' and *táto'oba* 'possibly/probably.'

To see the difference between these two classes of adjuncts, let us begin by examining low adjuncts. All of these adjuncts are obligatorily preverbal: no adjuncts may intervene between V and S or S and O.[9] This is demonstrated in (39-40) with *bàko'é* 'often' and *táto'oba* 'probably/possibly.' Note that the former is one of the

[9]The only exception to this are a set of clitics which correspond to phrasal adjuncts cross-linguistically. These obligatorily cliticize to the predicate, no matter its category. These clitics are *=ni'i* 'still,' *kâ* 'already,' and *túkú* 'again.' (i) demonstrates the pattern with *=ni'i* 'still.'

lowest adjuncts typologically that has a lexical correspondent in this language, while the latter is one of the highest (Cinque 1999).

(39) Typologically low adjunct *bàko'é* 'often'

a. **Bàko'é** xáxi'i'i tsǐna míí chútun.
often bite.PRES dog the cat
'The dog often bites the cat.'

b. *Xáxi'i'i **bàko'é** tsǐna míí chútun.

c. *Xáxi'i'i tsǐna **bàko'é** míí chútun.

d. *Xáxi'i'i tsǐna míí chútun **bàko'é**.

(40) Typologically high adjunct *táto'oba* 'probably/possibly'

a. **Táto'oba** kàni rà Julito ñá Maria.
probably hit.PAST he J. she M.
'Little Julio probably hit Maria.'

b. *Kàni **táto'oba** rà Julito ñá Maria.

c. *Kàni rà Julito **táto'oba** ñá Maria.

d. *Kàni rà Julito ñá Maria **táto'oba**.

When more than one low adjunct occurs, they both precede the verb and show the ordering predicted by the Cinque Hierarchy of Cinque (1999).

(41) a. Táto'oba kamá kàrakono kù' í.
probably quickly run.PAST sister my
'My sister probably ran quickly.'

---

(1) a. Náka'ma =ni'i ñá Maria.
weave.PRES =still she M.
'Maria is still weaving.'

b. Sókŏn =ni'i rà Juan.
tall =still he J.
'Juan is still tall.'

c. Mastro síki kúu =ni'i tát ì.
teacher funny COP.PRES =still father my
'My father is still a funny teacher.'

b. *Kamá táto'oba kàrakono kù' í.

c. Táto'oba ba'ă xíta amiga ña' ì.
probably well sing.PRES friend POSS my
'My friend is probably singing well.'

d. *Ba'ă táto'oba xíta amiga ña' ì.

Setting aside the specific syntax of these adjuncts, let us contrast these low adjuncts with what I refer to as high adjuncts. High adjuncts are time adjuncts like *koni* 'yesterday' or *nuhŭn iin tòò lo'o* 'in a little while/soon,' as well as place adjuncts like *yúkŭ* 'in the forest,' *bihkŏ* 'at the party,' or *Ñû Kă'nu* 'in San Martín Peras.' Unlike low adjuncts which are rigidly before the predicate, high adjuncts show a bit more flexibility. They may occur both initially, including before any low adjuncts, or finally.

(42) a. **Skwéla** táto'oba bàko'é káni rà Julito ñá Maria.
school probably often hit.PRES he J. she M.
'Little Julito probably hits Maria often at school.'

b. *Táto'oba **skwéla** bàko'é káni rà Julito ñá Maria.

c. *Táto'oba bàko'é **skwéla** káni rà Julito ñá Maria.

d. *Táto'oba bàko'é káni **skwéla** rà Julito ñá Maria.

e. *Táto'oba bàko'é káni rà Julito **skwéla** ñá Maria.

f. Táto'oba bàko'é káni rà Julito ñá Maria **skwéla**.

As can be seen in (42), high adjuncts are licit in only two positions. The first is in absolute initial position, to the left even of typologically high adjuncts like *táto'oba* 'probably/possibly.' The second is in absolute final position, occurring after all verbal arguments.

If more than one high adjunct occurs, only one may occur in initial position. Alternately, both may occur in final position, although there is a strong preference to have one initially.

(43) a. **Itsyààn** kani rà Julito ñá Maria **swkéla**.
tomorrow hit.IRR he J. she M. school
'Tomorrow little Julito will hit Maria at school.'

b. **Swkéla** kani rà Julito ñá Maria **itsyààn**.

c. *__Swkéla itsyààn__ kani rà Julito ñá Maria.

d. *__Itsyààn swkéla__ kani rà Julito ñá Maria.

e. ??Kani rà Julito ñá Maria **itsyààn swkéla**.

f. ??Kani rà Julito ñá Maria **swkéla itsyààn**.

This strict adverb ordering means that we do not have many of the usual tools available to us to diagnose the clause structure amongst V, S, and O (see Pollock 1989, among many others). Likewise, negation is exclusively a prefix on the verb. (44) demonstrates the allomorphs of negation in this language.

(44) a. **Kò** ní ka'abi rà abogado míí libro.
NEG PAST read he lawyer the book
'The lawyer did not read the book.'

b. **Kò** xí'ì ñá Maria míí tskwî.
NEG drink.PRES she M. the water
'Maria is not drinking the water.'

c. **Kă'abi** rà abogado míí tútu yó'o.
read.IRR.NEG he lawyer the paper this
'The lawyer won't read this paper.'

d. **Sĭ** rà brujo rà.
NEG.PRES.NOM he sorcerer he
'He is not a sorcerer.'

e. **Ĭ** kwă'àn kúu míí libro kân.
NEG.ADJ red COP.PRES the book that
'That book is not red.'

In (44a-b), we see *kò* is the negation marker for the past and present tenses.[10] In contrast, a tonal allomorph of negation is obligatory in the irrealis. This tone is a steep rising tone on the initial syllable, marked here as V̌.[11] Likewise, in (44d), we see that there is a special negation marker *sǐ* for nominal predicates in the present tense, while (44e) shows that adjectival predicates have a unique negation marker *ǐ*. As all of these negation morphemes are prefixal, we cannot use the syntax of negation to map the clausal syntax between V, S, and O either.

As such, I largely abstract away from the lower clausal syntax between V, S, and O in this language. Throughout, all that will be important is that S is structurally superior to O, which we diagnosed in (36) on the basis of reflexive binding.

With this in place, I will demonstrate what this dissertation is not about.

[10]Note as well that *kò* requires the *nì* allomorph of past, which, as mentioned around (21), is otherwise idiosyncratically selected by certain predicates. This is shown in (i).

(1) a. Kà'a rà Juân...
think.PAST he J.
'Juan thought...'

b. Nì ka'an rà Juân...
PAST talk he J.
'Juan talked...'

c. Kò ní ka'a rà Juân...
NEG PAST think he J.
'Juan didn't think.'

d. Kò ní ka'an rà Juân...
NEG PAST talk he J.
'Juan didn't talk...'

In (ia), we see that *ka'a* 'think' takes the tonal allomorph of past tense, while (ib) shows that *ka'an* 'talk' takes the past tense allomorph *nì*. In (ic-d) we see that negative *kò* requires the allomorph *nì*. Note that the tonal change of *nì* in *kò ní* is a regular sandhi process that affects sequential low tones in a prosodic word and is not relevant for our purposes.

[11]See Palancar et al. (2015) for similar tonal morphology involving both tense and negation in Yoloxóchitl Mixtec, which is spoken about 15 miles south of San Martín Peras across the boarder into the state of Guerrero.

1.2 What this dissertation is not about: the prosody of pronoun doubling

In previous work (Ostrove 2017b), I have shown that the distribution of pronoun doubling is sensitive to certain prosodic factors. I will present these findings briefly with the purpose of abstracting away from them in the remainder of the dissertation.

In many cases, pronoun doubling is obligatory. This is the case that we have seen so far, and further examples are provided in (45).

(45) a. **Rà** [doktór] rà Juán.
he doctor he J.
'Juan is a doctor.'

b. **Ñá** [sókŏn] ñá Maria.
she tall she M.
'Maria is tall.'

c. **Rí** [kárakono] tsĭna.
it.AML run.PRES dog
'The dog is running.'

In all of the examples in (45), failing to have an initial pronominal double results in ungrammaticality.

(46) a. *☒ [Doktór] rà Juán.

b. *☒ [Sókŏn] ñá Maria.

c. *☒ [Kárakono] tsĭna.

Critically, this distribution seems to be subject to both prosodic and syntacticosemantic factors. Focusing on the prosodic factors, consider what happens if the predicates in (45) are modified by an adjunct.

(47) a. ☒ [Doktór siki] rà Juán.
doctor funny he J.
'Juan is a funny doctor.'

b. ☒ [Sókŏn kwê'e] ñá Maria.
tall very she M.
'Maria is very tall.'

c. ☒ [Kamá kárakono] tsĭna.
fast run.PRES dog
'The dog is running fast.'

In (46), the interpretation of the subject is no less specific than in other examples we have seen. The sole elaboration in (46) over previous examples is that a modifier has been added to the predicate, such as *siki* 'funny' in (46a) or *kamá* 'quickly' in (46c). Interestingly, the addition of this element blocks pronoun doubling.

I consider the contrast between (45) and (47) to be nonsyntactic because I am aware of no theory of either φ-agreement or clitic doubling that would predict this sort of sensitivity to adjuncts. Furthermore, it is unclear that sensitivity to adjuncts should be built into the theory of either.

Interestingly, there is a clear prosodic generalization that can be made with respect to the contrast in (45-47). To see this, consider the phrasings of (47) in (48).[12] Here ϕ stands for a phonological phrase and ω stands for a phonological word.

(48) a. $^{\phi}($ $^{\omega}($ doktór $)$ $^{\omega}($ siki $)$ $)$ $^{\phi}($ rà Juán $)$

b. $^{\phi}($ $^{\omega}($ sókŏn $)$ $^{\omega}($ kwê'e $)$ $)$ $^{\phi}($ ñá Maria $)$

c. $^{\phi}($ $^{\omega}($ kamá $)$ $^{\omega}($ kárakono $)$ $)$ $^{\phi}($ tsĭna $)$

In (48), the adjunct and predicate form a phonological phrase. Bear in mind that this structure does not support pronoun doubling. Now, consider what the prosodic structure of the ungrammatical (46) would be.

(49) a. $*^{\phi}($ $^{\omega}($ doktór $)$ $)$ $^{\phi}($ rà Juán $)$

b. $*^{\phi}($ $^{\omega}($ sókŏn $)$ $)$ $^{\phi}($ ñá Maria $)$

[12] For details on the phonetic diagnostics which lead to this phonological phrasing, see Ostrove (2017b).

c. *$^{\phi}$($^{\omega}$(kárakono)) $^{\phi}$(tsǐna)

In comparing the ungrammatical (49) with the grammatical (48), we see that binarity of the initial prosodic phrase matters. If the leftmost prosodic phrase would not be binary without a doubled pronoun, then pronoun doubling is obligatory. This, I claim, is the basis for the observations summarized in (45-47). Likewise, if the leftmost prosodic phrase can be binary without a doubled pronoun, then pronoun doubling fails to occur. This is the case in (47).

This type of mandatory binarity is well-attested at various levels of the prosodic hierarchy in a variety of languages (Elordieta 1997, Elordieta 2007, Selkirk 2011, Bennett 2012), and such obligatory binary enforced for the leftmost prosodic phrase is a predicted pattern in prosodic typology. Therefore, I codify this restriction for SMP Mixtec in (50).

(50) MinBin(φ, L)The leftmost prosodic phrase in SMP Mixtec must be prosodically binary.

Now let us consider the prosodic status of the doubled pronoun. First, we know that it cannot be a prosodic word of the same level as the adjuncts in (48). This is because SMP Mixtec, like all other Mixtec languages, has a strict minimal binarity requirement on all prosodic words (Pike 1948, Pike and Cowan 1967, North and Shields 1977, Meacham 1991, Macaulay and Salmons 1995, Macaulay 1996, Gerfen 1996, Daly and Hyman 2007, DiCanio 2008, DiCanio and Bennett To appear, among many others). This minimal binarity requirement can be summarized using a MinBin constraint of the sort typical in prosodic phonology, shown in (51).

(51) MinBin(ω): Prosodic words in SMP Mixtec must be at leads bimoraic.
*CV, ✓CV:, ✓CVCV

(51) encodes the concept of the "bimoraic couplet" that is central to the prosody of every Mixtec language. Now, let us reconsider the prosodic shape of the pronouns in pronoun doubling constructions. (52-53) are reproduced from (24) and (25) above.

(52) Local Persons

	Sg		Pl
1	yù'u ∼ ì [jṵ̀ ∼ ḭ̀]	Incl	yé ∼ é [jé∼ é]
		Excl	ndú [ⁿdú]
2	yô'o ∼ ú [jô̰ ∼ ú]		ndó [ⁿdó]

(53) 3rd Persons

	Neutral	Feminine	Masculine	Liquid	Wooden	Animal
Sg	ñà ∼ yá [ɲà ∼ já]	ñá [ɲá]	rà [ɾà]	rá [ɾá]	tún [tṹ]	rí [rḯ]
Pl	nà [nà]	ná [ná]				

As is clear from (52-53), all of the pronouns are monomoraic. Therefore, none of these will be able to satisfy the MinBin(ω) constraint in (51) and therefore cannot be parsed as prosodic words. As such, doubled pronouns are presumably parsed as simple syllables. Consider the proposed prosodic representation of sentences like (45) will be as in (54).

(54) a. $^{\phi}$($^{\sigma}$(rà) $^{\omega}$(doktór)) $^{\phi}$(rà Juán)

b. $^{\phi}$($^{\sigma}$(ñá) $^{\omega}$(sókŏn)) $^{\phi}$(ñá Maria)

c. $^{\phi}$($^{\sigma}$(rí) $^{\omega}$(kárakono)) $^{\phi}$(tsĭna)

In (54), we see that the presence of the doubled pronoun satisfies the MinBin(φ, L) restriction in (50). Critically, it does so in an imperfect way. In comparing the prosodic structures in (54) with those in (48), we see that both satisfy (50), but the structures in (54) are worse in at least two respects. First, the leftmost prosodic phrases in (54) have a marked prosodic clitic (Selkirk 1995). Second, the two daughters of the leftmost prosodic phrase of (54) are not of the same prosodic category. This kind of unbalanced prosodic structure is disfavored by the Equal Sisters constraint of citemyrberg2010, myrberg2013.

This leads to an important prediction: if the MinBin(φ, L) restriction in (50) can be satisfied without the use of a prosodic clitic, the inherent markedness of that clitic will always block it. This seems to be what is going on in (48), where we saw that a leftmost prosodic phrase that is independently binary obviates the need for the doubled pronoun. Furthermore, this prosodic approach brings with it the advantage of

allowing us to keep sensitivity to syntactic adjuncts out of our φ-agreement and clitic doubling systems.

This approach has two further consequences that are worth noting, as their reflexes are present throughout the data in the rest of the dissertation. First, pronoun doubling in simple transitives is generally marked.

(55) a. ??**Rí** xàxi ntsibá'yi ndxùxi.
it.AML eat.PAST coyote chicken.hen
Intended: 'The coyote ate the hen.'

b. ☒ Xàxi ntsibá'yi ndxùxi.

c. ??**Rà**$_i$ tsyaa rà doktór$_i$ tútu.
he write.IRR he doctor paper
'The doctor will write a book.'

d. ☒ Tsyaa rà doktór tútú.

In (55), we see that pronoun doubling in simple transitive constructions is usually judged as degraded, while examples without pronoun doubling are consistently judged to be fully grammatical.

Critically, this dispreference towards pronoun doubling in transitives does not seem to be syntactic. First, as we will see amply throughout, pronoun doubling in simple transitives occurs frequently under the right pragmatic conditions. Second, if the prosodic environment changes, pronoun doubling in transitives becomes well-formed. For instance, if a monosyllabic word occurs to the left of the verb in a transitive, pronoun doubling again resurfaces as fully grammatical. In fact, failing to do pronoun doubling in these instances is strongly marked. Note that *na'à* 'early' in (56a) is monosyllabic and pronounced [na̱].

(56) a. **Rí** na'à xàxi ntsibá'yi ndxùxi.
it.AML early eat.PAST coyote chicken.hen
'The coyote ate the hen early.'

b. ?*☒ Na'à xàxi ntsibá'yi ndxùxi.

c. **Rà**$_i$ xa tsyaa rà doktór$_i$ tútú.
he PERF write.IRR he doctor paper
'The doctor will have written a book already.'

d. ?* ☒ Xa tsyaa rà doktór$_i$ tútú.

In (56) we observe that a monosyllabic word to the left of the transitive verb, either an adjunct like *na'à* 'early' in (56a) or an aspectual morpheme like the perfective marker *xa* in (56c), forces pronoun doubling to occur. Importantly, the generalization here does seem to be about prosodic size. For instance, if an adjunct that is a prosodic word or larger occurs in this position, pronoun doubling is blocked.

(57) a. *?**Rí** kamá xàxi ntsibá'yi ndxùxi.
it.AML quickly eat.PAST coyote chicken.hen
Intended: 'The coyote ate the hen quickly.'

b. ☒ Kamá xàxi ntsibá'yi ndxùxi.

c. ***Rà** nuhŭ iin tòò lo'o tsyaa rà doktór tútú.
he in a while little write.IRR he doctor paper
Intended: 'The doctor will write the book in a little while.'

d. ☒ Nuhŭ iin tòò lo'o tsyaa rà doktór tútú.

In comparing the prosodic size of the adjuncts in (57) to those in (56), we see that only monosyllabic adjuncts are compatible with pronoun doubling in transitive constructions.

Assuming that all of these are syntactically adjuncts, the sensitivity to prosodic size is rather unnerving. This is because in most syntactic theories since Zwicky and Pullum (1986), including minimalism (Chomsky 1993) and Distributed Morphology (Halle and Marantz 1993, 1994), as well as theories of syntax-to-prosody mapping, from Selkirk (1984) and Nespor and Vogel (1986) and the present-day Match theory (Selkirk 2009, 2011; Elfner 2012), prosodic information is unavailable to the syntax. Therefore,

it would be rather surprising to find a case of a syntactic process, either φ-agreement or clitic doubling, which is sensitive to this sort of fine-grained prosodic information.

Thankfully, the fully prosodic theory sketched above for intransitives can easily handle the distribution of pronoun doubling in transitives. This is because in simple transitives like (55), the finite verb seems to phrase with the subject. This is a common pattern in VSO languages like Connemara Irish (Elfner 2012), and the prosodic structures for (55) are provided in (58).[13]

(58) a. $^{\varphi}($ $^{\omega}($ xàxi $)$ $^{\omega}($ ntsibá'yi $)$ $)$ $^{\varphi}($ $^{\omega}($ ndxùxi $)$ $)$

b. $^{\varphi}($ $^{\omega}($ tsyaa =rà $)$ $^{\omega}($ doktór $)$ $)$ $^{\varphi}($ $^{\omega}($ tútú $)$ $)$

The prosodic structures in (58) are perfect relative to the constraints discussed informally above. In each, the leftmost prosodic phrase is binary, and the two daughters of that phrase are of the same prosodic category, satisfying Equal Sisters. Therefore, having a doubled pronoun would be marked, given the general markedness of prosodic clitics.

With this in mind, consider what the prosodic structures would be with a preverbal monosyllabic adjunct, both with a doubled pronoun and without. These structures are given in (59), based on (56a).

(59) a. $^{\varphi}($ $^{\sigma}($ na'à $)$ $^{\omega}($ xàxi $)$ $)$ $^{\varphi}($ $^{\omega}($ ntsibá'yi $)$ $^{\omega}($ ndxùxi $)$ $)$

b. $^{\varphi}($ $^{\omega}($ $^{\sigma}($ rí $)$ $^{\sigma}($ na'à $)$ $)$ $^{\omega}($ xàxi $)$ $)$ $^{\varphi}($ $^{\omega}($ ntsibá'yi $)$ $^{\omega}($ ndxùxi $)$ $)$

In (59a) without a doubled pronoun, we see that the monosyllabic adjunct would necessarily act as a prosodic clitic, as the strict word binarity restriction given in (50) would otherwise be violated. At the same time, if pronoun doubling occurs as in (59b), the two

[13]Note that the pronoun *rà* 'he' which is syntactically associated with *doktór* 'doctor' in (58b) prosodically phrases with the verb. This sort of prosody/syntax mismatch is common with clitics and will not be here. See the investigation of Kwak'wala in Anderson (2005) for a comprehensive discussion and review of this phenomenon.

clitics can lean on each other, forming a binary prosodic word of the sort that the language wants. Even better, this new prosodic word allows the leftmost prosodic phrase to both be binary and have daughters of the same prosodic category. This provides a fully prosodic understanding of why pronoun doubling is necessary with monosyllabic adjuncts.

Likewise, with ω or φ-sized adjuncts there will not be any prosodic need for the doubled pronoun, deriving the distribution in (57). These prosodic structures are given in (60).

(60) a. $^{\varphi}$($^{\omega}$(kamá) $^{\omega}$(xàxi)) $^{\varphi}$($^{\omega}$(ntsibá'yi) $^{\omega}$(ndxùxi))

b. $^{\varphi}$(nuhǔ iin tòò lo'o) $^{\varphi}$($^{\omega}$(tsyaa =rà) $^{\omega}$(doktór)) $^{\varphi}$($^{\omega}$(tútú))

In (60a), we see that the ω-sized adjunct serves the same prosodic function as the doubled pronoun and monosyllabic adjunct complex in (59b). This allows there to be a binary prosodic phrase at the leftmost edge of the sentence, without any prosodic clitics. Therefore, there is good prosodic reason not to have a doubled pronoun here, given the general markedness of prosodic clitics (Selkirk 1995). With this, we have derived the otherwise surprising adjunct size effect without violating the principle of Phonology-Free Syntax (Zwicky and Pullum 1986).

While the prosody of pronoun doubling is certainly an interesting issue, this dissertation will focus on the syntax of the phenomenon. Crucially, it is not the case that the prosody can fully derive the distribution of pronoun doubling. For instance, the requirement of prosodic doubling seen in the examples above can be violated if the syntactic and semantic requirements of pronoun doubling are not met. Compare the sentences in (61) to the ones above.

(61) a. * **Rí**$_i$ ntá'yi iin tsǐna$_i$, so xǐn ì ntsyâ rí
it.AML cry.PRES one dog but know.NEG.PRES I which it.AML
kúú rí.
COP.PRES it.AML

Intended: 'A dog is crying, but I don't know which one it is.'

b. ☒ Ntá'yi iin tsǐna, so xǐn ì ntsyâ rí kúú rí.

c. * **Nà**$_i$ na'à nì xǎ'ni iin nà$_i$ ña'a.
they early PAST break one them something
Intended: 'Someone early broke something.'

d. *☒ Na'à nì xǎ'ni iin nà ña'a.

In (61a-b), we see a simple intransitive predicate, *nta'yi* 'cry.' Given what we saw in (54), we would expect pronoun doubling to be required here. But, as we see in (61b), this is simply not the case, despite the fact that the resulting prosodic structure is unfavored in this language due to the otherwise strict constraint in (50).

(61c-d) demonstrates a similar point. Here we see a transitive with a monosyllabic adjunct *na'à* 'early.' Based on what we saw in (56), we predict that pronoun doubling should be required here. But, again, something else is going on here, leaving us necessarily with a prosodically imperfect structure in which we have a prosodic clitic, the monosyllabic adjunct, as discussed for (59a).

In this way, it seems clear that pronoun doubling has a crucial syntacticosemantic component, and its distribution cannot be determined solely by the phonology. The goal of this dissertation is to account for the syntactic component of pronoun doubling, with this brief discussion of the prosodic licensing sufficing for the remainder of the dissertation. In all of the data presented throughout, these prosodic licensing requirements have been controlled for and do not play a role in the patterns of of grammaticality presented below, though they will come up again in Chapter 5.

Chapter 2

The syntax of φ-agreement and clitic doubling

In this chapter, I will survey the literature on φ-agreement and clitic doubling. That said, distinguishing between φ-agreement and clitic doubling is frequently considered a difficult task. As Baker and Kramer (2018) state the problem, "It is an awkward fact that generative linguistics has had a hard time distinguishing reliably between pure agreement and clitic doubling." The issue, as far as I can tell, seems to be largely due to a lack of consensus on an analysis of clitic doubling.

I begin in §2.1 by examining φ-agreement, about which there is a much firmer consensus: Agree. Once we have examined the theoretical mechanics of Agree, I will consider a variety of ways in which the theory predicts that φ-agreement can be diagnosed empirically.

With this much in place, in §2.2 I contrast the behavior of clitic doubling systems with φ-agreement. This investigation yields a clear result, consistent with much earlier work on clitic doubling: clitic doubling systems show the syntactic characteristics of movement chains. With this, I propose that clitic doubling is cross-linguistically derived from a process called Move-and-Reduce, building on work by Alexiadou and

Anagnostopoulou (1997); Anagnostopoulou (2003); Harizanov (2014b); Kramer (2014), and an early manuscript of Baker and Kramer (2018) which I cite as Baker and Kramer (2016).

From this perspective, then, distinguishing between φ-agreement and clitic doubling is not necessarily difficult. The task is simply to look for the signature properties of a movement chain. If no movement chain can be diagnosed, then the system in question is best understood as φ-agreement. Likewise, if the properties of a movement chain can be uncovered, then the system in question is best analyzed as clitic doubling.

Naturally, some diagnostics to distinguish φ-agreement from clitic doubling are more reliable than others within this framework. (1) provides a ranked list of diagnostics that have been proposed in the literature.

(1) Ranking of diagnostics to distinguish clitic doubling from φ-agreement

		φ-agreement	Clitic doubling
High reliability:			
	a. Ability to affect variable binding?	**x**	✓
	b. Extraction from coordinate structures?	✓	**x**
Medium reliability:			
	c. Obeys the PIC?	✓	✓/**x**
	d. Presence of a default?	✓/**x**	**x**
Low reliability:			
	e. Sensitivity to specificity?	**x**/✓	**x**/✓
	f. Only one per clause?	**x**/✓	**x**/✓
	g. Semantic restrictions?	**x**/✓	**x**/✓
	h. Morphological similarity to D?	**x**/✓	**x**/✓

The diagnostics in (1a) and (1b) are highly reliable because they directly probe for the properties of a movement chain. Likewise, the diagnostics in (1c) and (1d) are less reliable because they attempt to diagnose movement chains in a more oblique way, and as such are prone to potential confounds. Finally, the diagnostics in (1e) through (1h) are of low reliability because they do not truly diagnose the core properties of movement chains.

2.1 The syntax of φ-agreement

It is no secret that some words never surface without the φ-features of a nominal in its clause. One well known example involves finite verbs in languages like Spanish.

(2) a. Yo$_i$ bat **-o**$_i$ los huevos con la batidora.
I beat -1SG.PRES the.PL eggs with the whisk
'I beat the eggs with the whisk.'

b. *Yo bat -$\emptyset$ los huevos con la batidora.

In (2) we see that finite T in Spanish must agree with a nominal in the clause. This pattern is referred to as φ-agreement. Consider the definition in (3) from Preminger (to appear).

(3) φ-Agreement: The phenomenon by which the φ-feature values (PERSON, NUMBER, GENDER/NOUN-CLASS) are transmitted from a noun phrase to a functional head (adapted from Preminger to appear, pg. 2).

In (3), we see that modern conceptions of φ-agreement treat the phenomenon as a process. In this process, the φ-features of a nominal projection are transferred to a functional head. Using the Spanish example in (2), the person and number features of *yo* 'I' are transmitted to finite T. This transmission gives rise to a configuration in which one set of φ-features, such as first person and singular in (2), occur twice: once in the nominal projection to which the φ-features are endemic, and once on the functional head which hosts the agreement.

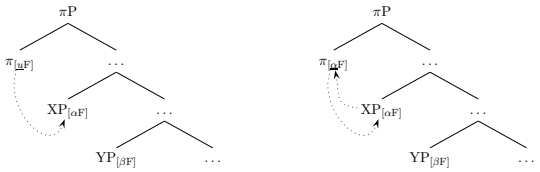
Throughout the history of modern syntactic thought on φ-agreement, various locality restrictions have been put forward to restrict what syntactic configurations are possible between the noun phrase and the functional head hosting the agreement, such as the subject and finite T in (2). Chomsky's AGREE operation has been particularly influential in these discussions. Since its original development in Chomsky (2000) and

Chomsky (2001), the notion of AGREE has undergone many reconceptualizations, but the core system largely remains intact. Here, I adopt a rather orthodox conception of AGREE.

In many ways, Chomskyian AGREE is, as it was intended to be, the minimal solution to the problem of φ-agreement. To see this, consider the general version of this system in provided in (4).

(4) Where XP is more local to π than YP

a. Search → b. Agree



The core of AGREE involves two syntactic objects: the probe and the goal. To return the example of Spanish from (2), finite T would be the probe while the subject would be the goal. The probe, π in (4), is a head which lexically lacks specification for some feature F. The basic idea of AGREE is that the probe must receive a value for F, just as finite T must be marked with φ-features.[1] The goal, XP in (4) and the subject in our Spanish example above, is a category which has a specification for the feature F that the probe needs. The purpose of the operation AGREE, then, is to assign the goal's value for F to the probe. This is necessarily done with syntactic mechanisms, given the syntactic character of φ-agreement.

As shown in (4), AGREE consists of three logically separable operations. The first of these is Search in (4a), during which the probe identifies the most local goal to the probe. It is this most local goal that the probe ultimately enters into an AGREE relation with. As a result of this AGREE relation, the probe's unspecified feature F

[1] I am intentionally vague about what 'must' means here. This is because for Chomsky, failure to assign a value for probe's unvalued feature renders the probe uninterpretable for either the phonology (or more specifically, the morphology) or the semantic component. Others, notably Preminger (2014), have argued that probes attempt to value their features, but the derivation does not crash if the syntactic derivation does not permit it. Our purposes here allow us to remain agnostic about these two options.

receives the feature value for F that its matching goal has. This final step of Valuation is shown in (4c), yielding two morphosyntactically distinct instantiations of the feature [αF], one on the probe and one on the goal. These multiple occurrences of the same feature are proposed to derive patterns of reduplication of φ-features.

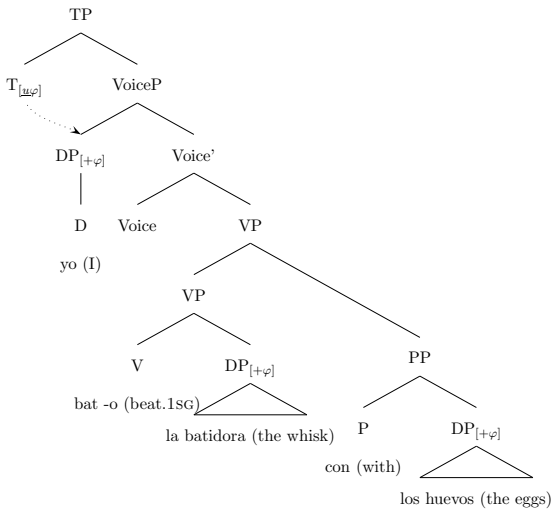
These specific properties of AGREE lead to a particular clustering of empirical characteristics of φ-agreement systems. Let us consider each in turn.

2.1.1 Property of φ-agreement #1: Locality determined syntactically

As discussed above, the nominal whose φ-features are replicated onto the probe is not randomly determined: it is strictly limited to the nominal which is structurally closest to the probe.

Let us return to the Spanish in (2). Let us assume the basic underlying structure for (2a) to be (5).

(5)



In (5), the subject *yo* 'I' is the highest nominal within the c-command domain of finite T. Therefore, Chomskyian AGREE requires that T enter into an AGREE relation with only this nominal, as it is the structurally closest. This is predicted to be the case

despite the presence of other nominals bearing φ-features that could satisfy T's $[u\varphi]$ because these other nominals are simply not as local as the subject.

This syntactic generalization holds water. For instance, if the structural configurations among the same nominals with the same φ-features are altered, it is still the case that the highest nominal, the subject, must be agreed with. This can be seen most clearly if (2) is passivized, as in (6).

(6) a. Los huevos$_i$ fu **-eron**$_i$ batidos con la batidora por mi.
the eggs be -3PL.PAST beaten with the whisk by me
'The eggs were beaten with the whisk by me.'

b. *Los huevos fu -e$_i$ batidos con la batidora$_i$ por mi.
the eggs be -3SG.PAST beaten with the whisk by me
Intended: 'The eggs were beaten with the whisk by me.'

c. *Los huevos fu -i$_i$ batidos con la batidora por mi$_i$.
the eggs be -1SG.PAST beaten with the whisk by me
Intended: 'The eggs were beaten with the whisk by me.'

In (6), just as in (2), we see that there is a syntactic generalization about the nominal which agrees with the verb. In both cases, it is the structurally highest nominal whose φ-features are exponed on finite T.

Additionally, it is not the case that AGREE must targets subjects. Rather, the crucial generalization is that only the structurally highest nominal in the probe's c-command domain must be agreed with. The classic example of this comes from Chomsky's (2001) investigation of expletive-*there* constructions in English and Icelandic. Let us consider the English data in (7), adapted from an inspired by Chomsky (2001).

(7) a. There are$_i$ expected to be caught [many fish]$_i$.

b. There is$_i$ expected to be caught [a fish]$_i$.

c. There are$_i$ expected to be [many people]$_i$ eating [their dinner] after 7.

d. *There is$_i$ expected to be [many fish] eating [garbage]$_i$.

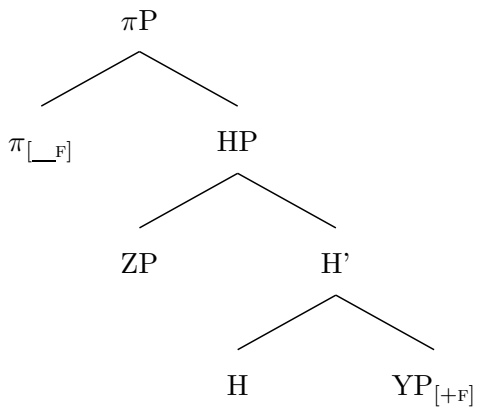
In (7a-b), we see that the matrix verb enters into a φ-agreement relation with the deeply embedded object, either 'many fish' in (7a), or 'a fish' in (7b). In (7c), we see that the matrix verb must agree with the structurally highest DP, in this case 'many people.' Furthermore, we know that 'many people' is structurally higher than 'their dinner' based on the availability of the bound-variable reading of 'their.'

This leads to a clear expectation about φ-agreement systems cross-linguistically: only the structurally highest available nominal can be targeted.

2.1.2 Property of φ-agreement #2: The locality of φ-agreement is phase based

Within a phasal approach to syntax, Chomsky (2001) proposes that syntactic operations such as AGREE are sensitive to units of structure referred to as phases. Consider the schema in (8).

(8) In which H is a phase head



The basic idea behind phase theory is that syntactic structure is built incrementally in chunks referred to as phases. A phase is built when a particular head in a language, such as H in (8), it induces a Spell-Out domain in its complement. In (8), this would mean that YP would be processed by the morphological and phonological systems. Practically, this means YP would exit the syntactic derivation. Naturally, this means

that YP would be be inaccessible to any further syntactic operations. This is codified in (9) from Chomsky (2001).

(9) The domain of [a strong phase head] H is not accessible to operations outside of HP; only H and its edge are accessible to such operations (Chomsky 2001, pg. 13)

With this in mind, let us return to (8). Here, we have a probe π which needs a value for a feature F. As such, it will examine its c-command domain to find a a constituent with a value for F. But in this context, H is a phase head, and as such, only the head H itself will be available. This effect is frequently referred to as the Phase Impenetrability Condition, or the PIC.

Assuming, along with Chomsky (2001), that AGREE is one of the outside syntactic operations referenced in (9), we predict that φ-agreement should respect the PIC. To see this, let us examine agreement into two possible phases cross-linguistically: finite CP and PP.[2]

2.1.2.1 No φ-agreement into finite CP

Finite CP is the best established candidate for phasehood cross-linguistically. At the same time, there are reported cases of φ-agreement across CP boundaries. This phenomenon is commonly referred to as Long Distance Agreement, or LDA. LDA is a φ-agreement relation established between a probe and a goal that are not contained within the same clause.

Because Long Distance Agreement crosses clause boundaries, it has the potential to be φ-agreement into a phase, namely CP. If it turns out to be the case that Long Distance Agreement can form dependencies into the Spell-Out domain of finite CP, it

[2]While the phasal status of P is controversial (see Bruening 2014), the reason why these categories were chosen, along with DP in §2.2.3.2, is that these seem to be the phasal categories in SMP Mixtec. See §5.1.

would be a strong challenge to the legitimacy of the PIC.

Long Distance Agreement (LDA) has been documented in several languages such as Tsez (Polinsky and Potsdam 2001), but it is best documented in Indo-Aryan languages, particularly Hindi. The basic LDA configuration in Hindi is demonstrated in (10). Throughout the discussion, bear in mind that overtly case-marked nominals in Hindi, such as nominals marked with the ergative marked *ne*, are categorically unavailable for φ-agreement (see Mohanan 1995 and Mahajan 1997 especially).

(10) Long Distance Agreement in Hindi (Boeckx 2004, Bhatt 2005)

a. Vivek -ne [kitaab$_i$ paṛh -**nii**$_i$] chaah -**ii**$_i$.
V. -ERG book.FEM read -INF.FEM want -PERF.FEM
'Vivek wanted to read the book.'

b. *Shahrukh [**ṭehnii**$_i$ kaaṭ -**nii**$_i$] chaah -**tii**$_i$ **thii**$_i$.
S.MASC branch.FEM cut -INF.FEM want -PERF.FEM be.PAST.FEM
Intended: 'Shahrukh wants to cut the branch.'

In (10a), the only matrix nominal is unavailable because it is marked with ergative case (Mohanan 1995, Mahajan 1997, Bhatt 2005). In this configuration, the matrix verb may agree with the embedded object *ṭehnii* 'branch.' Importantly, in (10b), the matrix subject *Shahrukh* is not overtly case-marked. Therefore, the matrix verb can, and in fact must, agree with it. In comparing (10a) to (10b) , we see that LDA is not available when the higher subject is available for agreement.

Crucial for our purposes, LDA in Hindi does not seem to involve φ-agreement across a phase boundary. First, the phasal properties of finite CP in Hindi have been well documented. See Manetta (2006, 2010) for thorough argumentation. Second, LDA in Hindi is impossible into a finite embedded clause.

(11) No LDA in Hindi into finite clauses (Boeckx 2010)

a. Firoz -ne soch **-aa** ki [Mona ghazal gaa -tii
F. -ERG think -PERF.3MASC that M. ghazal.FEM sing -HAB.FEM

hai].
be.PRES
'Firoz thought that Mona sings ghazals.'

b. *Firoz -ne soch **-ii**$_i$ ki [Mona$_i$ ghazal gaa -tii
F. -ERG think -PERF.FEM that M. ghazal.FEM sing -HAB.FEM
hai].
be.PRES
Intended: 'Firoz thought that Mona sings ghazals.'

In (11a), there are no nominals in the matrix clause for the matrix verb to agree with. This is a similar situation to the LDA case we saw in (10a). Therefore, if LDA were possible into a finite embedded clause, then we would expect the matrix verb to obligatorily agree with one of the embedded arguments, presumably the feminine subject 'Mona.' But the ungrammaticality of (11b) makes it clear that this sort of agreement into a finite clause is impossible. This derives the presence of default singular third singular masculine features on the verb in (11a).

These observations suggest that LDA in Hindi does not provide a counterexample to the PIC, and that the infinitival clauses in (10) are subphasal. Aside from LDA, I am aware of one other case that has been reported in the literature that shows apparent φ-agreement into a finite clause. This comes from Chukchi[3], a Chukotko-Kamchatkan language of eastern Siberia with about 10,000 speakers, as reported in Bošković (2007). Bošković argues that φ-agreement may occur into finite CPs in Chukchi in violation of the PIC on the basis of one sentence. This sentence is demonstrated in (12), with Bošković's glosses.

(12) ənan qəɬɣiɬu ɬəŋərkə -nin **-et** [iŋqun ∅- rətəmŋəv -nen -at **qora** **-t**].
he regrets -3 -PL that 3SG- lost -3 -PL reindeer -PL
'He regrets that he lost the reindeers.' (Bošković 2007, pg. 613)

[3]Bošković (2007) also proposes that Blackfoot provides an example of φ-agreement into finite CP, but see Legate (2005) for an alternative analysis of the Blackfoot data. I am aware of no alternative treatments of the Chukchi in (13-16).

In (12), Bošković claims that the matrix verb *qəɬɣiɬu* 'regret' agrees with the embedded object *qorat* 'reindeers.' From this, Bošković concludes that φ-agreement may cross finite clause boundaries.

That said, there are serious and troubling questions about the provenance of this example. First, (12) is the only Chukchi sentence that Bošković (2007) provides to support his conclusion that φ-agreement may occur into a finite clause. At the same time, as Bošković himself notes in a footnote[4], there is disagreement over the best translation of this sentence. Bobaljik (2006), working from the original Russian sources of this sentence, provides the alternate translation in (13).

(13) ənan qəɬɣiɬu łəŋ -ə -rkən -in -et [iŋqun rətəmŋəv -nen
he.ERG sorry/pity/regret AUX -EPEN -PAST -3>3 -PL because lose -3>3
-at **qora -t**].
-PL reindeer -PL
'He feels sorry (for them), that he lost (them), the reindeer.'

These differences in translation matter because Bošković's translation involves φ-agreement into a finite clause while Bobaljik's does not. If one accepts Bošković's translation, then one must also accept that φ-agreement must be able to reach into phases, whereas if one accepts Bobaljik's translation, then one does not necessarily have to accept this conclusion.

The tension centers around the best translation of the Chukchi word *inqun*. For Bošković, *inqun* corresponds to English 'that.' As such, he treats it as a complementizer which introduces a finite embedded clause. In contrast, Bobaljik translates *inqun* as 'because,' which would make the clause that follows it an adjunct clause.

The most complete Chukchi grammar in English, Dunn (1999), however, strongly suggests that both translations are incorrect. First, the presentation of this sentence as given by Bošković (2007) in (12) is surely misleading. This is because, as Dunn (1999) makes clear, Chukchi lacks finite embedding equivalent to English (Dunn 1999, pg. 84).

[4]Footnote 41, pg. 613

He reports that clauses may be semantically subordinated, "although there is not a syntactic distinction between conjunctive [finite] subordindation and coordination" (Dunn 1999, pg. 88). Dunn even reports that "Chukchi does not have any mechanisms for marking indirect speech: all quoted speech is direct" (Dunn 1999, pg. 91). In fact, none of the multiclausal sentences that Dunn provides are translated into English using finite subordination.

Therefore, the presentation in Bošković (2007) is certainly misleading: Chukchi cannot allow φ-agreement into a finite embedded clause because the language does not have any finite embedded clauses. That said, the translation provided by Bobaljik (2006) cannot be totally faithful, either. This is because the only word translated as 'because' in Dunn (1999) is *qeluq*, not *inqun*.

(14) ŋeekkeqej =ʔm **qeluq** =ʔm taŋ- ə- nm -ə -nen **qeluq**
girl.DIM.3SG.ABS =EMPH because =EMPH INTS- EPEN kill -EPEN -3>3 because
=ʔm ʔaqa- n- malaw -at -ə -ŋ.
=EMPH IMPOSS- CS- recover -CS -EPEN -VBASE
'The girl though he killed alas, because [she] was impossible to cure.' (Dunn 1999, pg. 247)

At the same time, Bobaljik (2006) is correct in translating the 'embedded' clause of this sentence as an adjunct clause. I could find only one instance of *inqun* in Dunn (1999), and it is translated as 'so that' introducing an adjunct clause.

(15) ənqorə neme, neme ənkə jawren -a =ʔm **inqun** peecwaɣ -jonr -at
then also also here next.year -ADV =EMPH so.that spring -wean -TH
-ə -k =ʔm əmə, neme qol ŋelwəl na- n- tomɣ
-EPEN- SEQ =EMPH also also QUANT.3SG.ABS herd.3SG.ABS 3PL- CS- exist
-aw -ə -n.
-CS -EPEN -3SG
'Then again, again there the next year after the spring weaning too, again they made another herd.' (Dunn 1999, pg. 132)

From this investigation, one thing seems clear: very little can be reliably con-

cluded about Chukchi, but there is little reason to believe that it has φ-agreement into finite embedded clauses.

Given the evidence presently available, then, so-called "Long-Distance Agreement" into finite embedded clauses, i.e., phases, does not seem to exist. Therefore, we can uphold the PIC with respect to finite embedded CPs.

2.1.2.2 No φ-agreement into PP

The inability of φ-agreement to target nominals which remain in situ within PP is taken for granted in most minimalist literature. This seems to be a holdover from older Government and Binding theories of φ-agreement in which P formed a barrier (Chomsky, 1986). But while rarely discussed in the minimalist literature[5], the inability to agree with a nominal embedded within a PP seems to fall out of P is taken to be a phase head (Abels, 2003). Consider the English pattern in (16).

(16) a. There seems$_i$ [to many people] to be [a cat$_i$] around.

b. There seem$_i$ [to many people] to be [several cats$_i$] around.

c. *There seem$_i$ [to many people$_i$] to be [a cat around].

In (16) we see that φ-agreement obligatorily looks past the PP experiencer in English, and must agree with the lower nominal. This receives a straightforward explanation of P is a phase head, making its complement inaccessible to φ-agreement outside of PP. See Deal (2015b) for a similar argument, though see Bruening (2014).

Another domain in which this same conclusion has been reached comes from patterns of φ-agreement in languages like Hindi. In Hindi, only unmarked arguments may agree with the verb, as mentioned briefly in §2.1.2.1.

(17) Hindi agreement (Mohanan 1995)

[5]Though see Deal (2015a).

a. Niinaa$_i$ baalak -ko uṭ^{h}aaegii$_i$.
R. boy -ACC lift.FUT.FEM.SG
'Nina will lift up the boy.'

b. Ravii -ne roṭii$_i$ k^{h}aayii$_i$.
R. -ERG bread eat.PERF.FEM.SG
'Ravi ate bread.'

c. Niinaa -ne baalikaa -ko uṭ^{h}aayaa.
Nina -ERG girl -ACC lift.PERF.MASC.SG
'Nina lifted up the girl.'

In (17a), we see that when the object is marked accusative but the subject is unmarked, the verb must agree with the unmarked subject. Likewise in (17b), when the subject is marked ergative but the object is unmarked, then the verb must agree with the unmarked object. In (17c) where both the subject and the object are case-marked, the verb cannot agree with either the subject or the object, and surfaces in a default masculine singular form.

Mohanan (1995), Mahajan (1997), and Polinsky (2016) suggest that this pattern is related to case suffixes, particularly the ergative suffix *-ne*, being of category P. Reason to think this, apart from historical reasons discussed in Polinsky (2016), comes from coordinated ergatives, demonstrated in (18).

(18) a. [laRki aur laRkaa] -ne
boy and girl -ERG

b. *laRki -ne aur laRkaa -ne
boy -ERG and girl -ERG

In (18) we see that the ergative marked *-ne* must adjoin to the entire coordinate structure. Polinsky (2016) suggests that this is the behavior we expect a postposition to demonstrate, rather than a case affix. Furthermore, if Polinsky is on the right track, then the inability of φ-agreement to target nominals with case suffixes becomes a straightforward consequence of the PIC without need for further investigation.

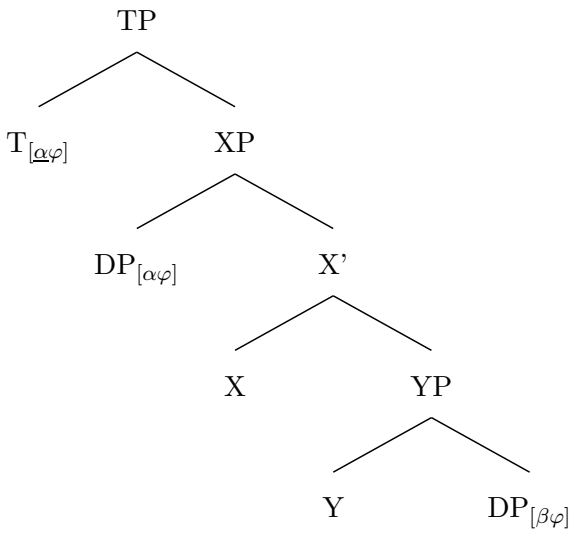
If this discussion is on the right track, then the impossibility of φ-agreement into PPs falls out directly from their their status as phase heads, so long as the PIC is in effect. This supports the conclusion that φ-agreement is indeed sensitive to the PIC.

This discussions leads to a cross-linguistic expectation: if a process that doubles the φ-features of a nominal within a clause is derived by φ-agreement, then we expect that it will obey the PIC. With this conclusion in place, let us consider further properties of φ-agreement systems cross-linguistically.

2.1.3 Property of φ-agreement #3: Inability to affect variable binding

Chomskyian AGREE has a further important consequence. To see this, let us consider the linguistic object that is produced by AGREE, namely a head π with some set of features. Consider (19), which demonstrates the end state after an AGREE has taken place.

(19)



On the now inactive probe T, we see a φ-feature bundle that is identical to the φ-features of the goal, in this instance the highest DP in its c-command domain.

This means, of course, that the end state of AGREE is the inactive probe of the same category, in this example T, that now has an interpretable φ-feature bundle.

Critically, these features are not the same thing as a full copy of the goal. In other words, AGREE does not turn a φ-probe into a DP by virtue of entering into an AGREE relation with something of category D.

Returning to (19), we predict that a valued φ-probe will not behave like full DP. One important facet of this defectiveness is the failure of valued φ-probes to alter binding relations. As is well-known, full phrasal movement can make possible variable-binding and anaphor-binding configurations that would be impossible in its absence. See, for instance, den Dikken (1995) and Rezac (2010) for discussion within this context. To see this, first consider the sentences in (20).

(20) a. * It seems to her$_i$ mother that every girl$_i$ is a genius.

b. Every girl$_i$ seems to her$_i$ mother $\_\_{}_i$ to be a genius.

In (20a), we see that in the absence of phrasal movement, the quantifier 'every girl' within the embedded clause may not bind the variable 'her mother' in the matrix clause. In (20b), though, we see that full phrasal movement does allow for this variable binding relation to occur.

Crucially, φ-agreement is not enough to yield new variable binding relations in the absence of full phrasal movement. Consider (21), which was inspired by similar sentences in den Dikken (1995).

(21) a. * There seem -$\emptyset_i$ to their$_i$ mother to be several girls$_i$ learning rocket science.

b. Several girls$_i$ seem -$\emptyset_i$ to their$_i$ mother to be $\_\_{}_i$ learning rocket science.

In (21a), the verb agrees with the quantified DP 'their mother,' yielding a copy of this DP's φ-features on T. Now, is φ-agreement behaved like full phrasal movement, we would expect this configuration to also yield the same variable binding possibilities that

we saw in comparing (20b) to (20a). But as the ungrammaticality of (21a) makes clear, this is not the case. Rather, full phrasal movement is still obligatory in order to yield these new variable binding possibilities. This is shown in (21b).

Similar data are also replicable with anaphor-binding. Consider (22), which demonstrates the same conclusion as (20) but with variable-binding substituted for anaphor-binding.

(22) a. * It seems to each other$_i$ that the boys$_i$ are kind.

b. The boys$_i$ seem to each other$_i$ __$_i$ to be kind.

In (22), we see that without movement of the DP 'the boys' to Spec,TP of the matrix clause, the reciprocal anaphor 'each other' may not be bound. While this conclusion is certainly not controversial since Chomsky (1980, 1981), it is important in the present context because it contrasts with φ-agreement. Consider the paradigm in (23), from den Dikken (1995).

(23) (den Dikken, 1995)

a. * There seem -$\emptyset_i$ to each other$_i$ to be some applicants$_i$ eligible for the job.

b. Some applicants$_i$ seem -$\emptyset_i$ to each other$_i$ to be __$_i$ eligible for the job.

In (23a), we see that, just as with variable-binding in (21a), φ-agreement is not enough to yield new anaphor-binding relations. Rather, full phrasal movement is required here as well.

From this, we see that φ-agreement may not yield new variable-binding or anaphor-binding configurations like full phrasal movement. This is a natural and welcome consequence within the present context. Because φ-agreement does not involve XP-movement, we strongly predict that φ-agreement should not be able to do the same things that XP-movement can. This is precisely what we are seeing here.

2.1.4 Property of φ-agreement #4: Agreement available into coordinate structures

In the last subsection, we saw that φ-agreement is unable to do things that XP-movement can. This is a natural consequence of φ-agreement being derived through mechanisms that do not involve XP-movement, such as in an AGREE framework. But this approach yields a parallel conclusion: Because φ-agreement is not derived through XP-movement, φ-agreement should be able to do things that XP-movement cannot.

In this section, I will show that this is the case by examining a set of well-known conditions on XP-movement referred to as 'island constraints' since Ross (1967). Consider one island context in particular: the Coordinate Structure Constraint, shown in (24).

(24) a. [Maria and José]$_i$ were awarded $\_\__i$ one million dollars.

b. * Maria$_i$ was awarded [$\_\__i$ and José] one million dollars.

In (24a), we see that an entire coordinate structure may undergo movement to Spec,TP, but in (24b) we see that only one conjunct, in this case the leftmost conjunct, cannot be extracted to the exclusion of the rest of the coordinate structure. This effect is referred to as the Coordinate Structure Constraint.

Now, the Coordinate Structure Constraint, like all island constraints, are conditions on XP-movement. This means that φ-agreement should not be subject to these constraints because, as we have seen above, φ-agreement is not derived through XP movement. This turns out to be the case.

A robustly attested phenomenon cross-linguistically is Closest Conjunct Agreement, or CCA. In CCA, a φ-probe expones the φ-features of the conjunct that is linearly closest to it. This is shown in (25) for a variety of languages, and see also McCloskey and Hale (1984), McCloskey (1986), Munn (1993, 1999), Aoun et al. (1994), Soltan (2006), Marušič et al. (2007), Benmamoun et al. (2009), Boškavić (2009), and Bhatt

and Walkow (2013), among many others.

(25) Closest COnjunct Agreement in Moroccan Arabic, Hindi, and Tsez (Benmamoun et al. 2009)

a. Ža$_i$ [ʕomar$_i$ w Kariim]. (Moroccan Arabic CCA)
came.3SG.MASC O. and K.
'Omar and Karim came.'

b. Main -ne [ek chaataa aur ek saaRii$_i$] khariidii$_i$. (Hindi CCA)
I -ERG an umbrella.M and a saree.F buy.F
'I bought an umbrella and a saree.'

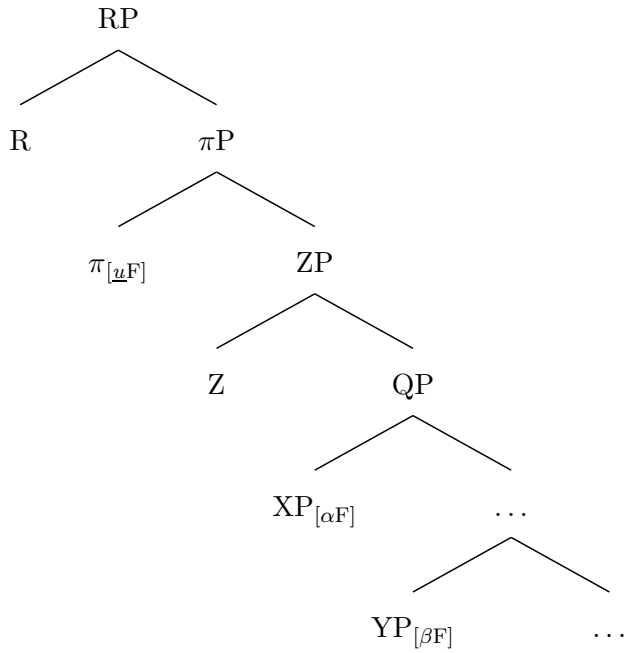
c. [Kid -no užI$_i$ -n] Ø$_i$- ik'is. (Tsez CCA)
girl.II -and boy.I -and I.SG- went
'A boy and a girl went.'

In each of the languages in (25), we see that may target the φ-features of only one conjunct. This phenomenon would be rather striking if φ-agreement were derived through XP-movement because the sentences in (25) would be predicted to be ungrammatical because they would violate the Coordinate Structure Constraint. But from the perspective of AGREE, the existence of Closest Conjunct Agreement is not as troubling. This is because AGREE does not involve XP-movement, and as such, does not predict φ-agreement to show the same island constraints as XP-movement.

2.1.5 Property of φ-agreement #5: Presence of a default

Consider the AGREE configuration in (26).

(26) Where R and Z are strong phase heads:



In (26), R and Z are strong phase heads. In this configuration, the PIC will block π from entering into an AGREE relation with any of the potential goals in the lower phase.

This results in a confound. The probe π cannot surface without a value for [$\underline{u}$F], because unvalued features in this system are uninterpretable at either the PF or LF interface (2000; 2001). This should result in an unpronounceable and ultimately ungrammatical sentence. At the same time, the system's metaphorical hands are tied, because there are simply no possible goals within the same phase as π.

Perhaps surprisingly, configurations such as (26) are permitted to occur in natural languages. This leads to an interesting empirical question: what do languages do to ultimately render structures like (26) grammatical? Preminger (2009, 2014) considers several such systems and examines the ways in which these "derivational timebombs," probes that cannot receive a value for an uninterpretable feature, are defused. Consider the diagnostic in (27) from Preminger (2009).

(27) **Preminger's (2009) diagnostic**: "Given a scenario where the relation $\mathcal{R}$ between an agreement-morpheme $\mathcal{M}$ and the corresponding full noun-phrase

$\mathcal{F}$ is broken - but the result is still a grammatical utterance - the proposed diagnostic supplies a conclusion about $\mathcal{R}$ as follows:

a. $\mathcal{M}$ shows up with default φ-features (rather than those of $\mathcal{F}$) $\Longrightarrow$ $\mathcal{R}$ is AGREE

b. $\mathcal{M}$ disappears entirely $\Longrightarrow$ $\mathcal{R}$ is clitic-doubling

For the time being, let us focus on (27a). Preminger proposes that the way languages resolve configurations like (26) is to insert a set of dummy φ-features. These "default" φ-features are not derived from a true AGREE relation, but are rather a kind of morphophonological EPP that are used to produce a pronounceable object, or an object that is interpretable at the PF interface.

To see how this works empirically, consider how Spanish finite-T agreement behaves in this sort of situation.[6]

(28) a. Parec **-e** [que los hombres están dormiendo].
seem -3SG.PRES that the men are sleeping
'It seems that the men are sleeping.'

b. *Parec -∅ [que los hombres están dormiendo].

c. *Parec **-en**$_i$ [que los hombres$_i$ están dormiendo].
seem -3PL.PRES that the men are sleeping
Intended: 'It seem that the men are sleeping.'

[6]See Alexiadou and Anagnostopoulou 1998 for arguments that there is no silent expletive in sentences like (28a). At the same time, Spanish dialects are subject to cross-linguistic variation on this point. Most European and Latin American varieties behave like (28). Interestingly, many Caribbean varieties, particularly Dominican Spanish, require an overt expletive in this sort of construction, like English.

(1) Data from Toribio (2000)

a. **Ello** llegan guaguas hasta allá.
there arrive.3PL.PRES buses up.to there
'There arrive buses there.'

b. **Ello** había mucha gente en *lay-a-way.*
there be.3PL.IMPERF many people on stand.by
'There were a lot of people on stand-by.'

In (28b-c), we see the pattern that we have come to expect. In (28b), failure to mark finite T with φ-features results in ungrammaticality, just as we saw in §2.1. Likewise, (28c) shows attempted φ-agreement between finite T and a nominal embedded within a finite CP phase. As we saw in §2.1.2.1, this is forbidden by the PIC.

Now let us consider (28a). Here finite T surfaces with third person singular φ-features. Importantly, there is no possible goal within its phase that could be the source of these φ-features. Therefore, the apparent third singular agreement is, instead, a set of default φ-features.[7] Due to the presence of these default features, the diagnostic in (27) correctly identifies this process as φ-agreement.

At the same time, this type of morphophonological diagnostic is inherently unreliable because it makes the critical assumption that default exponence will always be overt. Of course, φ-agreement within a paradigm need not be morphophonologically overt, as English present tense agreement makes clear.

(29) English present tense verb paradigm

	Singular	Plural
1st person	run -∅	run -∅
2nd person	run -∅	run -∅
3rd person	run **-s**	run -∅

In (29), we see that most of the cells in the English agreement paradigm are morphophonologically null. Given that the possibility of null exponence of agreement is robustly attested, there is no *a priori* requirement that default features be overt. Therefore, the diagnostic in (27) cannot be wholly reliable. The accurate conclusions of this diagnostic are given in (30).

(30) **Preminger's (2009) diagnostic**: "Given a scenario where the relation $\mathcal{R}$ between an agreement-morpheme $\mathcal{M}$ and the corresponding full noun-phrase

[7] One alternative analysis is that T agrees with the entire embedded CP, as proposed for Zulu by Halpert (2012). For the expository purposes of demonstrating Preminger's (2009) proposal, I will not consider this possibility.

$\mathcal{F}$ is broken - but the result is still a grammatical utterance - the proposed diagnostic supplies a conclusion about $\mathcal{R}$ as follows:

a. $\mathcal{M}$ shows up with default φ-features (rather than those of $\mathcal{F}$) $\Longrightarrow$ $\mathcal{R}$ is AGREE

b. $\mathcal{M}$ disappears entirely $\Longrightarrow$ $\mathcal{R}$ is clitic-doubling **or φ-agreement**.

Therefore, this diagnostic can only yield a reliable result if default features occur. In the absence of morphophonologically overt default features, we are left with a system that is compatible with φ-agreement with idiosyncratically silent default features (see Baker 2012), or with a clitic doubling system in which there are no default features.

2.1.6 Property of φ-agreement #6: Insensitivity to specificity or referentiality

Let us consider the structure of a basic φ-probe π in an AGREE system. π, by definition, lexically lacks specification for φ-features. This means that the only thing π needs is φ-features. Therefore, π will obligatorily interact with any nominal that bears φ-features, regardless of any other features of the nominal, such as specificity or referentiality.

Indeed, this insensitivity to specificity or referentiality is a core property of φ-agreement systems. Consider both the English and Spanish verb agreement systems in (31) and (32).

(31) English subject-verb agreement is insensitive to specificity or referentiality

a. No one$_i$ was$_i$ beating the eggs with the whisk.

b. No people$_i$ were$_i$ beating the eggs with the whisk.

c. Someone$_i$ was$_i$ beating the eggs, but I don't know who.

d. Some people$_i$ were$_i$ beating the eggs, but I don't know who.

e. That person$_i$ over there is$_i$ beating the eggs.

f. Those people$_i$ over there are$_i$ beating the eggs.

(32) Spanish subject-verb agreement is insensitive to specificity or referentiality

a. Nadie$_i$ bat **-ió**$_i$ los huevos con la batidora.
nobody beat -3SG.PAST the eggs with the whisk
'Nobody beat the eggs with the whisk.'

b. Cualquier persona$_i$ pued **-e**$_i$ batir los huevos con la batidora.
whichever person can -3SG.PRES beat the eggs with the whisk
'Any which person can beat the eggs with the whisk.'

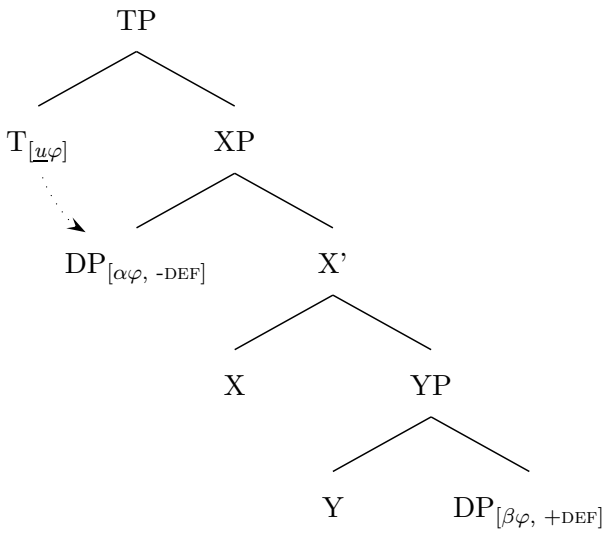
c. Unas personas$_i$ bat **-ieron**$_i$ los huevos con la batidora, pero no
some people beat -3PL.PAST the eggs with the whisk but not
sé quién.
know.1SG.PRES who
'Some people beat the eggs with a whisk, but I don't know who.'

d. Estas personas$_i$ que están sentadas en la parada de autobús
these people who are.3PL.PRES sitting at the bus stop
bat **-ieron**$_i$ los huevos.
beat -3PL.PAST the eggs
'These people who are sitting at the bus stop beat the eggs.'

(31) and (32) show a variety of nominals in English and Spanish respectively with different referential and definiteness properties. Crucially, all of them agree with the finite verb in the same way. In (31a-d) and (32a-c), we see several indefinite and non-referential subjects, like 'no one,' 'no people,' *nadie* 'nobody' and *cualquier persona* 'any which person.' As expected, all of these agree with finite T in the same way as the definite and specific subject in (31e-f) and (32d).

Again, this distribution of φ-agreement is exactly what Chomsykian AGREE predicts, assuming that finite T in both Spanish and English only bears uninterpretable φ-features. Consider the schematization in (33), which is compatible for both Spanish and English.

(33)



In (33), T bears uninterpretable φ-features which must be valued. Given AGREE, it will get feature values from the highest constituent in its c-command domain that bears interpretable φ-features. Importantly, the probe will not interact with any other features of that constituent, such as definiteness, formalized pre-theoretically as the [±DEF] feature in (33). So long as the highest DP can assign a value to T's [$\underline{u}\varphi$] feature, its value for [±DEF], or any other feature, will be immaterial.

While this diagnostic has a consistent theory-internal logic, by the end of this dissertation we will see that it cannot be reliable. This is because, as I will discuss in extensive detail in chapter 5, some φ-agreement systems cross-linguistically target topics. At the same time, languages place semantic restictions on which types of referential expressions can be topics. See Rizzi (1997), among many others. As such, we cannot take semantic restrictions as indicative of φ-agreement.

2.1.7 Property of φ-agreement #7: Only one per clause

Consider a diagnostic developed in Baker (2012). In all of the cases we have investigated so far, the probe has one single unvalued [$\underline{u}\varphi$] feature. This means that

the probe needs to enter into only one AGREE relation with only one nominal in order to value its unvalued feature. Once its unvalued feature receives a value under AGREE relation, there is simply no way that further φ-agreement with further nominals could be marked on the probe. See Baker (2012) for further discussion.

This leads to a strong empirical expectation: in φ-agreement systems, the number of morphological exponents should match the number of φ-probes. For instance, in a language like Spanish in which there is only one φ-probe, specifically finite T, we should only ever see one morphological exponent of φ-agreement. As we have seen in the above examples, this is indeed the case. (34) presents these data more explicitly with Spanish finite T-agreement.

(34) a. Yo$_i$ bat **-o**$_i$ los huevos con la batidora.
I beat -1SG.PRES the.PL eggs with the whisk
'I beat the eggs with the whisk.'

b. *Yo$_i$ bat -o$_i$ -en$_j$ los huevos$_j$ con la batidora.
I beat -1SG.PRES -3PL.PRES the eggs with the whisk
Intended: 'I beat the eggs with the whisk.'

c. *Yo$_i$ bat -o$_i$ -e$_j$ los huevos con la batidora$_j$.
I beat -1SG.PRES -3SG.PRES the eggs with the whisk
Intended: 'I beat the eggs with the whisk.'

In (34), we clearly see that only one DP may share its φ-features with the T probe.

At the same time, this diagnostic cannot be the whole story. This is because the morphosyntax and morphophonology of φ-agreement systems often yields patterns in which the φ-features of a single DP are replicated throughout the entire clause. One particularly clear case of this is Ibibio (Baker and Willie, 2010), who show that φ-agreement associated with finite T may occur both on auxiliaries in T and on lexical verbs.

(35) Ibibio agreement (Baker and Willie, 2010)

a. **N-** sʌk **n-** yem ebot odo.
1SG- AUX 1SG- seek goat the
'I am looking for the goat.'

b. ɔmmɔː **e-** mana **e-** nam.
they 3PL- do.again 3PL- do
'They are doing it again.'

Additionally, a common pattern cross-linguistically involves a single probe agreeing with multiple goals. See Béjar (2003) especially. Therefore, we can conclude little from the obligatory presence of only one agreement marker, as φ-agreement systems cross-linguistically do not need to show this restriction.

2.1.8 Summary of φ-agreement

In this section, we examined several key empirical properties of φ-agreement systems. From this investigation, we identified the following properties of φ-agreement systems, assuming Chomskyian AGREE.

(36)

	φ-agreement
Obligatorily target highest nominal?	✓
Obeys the PIC?	✓
Can affect variable binding?	**x**
Closest Conjunct effects?	✓
Presence of a default?	✓/**x**
Sensitivity to specificity or referentiality?	✓/**x**
One per clause?	✓/**x**

In (36), we see that some diagnostics are more reliable than others. As discussed in their respective sections, these are for different reasons. For example, the diagnostic of the presence of a default was shown in §2.1.5 to only be reliable if morphophonologically overt default φ-features were present. Others, like the sensitivity to specificity or the presence of only one marker per clause seem to be subject to cross-linguistic constraints on the specific type of φ-agreement involved.

That said, some diagnostics have emerged as particularly reliable. Chief among these are the inability of φ-agreement to yield new variable-binding or anaphor-binding possibilities, as well as the availability of Closest Conjunct affects, or more broadly, the lack of sensitivity to island constraints. With this in place, let us begin to investigate clitic doubling systems.

2.2 The syntax of clitic doubling: Move-and-Reduce

It has often been observed that not all patterns that replicate the φ-features of a nominal behave like φ-agreement. Consider Amharic, as reported by Kramer (2014). Amharic has a process in which the φ-features of the object are reproduced on the verb.

(37) (Kramer 2014, pg. 594)

a. Almaz tämari -w -in$_i$ ayy -ätʃtʃ -iw.$_i$
A.FEM student -DEF.MASC -ACC see -3FEM.SUBJ -3MASC.OBJ
'Almaz saw the male student.'

b. Almaz tämari -wa -n$_i$ ayy -ätʃtʃ -at.$_i$
A.FEM student -DEF.FEM -ACC see -3FEM.SUBJ -3FEM.OBJ
'Almaz saw the female student.'

In (37), we see that the a suffix on the verb bears the φ-features of the object. As Baker (2012) and Kramer (2014) observe, a reasonable first hypothesis is that this is a form of object φ-agreement.

But at the same time, this process in Amharic is different from the φ-agreement we saw in the previous section in several important ways. Most strikingly, object doubling in Amharic can yield new variable-binding possibilities. To see this, consider the paradigm in (38).

(38) Amharic object doubling affects variable-binding (Kramer, 2014)

a. Tigist$_i$ tämari -wa$_i$ -n ayy -ätʃtʃ.
T. student -her -ACC see -3SG.FEM

'Tɨgɨst$_i$ saw her$_i$ student.'

b. ?* Tämari -wa$_i$ Tɨgɨst$_i$ -ɨn ayy -ä.
student -her T. -ACC see -3SG.MASC
Intended: 'Her$_i$ student saw Tɨgɨst$_i$.'

c. Tämari -wa$_i$ Tɨgɨst$_i$ -ɨn ayy -**at**$_i$.
student -her T. -ACC see -3SG.MASC
'Her$_i$ student saw Tɨgɨst$_i$.'

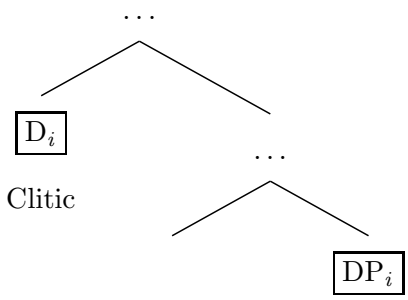
In (38a), we see a routine example in which the subject binds a variable in the object. Interestingly, (38b) shows that cataphoric constructions are judged as severely degraded in Amharic. Setting aside the source of the ungrammaticality of cataphora, (38c) shows that if the agreement morpheme in question targets the object, then variable binding becomes possible and the ban on cataphora is lifted. This ability to affect variable binding would be surprising if this construction in Amharic were φ-agreement, as discussed in §2.1.3.

To account for this, as well as the differences that we will see shortly, Kramer (2014) and other have proposed alternative analyses. Notable among these are a family of analyses collectively referred to as clitic doubling.

(39) **Clitic doubling**: The phenomenon by which a clitic pronoun [of category D] co-occurs with a full DP in argument position forming a discontinuous constituent with it. (Anagnostopoulou 2007)

In clitic doubling, the φ-features of a nominal are cross-referenced by a pronominal clitic of category D, rather than a non-referential bundle of φ-features (Torrego 1988, Uriagereka 1988, Uriagereka 1995). As we shall see, the referential nature of the pronominal clitic derives many of the differences between φ-agreement and clitic doubling. Consider the schematized representation of clitic doubling based on the definition in (39), shown in (40).

(40) Schematized Clitic Doubling



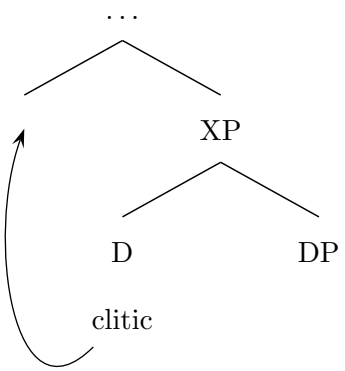
Doubled Nominal

In (40) we see a clitic pronoun of category D in a discontinuous position from the full nominal phrase that it cross-references. The question becomes how to analyze the dependency between the clitic pronoun and the doubled nominal.

There are three broad classes of analyses to derive (40). In the first of these, clitic doubling is simply a form of φ-agreement, as in Suñer (1988), Sportiche (1996), and Miller and Sag (1997). Since our focus here is on all the ways in which clitic doubling and φ-agreement differ from one another and how we could understand such contrasts, it makes little sense for us to pursue such a reductionist freamework. Additionally, with the exception of Suñer (1988), this sort of analysis is usually proposed for languages that do not actually have nominal doubling, but some kind of clitic movement, specifically French (Kayne 1975; Jaeggli 1982; Sportiche 1996; Miller and Sag 1997). Furthermore, these analyses do not consider the doubled clitic to be a pronoun, which makes the clustering of properties sensitive to binding and definiteness mysterious. As such, I will not consider agreement-based accounts of clitic doubling.

Among analyses of clitic doubling which distinguish it from agreement, two sorts of proposals can be distinguished. The first is commonly referred to as the 'Big-DP hypothesis,' associated with Torrego (1988, 1992), Uriagereka (1995), Franks and Rudin (2005), Roberts (2010), and Nevins (2011). The details among these analyses differ, but in each, the nominal and the clitic enter the derivation as a constituent,and are separated by subsequent movement of the clitic.

(41) Big-DP hypothesis

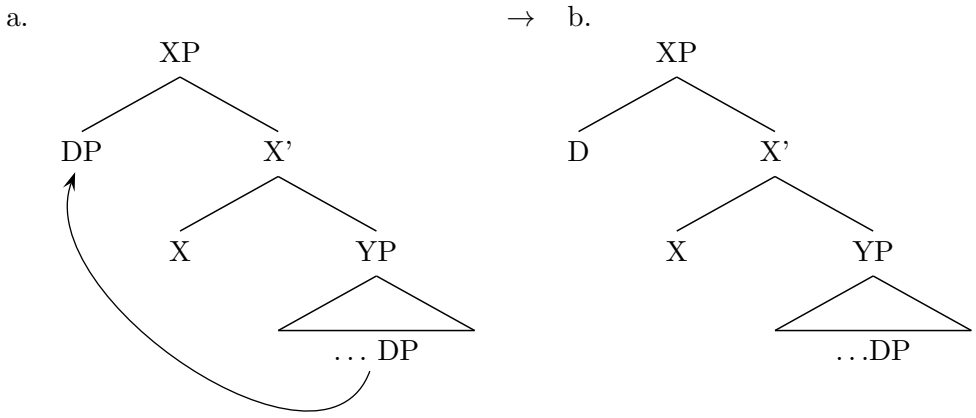


In (41), the clitic and the doubled nominal enter the derivation within some projecton XP. Analyses differ with respect to the identity of XP. For some, it is another DP projection, as in Uriagereka (1995), while for others it is a K(ase)P projection (Franks and Rudin 2005; Roberts 2010; Nevins 2011). The clitic then undergoes movement, stranding the rest of the DP. This derives the discontinuous aspect of clitic doubling constructions, as well as the dependency between the higher pronoun and the lower DP.

The second option still involves movement, but where Big-DP accounts rely on head movement out of the doubled nominal, "Move-and-Reduce" analyses propose that the entire doubled nominal undergoes phrasal movement to a higher position. After phrasal movement occurs, the higher copy of the doubled nominal undergoes a form of reduction.[8]

(42) Clitic doubling as Move-and-Reduce

[8](42) is adapted from an earlier, unpublished version of Baker and Kramer (2018).



In (42), the entire doubled nominal undergoes movement to a higher position. Then, the higher copy is reduced to the clitic pronoun. This straightforwardly accounts for the discontinuous aspect of clitic doubling construction, and reduces the dependency with the higher clitic to the standard dependency between movement copies.

In this way, Move-and-Reduce accounts of clitic doubling leverage an important aspect of the copy theory of movement (Chomsky 1995): two instances of a syntactic object provide the possibility for two occurrences at PF. Given this appeal to the basic workings of minimalist syntax, Move-and-Reduce approaches have found a natural home in minimalist analyses of clitic doubling, such as Anagnostopoulou (2007), Harizanov (2014b,a), Kramer (2014), and Baker and Kramer (2016).

Big-DP and Move-and-Reduce hypotheses share many characteristics, and both may be attested cross-linguistically. Furthermore, with the advent of bare phrase structure was proposed in Chomsky (1995), distinguishing head movement from phrasal movement on purely syntactic grounds becomes difficult. See Carnie (1995), Toyoshima (2001), Matushansky (2006), Vicente (2007), Harizanov (2016), and Harizanov and Gribanova (2017). Here, I will adopt Move-and-Reduce as the derivation for clitic doubling. This is largely from the perspective of theoretical parsimony. There is little independent evidence for the Big DPs that must be postulated and, at least on most recent versions, the analysis must tolerate excorporation of a head out of a larger phrase.

This has been difficult to motivate. For instance, Roberts (2010) provides one of the most explicit theoretical mechanisms for enacting just this head-excorporation. That said, his system is clearly built to handle French or Italian-style cases of clitic movement, rather than clitic doubling. To handle clitic doubling, Roberts relies on the so-called Kayne/Jaeggli Hypothesis.

(43) **The Kayne/Jaeggli Hypothesis**: Clitic doubling only doubles nominals that receive Case in some way other than the usual means, generally with a preposition. (Kayne 1975; Jaeggli 1982; Borer 1984; Suñer 1988)

The Kayne/Jaeggli Hypothesis was built to handle cases in Spanish and Romanian in which accusative clitic doubling seems to only target nominals which are marked with *a*, a preposition meaning 'to' that is also used in Differential Object Marking constructions (see Aissen 2003 for an overview). But as Suñer (1988) observes, there are well-attested instances of accusative clitic doubling in Spanish that do not double an *a*-marked nominal.

(44) Violations of the Kayne/Jaeggli Hypothesis (Suñer 1988, pg. 399)

a. Yo **la**$_i$ tenía prevista [esta muerte]$_i$.
I 3SG.F.O had.IMPF foreseen this death
'I had foreseen this death.'

b. Ahora tiene que seguir usándo **-lo**$_i$ [el apellido]$_i$.
now has.to.3SG continue using -3SG.M.O the last.name
'Now s/he has to go on using the last name.'

With these concerns in mind, I do not adopt a Big-DP analysis of clitic doubling, because the necessary theoretical mechanisms to make the system work are still largely murky. This, of course, is not necessarily a condemnation of the general approach, but an analysis that employs independently attested and necessary mechanisms is preferable. This is precisely why Move-and-Reduce accounts are appealing: they use only mechanisms that have independent and empirical support elsewhere.

The first necessary mechanism for Move-and-Reduce is the copy theory of movement, a core part of minimalist theory since Chomsky (1993). To see how this works, consider the pre-minimalist representation of movement in (45a) along with the copy theory compliant (45b).

(45) Representation of "Who do you think Joe talked to?"

a. Who$_i$ do you think t_i Joe talked to t_i

b. Who do you think ~~*who*~~ Joe talked to ~~*who*~~

Each of the representations in (45a) and (45b) has a way of marking the intermediate position of the wh-movement dependency. Where they differ is in the specific object which occupies the intermediate position.

In (45a), intermediate positions are occupied by a syntactic object that is formally different from the wh-expression 'who,' commonly referred to as a trace or empty category. Strictly speaking, the trace and the wh-expression do not stand in an identity relation. Rather, the two bear an identical index and are related to each other by an intricate theory of government and anaphor binding (see Chomsky 1980, 1986, Freidin and Lasnik 1981 and Haegeman 1994 for an overview).

The copy theory of movement, demonstrated in (45b), is similar in many ways. Like (45a), intermediate landing sites in a movement chain are occupied by a syntactic object which relates to the wh-expression 'who.' But where Government and Binding approaches appeal to binding to yield the correct interpretation of the moved wh-expression, (45b) does so in a much more straightforward way: the moved wh-expression is interpreted in the intermediate landing sites because a true copy of the wh-expression itself occurs there.

This approach to movement dependencies has several advantages. First, it trades the notion of government for the independently necessary idea of Merge (Chomsky 1993). Second, several patterns that were mysterious for theories of movement appealing

to traces receive a natural explanation (see Bošković and Nuñes 2007 for a detailed overview). The first of these are patterns of wh-movement in languages other than English. Consider Afrikaans, German, and Illasi (a Gallo-Italic language).[9]

(46) Afrikaans wh-movement (du Plessis, 1977)

a. **Met wie** het jy nou weer gesê **met wie** het Sarie gedog **met wie**
with who did you now again said with who did S. thought with who
gaan Jan trou?
go J. marry
'Whom did you say (again) did Sarie think Jan is going to marry?'

b. **Waaroor** dink jy **waaroor** dink die bure **waaroor** stry
whereabout think you whereabout think the neighbors whereabout argue
ons die meeste?
we the most
'What do you think the neighbors think we are arguing about the most?'

(47) German wh-movement (McDaniel, 1986)

a. **Wen** glaubt Hans **wen** Jakob gesehen hat?
whom thinks H. whom J. seen has
'Who does Hans think Jakob saw?'

b. **Wen** denkst Du **wen** sie meint **wen** Harald liebt?
whom think you whom she believes whom H. loves
'Who do you think that she believes that Harald loves?'

(48) Illasi wh-movement (Poletto and Pollock, 2004)

a. **S'** a -lo fato **che**?
what has he done what
'What has he done?'

b. **Ndo** e -lo ndat **endoe**?
where is -he gone where
'Where has he gone?'

[9]Identical patterns are also found in American and Brazilian Sign language (Nunes and Müler de Quadros 2004).

(46-48) demonstrate a variety of languages in which copies of wh-expressions are spelled out in multiple places. If intermediate representations of wh-movement were traces, it is not obvious how the trace would be converted to a phonological representation identical to the head of the chain.

Important for our purposes, it is well-documented that these sorts of "full copy" patterns are not the only option cross-linguistically. Various languages demonstrate a pattern in which intermediate copies receive a phonological exponent as well, but in a form that is somehow reduced. This option is referred to as partial copy doubling.[10] Two cases from the very distantly related Nùpe (Volta-Niger, Nigeria) and Seereer (Senegambian, Senegal) are provided below.[11]

(49) Partial copy doubling in Nùpe (Kandybowicz 2006, pg. 252)

a. **Zèé**$_i$ u: bè [ke **u:**$_i$ má du] na o?
who 3SG seem C 3SG know cook NA O
'Who does it seem ~~*who*~~ knows how to cook?'

b. ***Zèé**$_i$ u: bè [ke **zèé**$_i$ má du] na o?
who 3SG seem C 3SG know cook NA O

c. * **Zèé**$_i$ u: bè [ke ⊠$_i$ má du] na o?
who 3SG seem C 3SG know cook NA O

(50) Partial copy doubling in Seereer (Baier 2014)

a. **Xar**$_i$ xalaat -o [yee **ten**$_i$ Yande a lay -u [yee **ten**$_i$
what think -2SG.WH.AGR C 3SG Y. 3SBJ say -WH.AGR C 3SG

[10] Similar patterns are also found in Dinka Bor (van Urk, To Appear) and Finnish (Holmberg and Nikanne, 2008).

[11] These patterns differ from two other important patterns described in the literature. The first is true resumption of the Irish sort (see especially McCloskey 2002). Unlike Irish, "resumption" in these languages is sensitive to islands. This indicates that "resumption" here involves movement, but not in Irish. Second, partial copy doubling differs empirically from "*wh*-scope marking" in German, Romani, Hindi, and Warlpiri (McDaniel 1989, Dayal 1994, Fanselow and Mahajan 2000, Legate 2011) in that only intermediate copies may be reduced. Additionally, this sort of "direct-dependency" approach, as it is known in the *wh*-scope marking literature, has been demonstrated to be untenable for a variety of languages such as Hindi (Dayal 1994, 2000), Passamaquoddy (Bruening 2004), and Warlpiri (Legate 2011). See Bruening (2006) especially for differences between *wh*-scope marking and multiple copy doubling in Passamaquoddy, which demonstrates both patterns.

Jegaan a jaw -u $\_\__i$]]?
J. 3SBJ cook -WH.AGR
'What do you think [~~*what*~~ Yande said [Jegaan cooked ~~*what*~~]]?'

b. * **Xar**$_i$ xalaat -o [yee **xar**$_i$ Yande a lay -u [yee **xar**$_i$ Jegaan a jaw -u $\_\__i$]]?

c. **Xar**$_i$ xalaat -o [yee ⊠$_i$ Yande a lay -u [yee ⊠$_i$ Jegaan a jaw -u $\_\__i$]]?

In both Nùpe and Seereer, pronouns occur in precisely the positions which syntactic theory posits intermediate, successive cyclic movement. These pronouns, while not full copies like in (47-49), clearly demonstrate a related phenomenon.

This conclusion is not an artifact of wh-movement. A second construction in which reduced forms of multiple copies are spelled out is Predicate Clefting.[12] Predicate Clefts have been reported in a variety of genetically diverse languages, some of which are presented below (see Hiraiwa 2005, Kandybowicz 2006 and Vicente 2007 for especially thorough bibliographies).

(51) Predict Clefts in Vata (Kru, Koopman 1984)

a. **Lɛ̄** à **lɛ̄** saká.
eat we eat.PRES rice
'We are really EATING rice.'

b. **Nyɛ** ɔ ká yɔ -ɔ saká **nyɛ** kā mii.
give s/he FUT child -DEF rice give KA leave
'She will go GIVE rice to the child.'

(52) Predicate Clefts in Hebrew (Semitic, Landau 2006)

a. **Lirkod**, Gil lo **yirkod** ba-xayim.
to.dance G. not will.dance in-life
'As for dancing, Gil will never dance.'

b. **Liknot**, hi **kanta** et ha- praxim.
to.buy she bought ACC the- flowers

[12]Similar analyses have been put forth as well for a variety of constructions, such as demonstrative doubling in Greek (Grohmann and Panagiotidis, 2004). See Bošković and Nuñes (2007) for an overview.

'As for buying, she bought the flowers.'

(53) Predicate Clefts in Hungarian (Uralic, Vicente 2007)

a. **Olvasni**, **olvasott** János egy könyvet.
read.INF read.PAST.3SG J. a book.ACC
'As for reading, János read a book.'

b. **Úszni**, úszott Mari.
swim.INF swam.3SG M.
'As for swimming, Mari swam.'

In each of the sentences in (51-53), a normal, fully conjugated verb occurs in the expected position within the clause while a bare, unconjugated copy of the verb occurs in a left-peripheral position. Importantly, it has been clearly demonstrated in a number of different works that these constructions involve syntactic movement of a V or a VP (see especially Kandybowicz 2006, Landau 2006, and Vicente 2007).

There is a clear connection between the morphological characteristics of partial wh-copying in (49-50) and the Predicate Clefts in (51-53). Each pattern shows that multiple copies created by movement may have morphophonological exponence. Furthermore, both show that multiple copies in a movement chain need not be spelled out equally: intermediate copies may be morphologically reduced.

This conclusion is directly relevant for Move-and-Reduce analyses of clitic doubling. For these theories, the clitic is a partial copy of the doubled nominal itself. Crucially, this technology of multiple copy spell-out combined with copy reduction are well-attested processes cross-linguistically, as we saw above. Therefore, Move-and-Reduce accounts of clitic doubling do not need to appeal to any technology that is not independently attested cross-linguistically or theoretically. I take this to mean that Move-and-Reduce accounts are more appealing from the perspective of global theoretical parsimony than Big DP analyses.

Let us step through an illustrative derivation of a Move-and-Reduce analysis

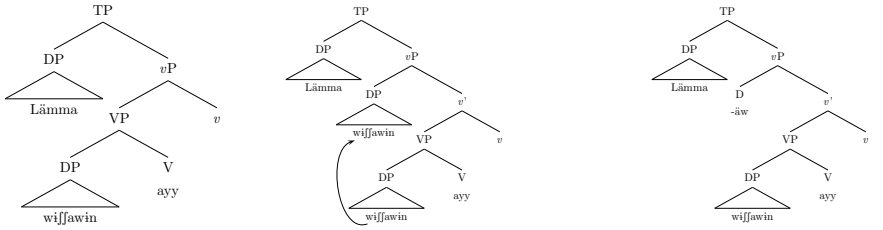
of clitic doubling. Consider a sample derivation for the Amharic in (54).[13]

(54) Lämma wɨʃʃa$_i$ -w -ɨn ayy **-äw**$_i$.
L. dog -DEF -ACC see.PERF -3MS.O
'Lemma saw the dog.'

First, following Kramer (2009, 2014), let us assume the syntactic structure in (55a), following movement of the subject to Spec,TP and abstracting away from accusative case assignment on the object. The object undergoes object shift to Spec,vP, shown in (55b). We will discuss the connection between clitic doubling and A-movement in §2.2.2.1, but for now let us continue though this derivation to (55c), which shows the reduction of the higher copy to the doubled clitic, while the lower copy is pronounced fully. At some later stage in the derivation, the clitic in Spec,vP will be morphologically unified with the verb, but on the syntactic end, the derivation in (55) suffices.

(55)

a. Structure of (54) b. Object shift from (55a) c. Redux of higher copy



In this way, we see that Move-and-Reduce analyses of clitic doubling differ substantively from φ-agreement in that Move-and-Reduce is fundamentally about syntactic movement, whereas φ-agreement is not.[14] It is this core distinction that derives

[13]This derivation was provided in an earlier, unpublished version of Baker and Kramer (2018).

[14]Many minimalist theories particularly since Chomsky (2000, 2001) have proposed that all syntactic movements are parasitic on AGREE. While this position is common, it is not without controversy. See Preminger (2014, to appear) especially. For the purposes of this dissertation, we can abstract away from these issues.

many of the differing empirical properties between these two systems. Let us consider these properties.

2.2.1 Property of clitic doubling #1: Dependencies behave like A-chains

Naturally, if Move-and-Reduce is a good representation of clitic doubling, the expectation will be that clitic doubling constructions should show properties associated with phrasal movement. This turns out to be the case. The connection between movement, particularly A-movement, and clitic doubling can be seen in two ways.

First, clitic doubling bleeds many of the signature properties of A'-movement. This is seen in Romanian. At a basic descriptive level, Romanian has two kinds of wh-movement. The first requires clitic doubling, and the second resists it. These are shown in (56) and (57) respectively.

(56) Romanian wh-movement with clitic doubling (Dobrovie-Sorin 1990, pg. 353)

a. Pe care băiat$_i$ **l-**$_i$ ai văzut?
pe which boy him- have.you seen
'Which boy did you see?'

b. * Pe care băiat ☒ ai văzut?

(57) Romanian wh-movement without clitic doubling

a. *Pe cine$_i$ **l-**$_i$ ai văzut?
pe who him have.you seen
Intended: 'Who have you seen?'

b. Pe cine ☒ ai văzut?

In (56) we see that D-linked wh-phrases such as *pe care* 'which' require clitic doubling, while in (57) we see that non-D-linked wh-phrases like *pe cine* 'who' may not be clitic doubled.

Interestingly, wh-constructions with clitic doubling like (56) seem to involve an A-movement chain, while those in (57) seem to involve only A'-movement. This can be seen in two ways. First, wh-movement without clitic doubling triggers Weak Crossover like A'-movement in other languages, but wh-movement with clitic doubling does not.

(58) Romanian wh-movement with clitic doubling does not trigger WCO (Dobrovie-Sorin 1990, pg. 358)

a. Pe care$_i$ **l-**$_i$ a certat mama lui$_i$ t_i?
pe which$_i$ him$_i$- has scolded mother his$_i$
'Which one$_i$did his$_i$ mother scold t_i?'

b. Pe al cui elev$_i$ îl nepredtătesc prietenii lui$_i$ t_i?
pe whose student$_i$ him$_i$ wrong friends his$_i$
'Whose student$_i$ do his$_i$ friends wrong t_i?'

(59) Romanian wh-movement without clitic doubling triggers WCO (Dobrovie-Sorin 1990, pg. 357-358)

a. *Pe cine$_i$ a certat mama lui$_i$ t_i?
pe who$_i$ has scolded mother his$_i$
Intended: 'Who$_i$ did his$_i$ mother scold t_i?'

b. *Ce copil$_i$ ar pedepsi părintii lui$_i$ t_i?
what child would punish parents his$_i$
Intended: 'What child$_i$ would his$_i$ parents punish t_i?'

The lack of WCO effects in (58) indicates that movement chains with clitic doubling involve A-movement. In contrast, the presence of WCO without clitic doubling in (59) indicates only A'-movement. In this way we see that clitic doubling constructions involve an A-movement component.

Furthermore, Dobrovie-Sorin (1990) (pg. 358) also observes that only wh-movement constructions without clitic doubling license parasitic gaps.

(60) Only wh-movement without clitic doubling licenses parasitic gaps

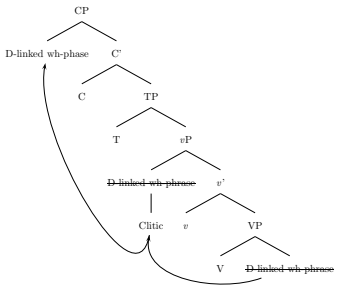
a. *Pe care$_i$ **l-**$_i$ ai apreciat t_i înainte de a cunoaste PG$_i$?
pe which$_i$ him$_i$- have.you appreciated before knowing
Intended: 'Which one did you appreciate before knowing?'

b. Pe cine$_i$ ☒ ai apreciat t_i înainte de a cunoaste PG$_i$?
pe who have.you appreciated before knowing
'Who did you appreciate before knowing?'

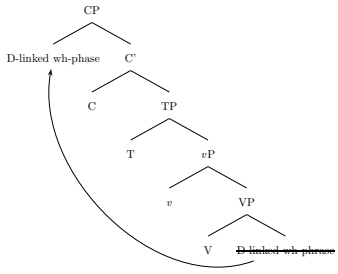
The inability to license a parasitic gap is further indication that clitic doubling involves A-movement (c.f. Engdahl 1983).

In this way, we see that clitic doubling in Romanian demonstrably involve A-movement. As Dobrovie-Sorin (1990) put it, "the inescapable conclusion is that some Romanian *wh*-structures do not rely on quantification, in contrast to *cine* ['who'] structures. This property converges with the distribution of [clitic doubling]" (pg. 359). This analysis is sketched in (61), which contrasts with the wh-movement process in (62) that does not involve A-movement or clitic doubling. In line with the analysis of Amharic presented in (55), I analyze the initial movement in (61) as movement to Spec,vP. See §2.2.2 for further discussion.

(61) Romanian wh-movement with clitic doubling



(62) Romanian wh-movement without clitic doubling



In (61), we see that wh-constructions with clitic doubling involve an extra step of A-movement to Spec,vP. I propose that this A-movement bleeds Weak Crossover and also fails to license a parasitic gap, despite secondary A'-movement to Spec,CP. In contrast,

in (62) we see that wh-constructions without clitic doubling involve only A'-movement straight to Spec,CP. This allows a parasitic gap to be licensed, as well as inducing Weak Crossover. Therefore, Romanian provides good evidence that clitic doubling constructions involve A-movement.

Further evidence that clitic doubling constructions involve a step of movement comes from their ability to yield variable-inding possibilities that would not be possible in their absence, just like XP-movement as discussed in §2.1.3. Consider the contrast in (63), repeated from (20-21) above.

(63) a. * It seems to her$_i$ mother that every girl$_i$ is a genius.

b. Every girl$_i$ seems to her$_i$ mother $\_\_{}_i$ to be a genius.

c. * There seem -$\emptyset_i$ to their$_i$ mother to be several girls$_i$ learning rocket science.

In (63a-b), we see that A-movement, in this case Subject-to-Subject Raising, yields new variable-binding possibilities that are impossible in its absence. Specifically, in (63b) we see that when the quantified DP 'every girl' undergoes A-movement to Spec,TP of the matrix clause it may bind the variable 'her' within the matrix experiencer. In (63a), we see that when no such movement takes place, this variable binding relation cannot be established. Likewise, (63c) demonstrates that φ-agreement in the absence of A-movement is insufficient to license variable binding.

Since at least Suñer (1988), it has been observed that clitic doubling constructions also make available variable-binding relations that may not occur in their absence, just like A-movement. Suñer's original examples from Porteño Spanish (Buenos Aires, Argentinian) are shown in (64).

(64) Porteño Spanish (Suñer 1988, pg. 421)

a. Todos$_i$ quieren a su$_i$ madre.
everyone likes their mother
'Everyone likes their mother.'

b. *?Su$_i$ madre quiere a todos$_i$.
their mother likes everyone
Intended: 'Their$_i$ mother likes everyone$_i$.'

c. Su$_i$ madre **los$_i$** quiere a todos$_i$.
their mother them likes everyone
'Their$_i$ mother likes everyone$_i$.'

(64a) demonstrates the unremarkable pattern of a subject binding a variable in object position. Likewise, the ungrammaticality of the (64b) can be readily derived if the quantified possessor attempts to bind the object variable, yielding a marked cataphoric interpretation in this language, similar to the Amharic in (38).

Interestingly, in (64c) we see that clitic doubling of the object greatly improves the configuration in (64b). While it remains largely unclear why clitic doubling has the ameliorating effect it does, these observations are important as an early demonstration of the potential for clitic doubling to expand anaphoric possibilities in interesting ways.

Similar observations quickly followed for a variety of languages, some involving the possibility of cataphora (frequently referred to as 'backwards pronominalization' in this literature), some involving bound variable anaphora. Kramer (2014) shows that clitic doubling in Amharic shows the first type of effect, repeated in (65) from (38).

(65) Amharic (Kramer 2014, pg. 604-605)

a. Tɨgɨst$_i$ tämari -wa$_i$ -n ayy -ätʃtʃ.
T.FEM student -her -ACC see -3FS.S
'Tigist$_i$ saw her student$_i$.'

b. *? Tämari -wa$_i$ Tɨgɨst$_i$ -ɨn ayy -ä.
student -her T.FEM -ACC see -3MS.S
Intended: 'Her$_i$ student saw Tigist$_i$.'

c. Tämari -wa$_i$ Tɨgɨst$_i$ -ɨn ayy **-at$_i$**.
student -her T.FEM -ACC see-(3MS.S) -3FS.O
'Her$_i$ student saw Tigist$_i$.'

Kramer notes that cataphora in Amharic is 'nearly ungrammatical' as seen in (65b), but that clitic doubling, as in (65c), repairs the violation.

Other cases that have been documented involve the binding of a pronoun by a quantifier. These come from Bulgarian, shown in (66), and Modern Greek, in (67).

(66) Bulgarian (Harizanov 2014a, pg. 1054-1055)

a. Petăr vărna [vsjaka kola]$_i$ [na sobstvenika i$_i$] včera.
P. returned every car to the.owner its yesterday
'Petăr returned every car to its owner yesterday.'

b. *Petăr vărna [na sobstvenika i$_i$] [vsjaka kola]$_i$ včera.
P. returned to the.owner its every car yesterday
Intended: 'Petăr returned every car to its owner yesterday.'

c. Petăr **ja**$_i$ vărna [na sobstvenika i$_i$] [vsjaka kola]$_i$ včera.
P. it returned to the.owner its every car yesterday
'Petăr returned every car to its owner yesterday.'

(67) Modern Greek (Alexiadou and Anagnostopoulou 1997)

a. *O Petros epestrepse [tu idioktiti tu$_i$]$_j$ [to kathe aftokinito]$_i$
the P.NOM returned.3SG the owner.GEN his the every car.ACC
xtes to vradi.
yesterday the night
Intended: 'Petros returned it$_i$ owner every car$_i$ last night.'

b. O Petros **to**$_i$ epestrepse [tu idioktiti tu$_i$]$_j$ [to kathe
the P.NOM it.ACC returned.3SG the owner.GEN his the every
aftokinito]$_i$ xtes to vradi.
car.ACC yesterday the night
'Petros returned its$_i$ owner every car$_i$ last night.'

Since Reinhart (1983), it has been widely believed that quantifier nominals can semantically bind pronouns only if they syntactically command them.[15] This is why the Bulgarian in (66b) and the Greek in (67a) are impossible. If the doubled clitics in (66c) and (67b) are reduced versions of the full quantified nominals that they double, as in

[15]Though see Baker (2012) for a skeptical review of the evidence for such a condition.

the Move-and-Reduce analysis pursued here, we understand how the 'Bound Anaphora Condition' can be met. This same approach can also be leveraged to explain why clitic doubling expands variable binding possibilities. In other words, the Modern Greek in (67b) will have a pre-Reduce configuration like (68).

(68) O Petros [*to kathe aftokinito*$_i$] epestrepse [to idioktiti tu$_i$] [to kathe aftokinito].
the every car the owner.GEN his

At this level of representation, the higher copy of *to kathe aftokinito* 'every car' can bind the pronoun within the indirect object *to idioktiti tu* 'its owner.'This allows for variable binding where, without movement, no such binding could occur.

In addition to ameliorating otherwise ill-formed anaphoric relations, clitic doubling can also induce violations in what would otherwise be well-formed structures. Consider the Modern Greek in (69).

(69) a. Sistisa [kathe gineka]$_i$ [ston melondiko andra tis$_i$]$_j$.
introduced.1SG every woman.ACC to.the future husband.DAT hers

'I introduced every woman to her husband.'

b. *__Tu__$_j$ sistisa [kathe gineka]$_i$ [ston melondiko andra tis$_i$]$_j$.
to.him.DAT introduced.1SG every woman.ACC to.the future husband.DAT hers

Intended: 'I introduced every woman to her husband.'

In (69a), we see that a grammatical sentence in which a quantifier in direct object position can bind a variable within the PP *ston melondiko andra tis* 'to her husband,' just as in English. Interestingly, in (69b), we see that clitic doubling the dative argument, ungrammaticality results.

I interpret this as clitic doubling recreating a backwards pronominalization

context. Consider the syntactic context that results before Reduction of the higher copy of the dative argument to the clitic, sketched in (70).

(70) [Ston melondiko andra tis$_i$] sistisa [kathe gineka]$_i$ [ston melondiko
to.the future husband her every woman
andra tis$_i$].

Crucially, (70) is the same sort of cataphoric context which we have already seen is ungrammatical in Modern Greek in (67a). Therefore, this pattern receives a natural account under a Move-and-Reduce analysis of clitic doubling.

The upshot of this discussion is that clitic doubling demonstrates many of the core properties of movement chains cross-linguistically. This receives a natural treatment under a Move-and-Reduce analysis of clitic doubling. Furthermore, this movement behavior can be used to distinguish clitic doubling from φ-agreement.

2.2.2 Property of clitic doubling #2: No clitic doubling out of coordinate structures

In §2.1.4, we saw that φ-agreement systems do not universally obey the Coordinate Structure Constraint, with Closest Conjunct Agreement being a well-attested pattern. Interestingly, this does not turn out to be the case in clitic doubling systems.[16] In this subsection, I will present what I believe are new data from Latin American Spanish, specifically Puerto Rican Spanish.[17]

In Puerto Rican Spanish, like other varieties of Latin American Spanish (Jaeg-

[16]Two works, van Craenenbroeck and van Koppen (2008) and Bošković (2018), report that clitic doubling may be marginal out of left conjuncts. I do not have anything to say about the Belgian Dutch from the latter work, though I was unable to replicate the result from the latter in Peruvian Spanish. Therefore, I will not engage with these works here for lack of data.

[17]Many thanks to Ivana Serrano for her judgments, enlightening discussion, and for connecting me with six other Caribbean Spanish speakers. Note that this description of Puerto Rican Spanish is apparently true for all Caribbean varieties of Spanish, as the same pattern was found with speakers of Panamanian, Columbian, and Cuban Spanish. Identical data were also found in Mexican Spanish, Central American Spanish, and Peruvian Spanish.

gli 1982), clitic doubling is required for pronominal direct objects but blocked for non-pronominal direct objects.

(71) Puerto Rican Spanish

a. **Lo**$_i$ vi a él$_i$ pero no **la**$_j$ vi a ella$_j$.
him see.1SG.PAST *a* him but not her see.1SG.APST *a* her
'I saw him, but I didn't see her.'

b. * ☒ Vi a él pero no ☒ vi a ella.

c. * **Lo**$_i$ vi a Juan$_i$ pero no **la**$_j$ vi a Maria$_j$.
him see.1SG.PAST *a* J. but not her see.1SG.APST *a* M.
Intended: 'I saw Juan, but I didn't see Maria.'

d. ☒ Vi a Juan pero no ☒ vi a Maria.

Interestingly, despite being required outside of coordination, clitic doubling is categorically banned from occurring out of coordinate structures. This is true no matter which conjunct the pronoun sits in.

(72) Puerto Rican Spanish

a. * El doctor **lo**$_i$ ayuda [a él$_i$ y a Juan].
the doctor him help.3SG.PRES *a* him and *a* J.
Intended: 'The doctor helps him and Juan.'

b. El doctor ☒ ayuda [a él y a Juan].
'The doctor helps him and Juan.'

c. * La maestra **lo**$_i$ vio [a Maria y a él$_i$].
the teacher him see.3SG.PAST *a* M. and *a* him
Intended: 'The teacher saw Maria and him.'

d. La maestra ☒ vio [a Maria y a él].
'The teacher saw Maria and him.'

(73) Puerto Rican Spanish

a. * El vampiro **lo**$_i$ mordió [a él$_i$ y a ella].
the vampire him bite.3SG.PAST *a* him and *a* her
Intended: 'The vampire bit him and her.'

b. El vampiro ☒ mordió [a él y a ella].

'The vampire bit him and her.'

c. * Una medusa **lo**$_i$ picó [a ella y a él$_i$].
a jellyfish him sting.3SG.PAST *a* her and *a* him
Intended: 'A jellyfish stung her and him.'

d. Una medusa ☒ picó [a ella y a él].

'A jellyfish stung her and him.'

In (72), we see that if a pronoun is coordinated with a non-pronominal DP in direct object position, clitic doubling out of the coordinate structure is ungrammatical. Likewise, when both conjuncts are pronouns, clitic doubling out of the coordinate structure is equally ungrammatical. This is seen in (73).

Interestingly, "resolved" clitic doubling does not improve grammaticality in any of these circumstances.

(74) Puerto Rican Spanish

a. * El doctor **los**$_{i+j}$ ayuda [a él$_i$ y a Juan$_j$].
the doctor him help.3SG.PRES *a* him and *a* J.
Intended: 'The doctor helps him and Juan.'

b. * La maestra **los**$_{i+j}$ vio [a Maria$_j$ y a él$_i$].
the teacher him see.3SG.PAST *a* M. and *a* him
Intended: 'The teacher saw Maria and him.'

c. * El vampiro **los**$_{i+j}$ mordió [a él$_i$ y a ella$_j$].
the vampire him bite.3SG.PAST *a* him and *a* her
Intended: 'The vampire bit him and her.'

d. * Una medusa **los**$_{i+j}$ picó [a ella$_j$ y a él$_i$].
a jellyfish him sting.3SG.PAST *a* her and *a* him
Intended: 'A jellyfish stung her and him.'

Finally, disjunction demonstrates the same behavior.

(75) Puerto Rican Spanish

a. * La maestra **lo(s)**$_{i(+j)}$ castigará [a él$_i$ o a Juan$_j$].
the teacher him/them punish.3SG.FUT *a* him or *a* J.
Intended: 'The teacher will punish him or Juan.'

b. La maestra ☒ castigará [a él o a Juan].
'The teacher will punish him or Juan.'

c. * Una abeja la$_i$/los$_{i+j}$ picó [a Juan$_j$ o a ella$_i$].
a bee her/him sting.3SG.PAST *a* J. or *a* her
Intended: 'A bee stung Juan or her.'

d. Una abeja ☒ picó [a Juan o a ella].
'A bee stung Juan or her.'

e. * El doctor **lo(s)**$_{i(+j)}$ va a ayudar [a él$_i$ o a ella$_j$].
the doctor him/them go.3SG.PRES to help *a* him or *a* her
Intended: 'The doctor is going to help him or her.'

f. El doctor ☒ va a ayudar [a él o a ella].
'The doctor is going to help him or her.'

This complete inability to clitic double out of coordinate or disjunctive structures is exactly what we expect if clitic doubling is derived through movement. This is because, like all movement, the movement hypothesized to underpin clitic doubling constructions should obey island constraints such as the Coordinate Structure Constraint (Ross 1967). Unsurprisingly, the Coordinate Structure Constraint is indeed active in Puerto Rican Spanish. This is demonstrated for subject movement of an unaccusative predicate in (76a-b) as well as passivization in (76c-d).

(76) a. [Juan y Maria]$_i$ cayeron __$_i$.
J. and M. fall.PAST.3PL
'Juan and Maria fell.'

b. * Juan$_i$ cayó/ cayeron [__$_i$ y Maria].
J. fall.PAST.3SG/ fall.PAST.3PL and M.
Intended: 'Juan fell and Maria.'

c. [Juan y Maria]$_i$ fueron besados __$_i$.
J. and M. be.PAST.3PL kissed.3PL.MASC
'Juan and Maria were kissed.'

d. * Juan$_i$ fue/ fueron besado(s) [
J. be.PAST.3SG/ be.PAST.3PL kissed.3SG.MASC/3PL.MASC
__—sub$i$ y Maria].
and M.
Intended: 'Juan was kissed and Maria.'

Therefore, this sensitivity to coordination is straightforwardly predicted within our system of clitic doubling.

2.2.3 Property of clitic doubling #3: Locality determined by movement

In considering the locality conditions of φ-agreement in §2.1.2, we saw that φ-agreement obeys the Phase Impenetrability Condition. The definition of the PIC is repeated in (77).

(77) **Phase Impenetrability Condition**: In which HP is a strong phase with a head H, the domain of H is not accessible to operations outside HP; only H and its *edge* are accessible to such operations. (Chomsky 2001, pg. 13)

As (77) makes clear, elements on the edge of the phase are accessible to syntactic operations outside of the phase, such as φ-agreement or further movement operations.

Crucially for our purposes, A-movement frequently displaces nominals to the phase edge, rendering them accessible to further extraction. Let us consider cases of A-extraction out of three phases: finite CP and PP, as well as DP.[18]

[18] Again, the reason why these phases were chosen for discussion is because there seem to be the phasal categories in SMP Mixtec. See §5.1.

2.2.3.1 A-movement can extract nominals from finite CPs

To begin, English famously forbids A-movement out of finite CPs.

(78) a. Maria$_i$ seems [t_i to be a genius].

b. * Maria$_i$ seems [that $_i$ is a genius].

c. It seems [Maria is a genius].

In comparing (78a) to (78b), we see that Subject-to-Subject movement is permitted only from nonfinite clauses, while it may not occur out of a finite CP. (78c) confirms that the predicate 'seem' permits a finite CP complement. Therefore the ungrammaticality of (78b) must be the result of A-movement from a finite CP.

At the same time, this restriction is not universal. First consider Japanese, which has an A-movement process that moves embedded subjects to direct object position of the matrix clause (Kuno 1976, Tanaka 2002, inter alia). First consider (79). All Japanese data in this discussion is from Kuno (1976) through Tanaka (2002).

(79) a. John -ga **orokanimo** [Bill -ga tensai -da -to] omot -teiru.
J. -NOM stupidly B. -NOM genius -COP -C think -PROG
'John stupidly thinks that Bill is a genius.'

b. * John -ga Bill -ga **orokanimo** tensai -da -to omot -teiru.

In (79), the matrix adverb *orokanimo* 'stupidly' must precede the embedded nominative subject *Bill -ga*. Therefore, we see that the embedded subject must remain in situ when marked nominative. With this in mind, consider (80). The translation is provided by Tanaka (2002).

(80) John -ga **Bill -o**$_i$ orokanimo [t_i tensai -da -to] omot -teiru.
J. NOM B. -ACC stupidly genius -COP -C think -PROG
'John thinks of Bill stupidly as a genius.'

In (80), we see that the nominal that was the embedded subject in (79), *Bill*, now occurs to the left of the matrix adverb *orokanimo* 'stupidly.' Additionally, it is marked with

accusative case, rather than nominative. Therefore, it clearly has been displaced from the embedded clause.

Additionally, there is good reason to consider this movement out of the finite embedded clause to be A-movement because this movement affects variable binding. Consider (81).

(81) a. * John -ga$_i$ kare -o$_i$ hihansita.
J. -NOM he -ACC criticized
Intended: 'John$_i$ criticized him$_i$.'

b. John -ga$_i$ [kare -ga$_i$ baka -da -to] omot -teiru.
J. -NOM he -NOM fool -COP -C think -PROG
'John$_i$ thinks that he$_i$ is a fool.'

In (81a), a pronoun in object position may not be coreferent with an R-expression in subject in a classic Condition B effect. Likewise, in (81b) we see that when the pronoun and the R-expression are separated by a finite clause boundary, they may be coreferent.

Importantly, when a pronoun in embedded subject position raises to object position of the matrix clause, they may no longer corefer.

(82) * John -ga$_i$ kare -o$_i$ [t$_i$ baka -da -to] omot -teiru.
J. -NOM he -ACC fool -COP -C think -PROG
Intended: 'John$_i$ thinks of him$_i$ as a fool.'

From this, we can conclude that this displacement of embedded subjects to matrix objects out of finite clauses is indeed a form of A-movement. Therefore, it seems inescapable that A-movement may cross a finite clause boundary.

In addition to Raising to Object in Japanese, the literature abounds of other examples of "Hyperraising," or A-movement out of finite clauses. Further cases of Raising to Object out of a finite embedded clauses have been reported in Janitizo P'urhepecha, an isolate spoken in Michoacán, Mexico.

(83) A-movement to Object in Janitzio P'urhepecha (Zyman 2017)

a. Ueka -sïn -∅ -di =sï [eska **Xumo** u -a -∅ -ka ma
want -HAB -PRES -IND3 =PS that X. make -FUT -PRES -SJV a
k'umanchikua].
house
'They want Xumo to build a house.'

b. Ueka -sïn -∅ -di =sï **Xumo -ni**$_i$ [eska t_i u -a -∅ -ka
want -HAB -PRES -IND3 =PS X. -ACC that make -FUT -PRES -SJV
ma k'umanchikua].
a house
'They want Xumo to build a house.'

In (83), we see that the embedded subject *Xumo* may receive (unmarked) nominative case within the embedded CP to the left of the complementizer *eska* 'that,' or it may be marked accusative and occur within the matrix clause, to the right of *eska*. Just like with Japanese, much of the same evidence applies here that the movement raises the embedded subject into the matrix clause. For instance, raised objects may occur to the left of matrix adverbs, as in (84).

(84) Emilia ueka -sïn -∅ -di Xumo -ni$_i$ **minstita -ni jingoni** [eska t_i
E. want -HAB -PRES -IND.3 X. -ACC heart -ACC with that
jaruata -a -∅ -ka pauani].
help -FUT -PRS -SUBJ tomorrow
'Emilia wants Xumo with all her heart [to help her tomorrow].'

Additionally, just like Japanese, the position to which the raised subject moves seems to be an A-position. Consider the contrast in (85).

(85) a. Ueka -pirin -∅ -ga =ni eska [**Xumu -eri**$_i$ k'uinchikua jimbo]
want -COND -PRS -IND.1 =1SS that X. -GEN party in
ima$_i$ intsïmpe -pirin -∅ -ga ujtsïkukate -echa -ni.
he serve -COND -PRS -SBJV pastry -PL -ACC
'I'd like it if [at Xumo's$_i$ party] he$_i$ served pastries.'

b. * Ueka -pirin -∅ -ga =ni **ima**$_i$ -ni eska [**Xumu -eri**$_i$
want -COND -PRS -IND.1 =1SS he -ACC that X. -GEN
k'uinchikua jimbo] t_i intsïmpe -pirin -∅ -ga ujtsïkukate -echa
party in serve -COND -PRS -SBJV pastry -PL

-ni.
-ACC
Intended: 'I'd like him$_i$ to, at Xumo's$_i$ party, serve pastries.'

In (85a), we see that no Condition C violation occurs with the embedded subject *ima* 'he' remains in situ in the embedded clause. In contrast, in (85b) when *ima* 'he' undergoes Hyperraising, a Condition C violation occurs. This indicates that Hyperraising in Janitzio P'urhepecha is a form of A-movement.

In addition to these Subject-to-Object movements, the Nguni Bantu languages allow Subject-to-Subject movement out of finite clauses (Zeller 2006, Halpert 2012).

(86) A-movement to Subject in Zulu (Halpert 2012)

a. Ku- fanele [ukuthi **abantwana** ba- fund -e].
C.17- ought that child.2 C.2- study -SUBJ
'It is necessary that the children study.'

b. **Abantwana** ba- fanele [ukuthi ba- fund -e].
child.2 C.2- out that C.2- study -SUBJ
'The children must study.'

In (86) we see that the embedded subject *abantwana* 'children' may raise to subject position of the matrix clause and with the matrix verb. Just as in Japanese and Janitzio P'urhepecha, this Subject-to-Subject movement out of the finite embedded clause affects variable binding. Consider (87).[19]

(87) (Halpert 2012)

a. Ku- fanele [ukuthi [ngo- buhlakana buka Sipho$_i$] *pro*$_i$ a$_i$- m$_j$- siz
C.17- ought that out- wisdom POSS S. he AGR- AGR- help
-e uThemba$_j$].
-AUG T.
'It is necessary that out of Sipho's$_i$ wisdom, he$_i$ helps Themba.'

[19] I altered Halpert's original glosses in this example for ease of reading.

b. * *pro*$_i$ u$_i$- fanele [ukuthi [ngo- buhlakana buka Sipho$_i$] t_i a$_i$-
he AGR- necessary that out- wisdom POSS S. AGR-
m$_j$- siz -e uThemba$_j$].
AGR- help -AUG T.
Intended: 'He$_i$ must, out of Sipho's$_i$ wisdom, help Themba.'

(87) provides similar data to the Janitzio P'urhepecha in (85): if an embedded subject, in this case a silent pronoun, undergoes Subject-to-Subject Raising to the matrix clause, as evidenced by the matrix agreement, it binds into the embedded clause. This yields a Condition C violation in (87b), and further solidifies that this process is A-movement.

With the Japanese, Janitzio P'urhepecha, and Zulu in mind, it is an undeniable fact that A-movement may cross finite clause boundaries.

2.2.3.2 A-movement can extract nominals from DP

DP is widely considered to be a phase (see Abels 2003, Heck and Zimmermann 2004, and Svenonius 2004), yet many languages allow extraction of possessors from DP.

This phenomenon, referred to as Possessor Raising, is a variety of a general phenomenon referred to as external possession. In some languages, external possession does not seem to involve movement, but for others, it demonstrably involves movement. Here I will focus on a particularly clear argument for A-movement of possessors in one language: Nez Perce. All Nez Perce data and analysis come from Deal (2013).

First consider the behavior of Nez Perce objects. As Deal demonstrates, transitive objects are marked with distinct case morphology in *-ne*, referred to objective case.

(88) a. Hi- pnim -se -∅ picpic.
3SUBJ- sleep -IMPERF -PRES cat.NOM
'The cat is sleeping.'

b. Ciq'aamqal -nim pee- tw'ehke'yk -se -∅ picpic -**ne**.
dog -ERG 3/3- chase -IMPERF -PRES cat -OBJ
'The dog is chasing the cat.'

A second important aspect of Nez Perce transitive objects is that they control object agreement for both number and person.

(89) a. *pro* cewcew -tee'nix.
we call -HAB.PRES.PL
'We call/make phone calls.'

b. *pro* **'e**$_i$- cewcew -tee'nix 'ip$_i$ -ne.
we 3OBJ- call -HAB.PRES.PL 3SG -OBJ
'We call him/her.'

c. *pro* **'e**$_i$- **nees**$_i$- cewcew -tee'nix immu$_i$ -ne.
we 3OBJ- O.PL- call -HAB.PRES.PL them -OBJ
'We call them.'

With this in mind, let us consider the behavior of a class of possessors which Deal refers to as inalienable possessors that modify logical objects. Interestingly, it is the inalienable possessor, not the possessum, which behaves like the transitive object.

(90) a. *pro* hi- **nees**$_i$- hex -ne'ny -∅ -e **ma- may'as**$_i$ **-na**
he 3SUBJ -O.PL- see -μ -P -REM.PAST PL- child -OBJ
pist.
father.NOM.
'He saw the children's father.'

b. Himiis -nim **pee**$_i$- p -e'ny -∅ -e hoq'hoq' -na siis.
wolf -ERG 3/3- eat -μ -P -REM.PAST pig -OBJ soup.NOM
'The wolf ate the pig's soup.'

c. * Himiis -nim pee- p -e'ny -∅ -e hoq'hoq' siis **-na**.
wolf -ERG 3/3- eat -μ -P -REM.PAST pig soup -OBJ
Intended: 'The wolf ate the pig's soup.'

In (90), we see that inalienable possessors, like *mamay'as* 'children' and *hoq'hoq'* 'pig' behave like transitive objects. This can be seen in that they control object agreement on the verb as in (90a), and in that only the inalienable possessor may be marked with objective case. This can be seen in comparing (90b-c).

Furthermore, Deal argues that the inalienable possessor and the possessum do not form a constituent. This can be seen in that they need not be adjacent, as in (91), and the possessor may undergo wh-movement without pied-piping the possessum, as in (92).

(91) a. Angel -nim paa- 'yaxˆ -na'ny -∅ -a **Tatlo -na** taaqmaał.
A. -ERG 3/3 find -μ -P -REM.PAST T. -OBJ hat.NOM
'Angel found Tatlo's hat.'

b. Angel -nim Tatlo -na paa- 'yaxˆ -na'ny -∅ -a taaqmaał.

c. Angel -nim taaqmaał. paa- 'yaxˆ -na'ny -∅ -a **Tatlo -na**.

(92) a. **'Isii**$_i$ **-ne** *pro* 'e- sewleke'yk -ey' -se -∅ t$_i$ 'aatoc?
who -OBJ you 3OBJ- drive -μ -IMPERF -PRES car.NOM
'Whose car are you driving?'

b. **'Isii**$_i$ **-ne** *pro* 'aw- 'yaxˆ -na'ny -∅ -a t$_i$ 'iniit?
who -OBJ 3OBJ- find -μ -P -REM.PAST house.NOM
'Whose house did you find?'

Crucially, as Deal (2013) observes, this behavior is not observed in sentences without Possessor Raising. For instance, in ditransitives an inalienable possessor in the direct object may not undergo possessor raising. This can be seen in (93), where the possessor in the direct object is marked with genitive case rather than objective case. We will discuss this shortly, but for now let us simply observe the tight connection between possessor raising and the separability demonstrated in (91-92).

(93) a. *pro* 'ew- 'nii -se -∅ Tatlo -na **Angel -nim** taaqmaał.
I 3OBJ- give -IMPERF -PRES T. -OBJ A. -GEN hat.NOM
'I'm giving Tatlo Angel's hat.'

b. * *pro* 'ew- 'nii -se -∅ **Angel -nim** Tatlo -na taaqmaał.

c. * *pro* **Angel -nim** 'ew- 'nii -se -∅ Tatlo -na taaqmaał.

(94) a. [Isii -nm ciickan]$_i$ *pro* 'ew- 'nii -∅ -ye t$_i$ 'aayat -ona?
who -GEN blanket.NOM you 3OBJ- give -P -REM.PAST woman -OBJ

‘Whose blanket did you give to the lady?’

b. * [Isii -nm]$_i$ *pro* ’ew- ’nii -∅ -ye t$_i$ ciickan ’aayat -ona?
who -GEN you 3OBJ- give -P -REM.PAST blanket.NOM woman -OBJ
‘Whose blanket did you give to the lady?’

In (93) we see that in situ possessors marked with genitive case must be linearly adjacent with their possessum, unlike the raised, objective marked possessors in (91). Likewise, questioning an in situ possessor in (94), pied-piping the possessum of the possessor, unlike with the raised possessor in (92).

Deal (2013) argues that this process is true A-movement based on locality of the sort we examined above. First, Deal demonstrates that in ditransitives, the indirect object is structurally higher than the direct object. This is done by showing that the indirect object may bind a variable in the direct object, but not the other way around.

(95) a. Pinooc -nim$_i$ pee- kiwyek -∅ -e Elwit’et$_i$ -ne [’ip$_{i,\ j}$ -nim hipt].
P. -ERG 3/3- feed -P -REM.PAST E. -OBJ 3SG -GEN food.NOM
‘Pinooc$_i$ fed Elwit’et$_i$ her$_j$/his$_i$ food.’

b. Pinooc -nim$_i$ pee- kiwyek -∅ -e ip$_{*i,\ *j}$ -ne [Elwit’et$_j$ -nim hipt].
P. -ERG 3/3- feed -P -REM.PAST 3SG -OBJ E. -GEN food.NOM
‘Pinooc$_i$ fed him/her/it$_{*i,\ *j}$ Elwit’et$_j$’s food.’

With this established, Deal observes that possessors may raise only from the structurally higher indirect object.

(96) a. *pro* ’ew- ’nii -yey’ -se -∅ Angel -ne pike taaqmaał.
I 3OBJ- give -μ -IMPERF -PRES A. -OBJ mother.NOM hat.NOM
‘I’ve giving Angel’s mother a hat,’ ‘*I’m giving a/the mother Angel’s hat.’

b. 'Aayat -om hi- kiwyek -ey' -se -∅ 'iin -e picpic
woman -ERG 3SUBJ- feed -μ -IMPERF -PRES my -OBJ cat.NOM
cuu'yem.
fish.NOM
'The woman fed my cat the fish,' '*The woman fed a/the cat my fish.'

If we adopt the conclusions of Deal (2013) at face value, then it must also be the case that A-movement can target possessors.

2.2.3.3 A-movement can extract nominals from PP

One need not look far cross-linguistically to find examples of A-movement out of PPs. This can be seen routinely in English passive movement in constructions routinely refered to as 'pseudo-passives,' shown for a variety of prepositions in (97).

(97) a. We frequently rely [on those women].
Those women$_i$ are frequently relied [on t$_i$].

b. I always wait [for Julio].
Julio$_i$ is always waited [for t$_i$].

c. Those girls always gossip [about Brad].
Brad$_i$ is always gossiped [about t$_i$].

d. Several people contributed [to this project].
This project$_i$ was contributed [to t$_i$] by several people.

e. Dogs have eaten [at that meat].
That meat$_i$ has been eat [at t$_i$] by dogs.

In (97), we see that Passivization, a classical A-movement process in English, may extract the complement of a preposition from its containing PP.

Unfortunately, the cross-linguistic picture for A-movement processes out of prepositional phrases is woefully small. This is for two reasons. First, only a small

number of languages allow the prerequisite preposition stranding. But in these languages, A-movement out of PP is often attested. One such case is found in Frisian (Hoeksema 1995, through Abels 2003), shown in (99), while another is the Norwegian in (98), which shows similar prepositional pseudo-passives to English (Lødrup 1991, Truswell 2009).

(98) Norwegian

a. De$_i$ m*å* bli passet bedre p*å* t$_i$.
they must be looked better after
'They must be looked after better.'

b. Han$_i$ ble ledd av t$_i$.
he was laughed at
'He was laughed.'

(99) De bern$_i$ wurdt net nei t$_i$ harke.
the children are not to listened
'The children$_i$ are not listened to t$_i$.' (Frisian, Hoeksema 1995)

The second reason that A-movement out of PPs is rare cross-linguistically is that a variety of languages which generally allow preposition stranding under A'-movement prohibit it under A-movement. This is especially well-attested in the Scandinavian languages besides Norwegian. The Object shift data are from Vikner (2005), while the pseudopassive data are from Truswell (2009). See 2.2 of this chapter for a discussion of how Object Shift is a form of A-movement.

(100) Danish

a. Hvem$_i$ har Peter snakket [med t$_i$]?
who has P. talked with
'Who has Peter talked with?'

b. * Han blev grinet [af t$_i$].
he was laughed at
Intended: 'He was laughed at.'

c. * Hvorfor læste Peter den her bog$_i$ aldrig [i t$_i$]?
why read P. this here book never in
Intended: 'Why did Peter never read in this book?'

(101) Icelandic

a. Hvern$_i$ hefur Pétur talað [við t$_i$]?
who has P. talked with
'Who has Peter talked with?'

b. * Ég tel Vigdísi$_i$ vera oftast talað vel [um t$_i$].
I believe V. be most.often spoken well of
Intended: 'I believe Vigdis to be most often spoken well of.'

c. * Af hverju las Pétur henni$_i$ aldrei [í t$_i$]?
why read P. it never in
Intended: 'Why did Peter never read in this book?'

In the (a) examples of (100-101), we see that both Danish and Icelandic allow preposition stranding under A'-extraction. In contrast, the (b) examples show that these languages do not allow extraction from PPs in passivization[20], while the (c) examples show that complements of prepositions may never undergo Object shift.

All in all, the cross-linguistic picture is complex for a variety of reasons. Despite

[20]The Icelandic example in (101b) is a bit convoluted because Icelandic does superficially allow prepositional passives of the English sort.

(1) Icelandic "prepositional passive" (Maling and Zaenen 1985)

a. Þessa konu$_i$ er oftast talad vel um t$_i$.
that woman.ACC is usually spoken well of
'That woman is usually spoken well of.'

b. Þennan ref hefur aldrei verid skotid á t$_i$.
that fox.ACC has never been shot at
'That fox has never been shot at.'

Despite the appearance of (i), Maling and Zaenen (1985) demonstrate convincingly that these constructions are a form of A'-movement that does not target subject position like a true passive. An important piece of their evidence comes from examples like (100b), where the apparent subject of the supposed prepositional passive fails to undergo further A-movement, such as Raising to Object. This indicates both that these apparent prepositional passives do not target true subject position, and also that this is A'-movement, meaning the violation here is Improper Movement.

this, it is an unescapable fact that A-movement out of prepositional phrases does occur in languages like English, Frisian, and Norwegian.

2.2.3.4 Conclusion of investigation of A-movement out of phases

If these conclusions within this subsection are on the right track, then we are left with a clear conclusion: A-movement can target nominals within phases (Chomsky 2000, 2001). This should not be surprising from a theoretical perspective, as phase theory was designed to allow movement out of phasal categories, so long as certain locality requirements on that movement are met, such as raising first to the specifier of the phase head. Therefore, this ability to escape a phase by A-movement is a predicted consequence (see Zyman 2018).

If Move-and-Reduce is a good model of clitic doubling, then we are left with a clear expectation, codified in the form of a diagnostic diagnostic in (102).

(102) **Diagnostic from locality**: Given a scenario where a relation $\mathcal{R}$ occurs between an agreement-morpheme $\mathcal{M}$ and the corresponding full noun-phrase $\mathcal{F}$, if $\mathcal{F}$ occurs within a phase $\mathcal{P}$ while $\mathcal{M}$ occurs outside $\mathcal{P}$:

a. $\mathcal{M}$ may cross-reference $\mathcal{F} \Longrightarrow \mathcal{R}$ is clitic doubling

b. $\mathcal{M}$ may not cross-reference $\mathcal{F} \Longrightarrow \mathcal{R}$ is consistent with either clitic doubling or Agree

Importantly, the diagnostic from locality only works if the process in question may cross phase boundaries. This is because A-movement processes do not uniformly allow movement out of phasal categories. A clear example of this comes from the English passive. Above, we saw that passive may freely extract a nominal from within an argument PP. These data are repeated in (103) from (97).

(103) a. We frequently rely [on those women].
Those women$_i$ are frequently relied [on t$_i$].

b. I always wait [for Julio].

   Julio$_i$ is always waited [for t$_i$].

c. Those girls always gossip [about Brad].

   Brad$_i$ is always gossiped [about t$_i$].

d. Several people contributed [to this project].

   This project$_i$ was contributed [to t$_i$] by several people.

e. Dogs have eaten [at that meat].

   That meat$_i$ has been eat [at t$_i$] by dogs.

While passivization may extract nominals from PPs, it may not extract DPs from finite CPs. This can be seen in examining the behavior of a predicate like *think*, which may select for both finite and non-finite CPs and may also be passivized.

(104) a. Everyone thought [that Jack was guilty].

b. Everyone thought [Jack to be guilty].

c. It was thought by everyone [that Jack was guilty].

d. * Jack$_i$ was thought by everyone [that t$_i$ was guilty].

e. Jack$_i$ was thought by everyone [t$_i$ to be guilty].

In (104a-b), we see that *think* may select either a finite CP complement or a non-finite complement, yielding an ECM configuration. In (104c) we see that this predicate may also undergo passivization when it selects a finite CP complement. When this occurs, a DP from within the finite embedded clause may not undergo passivization, as seen in (104d). Instead, an expletive must be inserted into the matrix subject position, as in (104c). (104e) demonstrates that passivization may only extract a nominal from a non-phasal, non-finite complement.

Therefore, it is not the case that every A-movement process treats nominals within different phases equally. Stated differently, not all phases pattern identically

with respect to A-movement.

Likewise, some A-movement processes may not extract nominals from any phasal categories. One such A-movement process from English is Subject-to-Subject Raising in (105).

(105) a. It seems ^{PP}[to everyone] [that Mary's father is guilty].

b. * Everyone$_i$ seems [to t$_i$] [that Mary's father is guilty].

c. * Mary's father$_i$ seems [to everyone] [that t$_i$ is guilty].

d. Mary's father$_i$ seems [to everyone] [to be guilty].

e. * Mary$_i$ seems [to everyone] ['s father to be guilty].

In (105a-c) we have a raising predicate, *seem*, which selects a PP experiencer *to everyone* and a finite CP clause. Both of these complements are phasal, and we see that nominals from either cannot undergo Subject-to-Subject Raising. In (105d), we see that a subject embedded within a non-finite complement of *seem*, in this case *Mary's father*, may undergo Subject-to-Subject Raising around the PP experiencer *to everyone*. Importantly, only the entire DP *Mary's father* must raise: the possessor *Mary* may not raise independently.[21]

Therefore, we see that A-movement processes behave differently with respect to extraction from phases. Some A-movement processes, like the English passive, allow extraction only from PP. Others, like Subject-to-Subject Raising, do not allow extraction from any phases. Therefore, if we apply the diagnostic in (102) in isolation and see that a process may not cross phase boundaries, this result is consistent both with φ-agreement, but also possibly clitic doubling.

[21] (105e) has the grammatical parse in (i). This parse is irrelevant for our purposes, as it points to the same conclusion as (105d).

(1) Mary$_i$ seems [to everyone's father] [t$_i$ to be guilty].

Despite this uncertainty, the diagnostic in (102) supports an analysis of clitic doubling in languages with independent evidence for it. One such language is Amharic. In Amharic, clitic doubling may cross-reference nominals within phases. First, nominals embedded within PPs may be doubled, so long as a reduced form of the preposition is doubled as well.

(106) Amharic prepositional object markers (Kramer 2014)

a. Dañña -w **bä-** **Aster**$_i$ färrädä **-bbat**$_i$.
judge -DEF.M against -A. judge.3MS.S **-against.3fs.o**
'The judge judged against Aster (=he convicted her).'

b. Dañña -w **lä-** **Aster**$_i$ färrädä **-llat**$_i$.
judge -DEF.M for -A. judge.3MS.S **-for.3fs.o**
'The judge judged in Aster's favor (=he acquitted her).'

Additionally, as Kramer (2014) observes, "prepositional object markers behave like 'normal' (non-prepositional) object markers ... and supports the analysis of all object markers as clitics in Amharic" (pg. 627-628). The primary argument she provides is that the configurational locality restrictions we see on A-movement apply to these doubled clitics as well. First, she follows McGinnis (2008) and assumes that benefactives are structurally higher than other kinds of applicative arguments, such as instrumentals. With this in place, consider (107).

(107) a. Gɨrma lä- Almaz$_i$ dädʒdʒ -u -n bä- mät'rägiya -w t'ärräg-
G. for -A. doorway -DEF -ACC with- broom -DEF sweep.3MS.S
-ä -llat$_i$.
-BEN -FOR.3FS.O
'Girma swept the doorway with the broom (=instrument) for Almaz (=beneactive).'

b. * Gɨrma lä- Almaz dädʒdʒ -u -n bä- mät'rägiya$_i$ -w
G. for -A. doorway -DEF -ACC with- broom -DEF
t'ärräg -ä -bbät$_i$.
sweep.3MS.S -INST -WITH.3FS.O

Intended: 'Girma swept the doorway with the broom (=instrument) for Almaz (=benefactive).'

In (107), we see that only the structurally higher benefactive argument *lä-Almaz* 'for Almaz' may undergo clitic doubling, not the structurally lower instrumental *bä-mät'rägiyaw* 'with the broom.' In this way, we see that this clitic doubling is subject to the same configurational restrictions as other kinds of A-movement.

Additionally, Kramer (2014) notes that clitic doubling in Amharic is obligatory when the direct object contains an inalienable possessor. This is comparable to Nez Perce as discussed in Deal (2013), where inalienable possessors in direct objects must undergo A-movement.

(108) Bärr -u t'at -e$_i$ -n k'ärät't'äf -ä -**ñn̈**$_i$.
door -DEF finger -my -ACC pinch.3MS.S -1S.O
'The door pinched my finger.' (Kramer 2014)

Therefore, it does seem to be the case that Amharic allows clitic doubling into phases. Again, this receives a natural explanation when we consider the tight connection between clitic doubling and A-movement. Above, we saw that A-movement may extract possessors, particularly inalienable possessors from the structurally highest object (Deal 2013). Therefore, seeing this same pattern replicated in Amharic clitic doubling, despite this language lacking canonical external possession constructions (Haspelmath 1999), provides welcome confirmation that the diagnostic in (102) is on the right track.

2.2.4 Property of clitic doubling #4: Semantic restrictions

It has long been observed that some A-movement constructions are sensitive to the definiteness or specificity of the nominal. Of particular importance with respect to the literature on clitic doubling is object shift.

Object shift is a phenomenon in North Germanic languages, especially Icelandic, in which definite and specific objects occur in a demonstrably higher position

than indefinite and nonspecific objects (Holmberg 1986, Diesing 1992, 1997, Vikner 1990, 1994, 2005, Collins and Thráinsson 1996, Bobaljik and Thráinsson 1998, Thráinsson 2001, 2007, inter alia). Consider the position of the object in (109) with respect to negation.

(109) Object Shift in Icelandic (Thráinsson 2001)

a. Nemandinn las ekki **bókina**.
student.the read not book.the
'The student didn't read the book.'

b. Nemandin las **bókina**$_i$ ekki t_i.

c. *Nemandinn las ekki **hana**.
student.the read not it
Intended: 'The student didn't read it.'

d. Nemandin las **hana**$_i$ ekki t_i.

e. Hún keypti ekki **kaffi**.
she bought not coffee
'She didn't buy coffee.'

f. * Hún keypti **kaffi**$_i$ ekki t_i.

In (109a-b), the definite DP object *bókina* 'the book' may occur either in its base position to the right of negation or in a derived position to its left. This derived position is not available for indefinite objects like *kaffi* 'coffee' as shown in (109e-f), but it is obligatory for pronominal objects like *hana* 'it' in (109c-d). In this way, we see that object shift is closely tied to specificity and refentiality (see Diesing 1992, Alexiadou and Anagnostopoulou 1997, and Vikner 2005 for an overview).

Much work, particularly on Icelandic and Danish, has demonstrated that object shift is A-movement. One piece of evidence is that object shift fails to license a parasitic gap while A'-movement does.

(110) Parasitic gaps in Danish (Vikner 2005)

a. Hvad for en bog$_i$ stillede alle t_i hen på reolen [uden at læse
which book put all onto bookcase-the without to read
PG$_i$ først]?
first
'Which book did everyone put on the shelf without reading first?'

b. *Alle stillede den$_i$ straks t_i hen på reolen [uden at læse PG$_i$
all put it at.once onto bookcase-the without to read
først]?
first
Intended: 'They all put it onto the bookcase without reading PG first.'

In (110a), we see that wh-movement licenses a parasitic gap. This is the expected pattern since Engdahl (1983). In contrast, (110b) demonstrates that object shift fails to license a parasitic gap. This is one indication that object shift is not A'-movement like the wh-movement in (110a), but A-movement.

Therefore, we see that object shift is an A-movement process which only targets specific or referential nominals. This means that if clitic doubling is derived by A-movement, we might expect clitic doubling processes cross-linguistically to be sensitive to these same concepts of specificity and referentiality.

As it turns out, there is a tight parallelism between the nominals that may be clitic doubled and the nominals that can undergo object shift in languages like Icelandic. This was first observed in Uriagereka (1995). Similar ideas were put forward in Sportiche (1996), Alexiadou and Anagnostopoulou (1997), Suñer (2000), Kallulli (2000), and by the time of Nevins (2011), the isomorphism between clitic doubling and object shift was well-established. To see this, compare (111) to (109).

(111) Clitic Doubling in Rioplatense Spanish (Jaeggli 1982)

a. ☒ Vimos a Guille.
we.saw G.
'We saw Guille.'

b. **Lo**$_i$ vimos a Guille$_i$.
him we.saw G.

'We saw Guille.'

c. *⊠ Vimos a él.
we.saw him
Intended: 'We saw him.'

d. **Lo**$_i$ vimos a él$_i$.

e. No ⊠ ví a ningún chico.
NEG I.saw no boy
'I didn't see any boys.'

f. *No **lo(s)**$_i$ vi a ningún chico$_i$.
NEG him I.saw no boy

In (111), we see an identical distribution between clitic doubling in Spanish and object shift in Icelandic.[22] In (111a-b) we see that clitic doubling is optional with a definite object, in this case the proper name *Guille*. This matches the optionality of object shift with a definite object in (109a-b). Likewise, (111c-d) show that clitic doubling is obligatory with pronominal objects, just as object shift is obligatory for pronominal objects (109c-d). Finally, (111e-f) demonstrate that an indefinite object, like the negative DP *ningún chico* 'no boy' may not be clitic doubled, just as an indefinite object in Icelandic cannot undergo object shift (109e-f).[23]

This sensitivity to specificity is not a quirk of Rioplatense Spanish. Amharic, Modern Greek, and Castillian Spanish each demonstrate a similar restriction on clitic doubling.

[22]I refer only to accusative clitic doubling, not to dative clitic doubling, which has been argued to be a form of agreement. See Uriagereka (1995) and Bleam (2000).

[23]Baker and Kramer (2018) observe that the parallelism between object shift and clitic doubling in Amharic is not perfect. Specifically, they observe that universally quantified DPs may undergo object shift in Icelandic, but these cannot be clitic doubled in Amharic. From this, they conclude that clitic doubling in Amharic cannot be derived by A-movement, unlike in other languages.

This conclusion seems premature to me for two reasons. First, as we will see in Chapter 5, universally quantified DPs can refer to members of a contextually salient set, or not. This sort of fine-grained pragmatic detail is not provided in the data they discuss. Second, it is well-documented that languages differ in terms of which nominals count as specific or referential enough for particular processes. See, for instance, the exchange between Cinque (1990) and Chung (1994), as well as the typology of differential object marking in Aissen (2003). Therefore, the details of the behavior of one A-movement process in Icelandic do not strike me as relevant in Amharic. Rather, the overall parallelism is certainly still present, and I consider this the most important aspect.

(112) Amharic (Kramer 2014, pg. 601)

a. Almaz doro wät' bäll-tʃtʃ -☒.
A.F chicken stew eat -3FS
'Almaz ate chicken stew.'

b. Almaz doro wät' -u -n bäll -tʃtʃ -**i̶w**.
A.F chicken stew -DEF.M -ACC eat -3FS.S -3MS.O
'Almaz ate **that** chicken stew.'

(113) Modern Greek (Alexiadou and Anagnostopoulou 1997)

a. **To**$_i$ diavasa to vivlio$_i$ me prosohi.
it.ACC read.1SG the book.ACC carefully
'I read the book carefully.'

b. ***To**$_i$ diavasa kapjo vivlio$_i$ me prosohi.
it.ACC read.1SG some book.ACC carefully
Intended: 'I read some book carefully.'

(114) Castillian Spanish

a. Pedro la vio.
P. her saw
'Pedro saw her.'

b. Pedro la vio a María.
P. her saw M.
'Pedro saw María.'

"Assuming that the individual referred to by *la* in [114a] and *María* in [114b] is the same, sentences [114a] and [114b] are truth-conditionally equivalent. Nevertheless, they are not equivalent at a pragmatic level. In some dialects, sentence [114b] has a distinctive emphatic character." (Gutiérrez-Rexach 1999, pg. 318, footnote 6)

A Move-and-Reduce approach to clitic doubling provides a natural explanation for these data. As Icelandic object shift demonstrates, only specific or referential

nominals undergo some A-movement processes. If one of these discourse sensitive A-movement processes is the source of movement for these clitic doubling processes, then this effect is a predicted consequence.

At the same time, few solid conclusions can be reached if a process is found to show semantic restrictions. This is for two reasons. First, not all A-movement processes show semantic restrictions like object shift. For instance, English Passivization affects all DPs, no matter their specificity.

(115) a. Maria$_i$ was kissed $\_\__i$.

b. Some women$_i$ were kissed $\_\__i$.

c. A woman$_i$ was kissed $\_\__i$, but I don't know who.

d. Nobody$_i$ was kissed $\_\__i$.

In (115), we see that a variety of DPs undergo Passivization, no matter how specific or referential they are. Therefore, if a process does not demonstrate semantic restrictions, it could still be derived by A-movement, albeit a different kind of A-movement than object shift.

In addition, as we will explore extensively in chapter 5, φ-agreement systems, most notably topic agreement systems, frequently demonstrate semantic restrictions as a by-product of the semantic restrictions on topics. Therefore, even if a process does demonstrate semantic restrictions of the sort discussed in this subsection, it could still be the case that that process is derived by **?**.

2.2.5 Property of clitic doubling #5: Morphological similarity to definite determiners

Uriagereka (1995), developing earlier unpublished work by Torrego (1988), notices that doubled clitic and determiners are morphologically identical in various Iberian Romance languages. Consider the following data from Galician.

(116) Galician nominal doubles and definite determiners (Uriagereka 1995, pg. 81)

	Doubled clitic	Determiner
Masculine/singular	o	o
Feminine/singular	a	a
Masculine/plural	os	os
Feminine/plural	as	as

Furthermore, Uriagereka contends that this morphological identity is not accidental. Rather, it reflects that doubled clitics are of category D, like determiners. Since this work, morphophonological identity with elements of category D has been taken as evidence for clitic doubling (see Kramer 2014 for a review).

At the same time, this kind of morphophonological diagnostic is famously unreliable (see Nevins 2011 especially). A particularly clear example comes from Yuan's (2017a) cross-dialectal investigation of Inuit languages. Each of these languages shows intricate verbal morphology in which subject and object nominal doubling form a portmanteau with mood. (117-118) demonstrate the Inuktitut form.

(117) a. Taku -**jara**.
see -1S.S/3S.O.DECL
'I saw her.'

b. Taku -**jait**.
see -2S.S/3S.O.DECL
'You saw her.'

c. Taku -**jarma**.
see -2S.S/1S.O.DECL
'You saw me.'

(118) a. Taku -**vigu**?
see -1S.S/3S.O/INT
'Did I see her?'

b. Taku -**viuk**?
see -2SG.S/3S.O/INT
'Did you see her?'

c. Taku -**vinga**?
see -2S.S/1S.O/INT
'Did you see me?'

As can be seen in (117-118), these languages have intricate morphology that cross-references both the subject and the object while simultaneously showing allomorphy for Mood. Importantly, Yuan argues convincingly from syntactic evidence that in some Inuit languages, such as Kalaallisut (formerly known as West Greenlandic), this mor-

phology is a form of φ-agreement, while in the closely related Inuktitut, this same morphology is best treated as φ-agreement with the subject but clitic doubling of the object. At the same time, the morphology between these two languages is nearly identical, aside for regular sound correspondences.

(119) Inuktitut

a. **-jara** '1S.S/3S.O/DECL'

b. **-jait** '2S.S/3S.O/DECL'

c. **-jarma** '2S.S/1S.O/DECL'

(120) Kalaallisut

a. **-vara** '1S.S/3S.O/DECL'

b. **-vait** '2S.S/3S.O/DECL'

c. **-varma** '2S.S/1S.O/DECL'

In addition to these empirical concerns, there is a conceptual argument against appealing to morphophonological diagnostics. For instance, it is a well-known that φ-agreement systems frequently develop from clitic doubling (Givón 1971, Lehmann 1988, Corbett 1995, 2006, Hopper and Traugott 1993, Roberts and Roussou 2003, Fuß 2005). Therefore, morphophonological similarity between morphemes that double the φ-features of an argument and definite determiners could reflect a diachronic connection between the two, rather than provide evidence for a synchronic analysis.

2.3 Summing up the differences between φ-agreement and clitic doubling

Based on the investigations from the previous two sections, we can establish the following diagnostics to distinguish φ-agreement from clitic doubling. These are shown in (121), repeated from (1) above, with their respective rankings.

(121) Ranking of diagnostics to distinguish clitic doubling from φ-agreement

		φ-agreement	Clitic doubling
High reliability:			
	a. Ability to affect variable binding?	**x**	✓
	b. Extraction from coordinate structures?	✓	**x**
Medium reliability:			
	c. Obeys the PIC?	✓	✓/**x**
	d. Presence of a default?	✓/**x**	**x**
Low reliability:			
	e. Sensitivity to specificity?	**x**/✓	**x**/✓
	f. Only one per clause?	**x**/✓	**x**/✓
	g. Semantic restrictions?	**x**/✓	**x**/✓
	h. Morphological similarity to D?	**x**/✓	**x**/✓

The core and most reliable of these diagnostics are derivative of the central analysis of clitic doubling and φ-agreement: φ-agreement is derived by AGREE, which does not form a movement chain, while clitic doubling is derived by a movement chain. Therefore, to argue for or against clitic doubling, one must examine if movement can be diagnosed in the relevant construction.

With this, we are left with two (and a half) logical possibilities to derive any multiple exponence of φ-features. These are presented in (122).

(122) Multiple exponence of φ-feature bundles can be derived by:

a. Providing a value to some node's unvalued φ-features (φ-agreement).

b. Spell-out of multiple copies in a movement chain (clitic doubling).

i. This movement chain could be formed by A-movement.

ii. This movement chain could be formed by A'-movement.

Now that we have identified the two (and a half) possible analyses of pronoun doubling in SMP Mixtec, let us begin the examination of its properties. Naturally, this requires a close examination of movement processes in this language.

Chapter 3

Diagnosing A-movement in San Martín Peras Mixtec

In chapter §2, we saw that the hallmark of clitic doubling systems cross-linguistically is that they form an A-chain. See Sportiche (1996), Alexiadou and Anagnostopoulou (1997), Preminger (2009), Harizanov 2014b,a, Kramer (2014), *inter alia*. Therefore, if pronoun doubling in SMP Mixtec is derived through clitic doubling of the same sort, we expect to likewise find the properties of A-movement in these constructions. Naturally, the gold-standard is to find A-movement processes internal to this language, and compare their properties to those of pronoun doubling. This is especially the case because, as we saw in §2, A-movement shows a wide range of variation wining and across languages. As such, we expect to see that variation mirrored in clitic doubling constructions, if they are truly to be understood in terms of A-movement.

That said, finding A-movement in this language is easier said than done. This is because SMP Mixtec lacks all of the canonical raising constructions found in more familiar languages. First, SMP Mixtec lacks a passive. Instead, this language uses a construction which I call the *nà*-impersonal construction. These are roughly equivalent to the English passive, or the *se*-passive of Spanish (see Suñer 1976 and D'Alessandro

2004 for a more recent discussion).

(1) a. Ndànuûn nà tsǐn nuhŭn *kama*.
put.on.SG.PAST they mouse on bed
'The mouse was put on the bed.'

b. Nì ixăni rà xà'ani nà rà.
PAST dream he kill.PAST they him
'He dreamed that he was killed.'

(1) exemplifies the *nà*-impersonal. Here, we see that subject position is occupied by a nonreferential *nà* 'they,' yielding an impersonal interpretation. Therefore, the language lacks any kind of passive-like construction with displacement.

Additionally, SMP Mixtec is rigidly XVS(O) in all clauses, with virtually all adjuncts occurring preverbally. Consider (2), which summarizes the discussion from §1.1.2 in chapter 1.

(2) a. **Bàko'ĕ** kú'u (***bàko'ĕ**) rà mástro (***bàko'ĕ**).
often be.sick.PRES he teacher
'The teacher is often sick.'

b. **Inkatúkú** kànakààba (***inkatúkú**) ñá lo'o (***inkatúkú**).
again fall.PAST she little
'The girl fell again.'

c. Rà$_i$ **xa** ntùba'á (***xa**) rà lo'o$_i$ (***xa**).
he PERF get.better.PAST he little
'The boy has already gotten better.'

d. Rí$_i$ **skă** nì xi'i (***skă**) ndxùxi$_i$ (***skă**).
it.AML just PAST die hen
'The hen just died.'

e. **Táto'oba** íntu'u (***táto'oba**) ñá yata (***táto'oba**).
probably be.sitting.PRES.SG she old
'The old woman is probably sitting.'

f. **Táxakán** ndàkinúú (***táxakán**) míí *láncha* (***táxakán**).
then sink.PAST the boat

'Then the boat sank.'

(2) demonstrates a variety of adjuncts in SMP Mixtec from the lowest to the highest projections of the hierarchy in Cinque (1999) coupled with a variety of unaccusative predicates.[1] These range from low adjuncts like *bàko'ĕ* 'often' and *inkatúkú* 'again' in (2a-b) to high adjuncts like *táto'oba* 'probably' and *táxakán* 'then' in (2e-f). In each case, we see that the adjunct must be initial, Therefore, using adjunct placement as a benchmark for clausal structure is not straightforward. This makes it difficult to identify any A-movement of unaccusative subjects of the EPP-driven sort common to languages like English.

Finally, SMP Mixtec lacks all canonical Subject Raising predicates that are common cross-linguistically. Rather, the language uses other syntactic devices to express similar ideas. Consider (3).

(3) a. Ñá$_i$ keba'a ñá Maria$_i$ ká' ìn.
she win.IRR she M. think.PRES I
Provided for 'It seems that Maria will win,' lit. 'I think that Maria will win.'

b. Ná'a xíka kwê'e íyo yúkŭ.
look.PRES far very is.PRES.SG mountain
'The mountains seem to be very far away,' lit. 'The mountains look like they are far away.'

c. Kìnì ná' ì bítsí.
ugly look.PRES I now
'I look ugly now.'

d. Táto'oba keba'a ñá Maria.
probably win.IRR she M.
'It is likely that Maria will win,' like 'Maria will probably win.'

e. Kìxă rí leso taxa'a rí.
begin.PAST it.AML rabbit dance.IRR it.AML
'The rabbit started/began to dance.'

[1]See chapter 4 for arguments that these predicates are unaccusative.

In (3a-b), we see the lexical tools SMP Mixtec has to express 'seem.' These involve either a more direct 'I think' statement, as in (3a), or the predicate *na'a*, which speakers translate as *mirarse*, 'to look' rather than 'seem.' This is supported by its use in (3c).

In (3d), we see that SMP Mixtec expresses *es probably* 'it is likely' by using the same adjunct *táto'oba* 'probably' that we saw in (2e). Finally, predicates like *kixă* 'start,' which can involve Subject Raising in English, demonstrably involve control. Control structures, and the syntax of embedded clauses more broadly, will be discussed in more detail below in §3.1.

The lack of such A-processes makes it difficult to generate expectations for what characteristics pronoun doubling should display if it were derived through A-movement. That said, despite lacking cross-linguistically common instances of A-movement, SMP Mixtec has a different movement process which I will demonstrate below exhibits many of the core properties of A-movement. I term this quantifier fronting. Consider (4).

(4) a. Íntoso mií sâ nda'ă mií ítŭn kân.
be.on.PRES.PL the bird in the tree that
'The birds are in that tree.'

b. [Nda'ă iin ítŭn]$^{\exists}$ íntoso [ntskû sâ]$^{\forall}$ __.
in one tree be.on.PRES.PL all bird
'All the birds are in a tree.' ($\exists > \forall$)

c. [Ntskû sâ]$^{\forall}$ íntoso __ [nda'ă iin ítŭn]$^{\exists}$
all bird be.on.PRES.PL in one tree
'All the birds are in a tree.' ($\forall > \exists$)

(4a) demonstrates the basic word order of SMP Mixtec: the verb comes first, followed by the subject and then the complements, in this case the prepositional phrase *nda'ă mií ítŭn kân* 'in that tree.' In (4b-c), we see that quantified expressions are fronted to preverbal position, where they obligatorily take scope over any other quantificational elements in the sentence.

In this chapter, I will demonstrate that quantifier fronting is a form of A-movement. This allows us to examine its particular properties to see if they can be replicated in pronoun doubling. Ultimately, we will see that they cannot, and pronoun doubling systematically fails to show the core properties of an A-chain. While this is not fatal to a Move-and-Reduce clitic doubling analysis of pronoun doubling, as this technology does not require that the prerequisite movement be A-movement, it would set pronoun doubling apart from other clitic doubling systems cross-linguistically.

This chapter is organized as follows. First, I provide the necessary syntactic background on embedded clauses in this language in §1. This includes an investigation of embedded irrealis clauses, of which I argue some are control structures. With this necessary clausal architecture established, §2 presents the A-movement properties of quantifier fronting, as well as other distinguishing characteristics it displays language-internally. §3 then applies this insight to pronoun doubling, to see if pronoun doubling can be demonstrated to show the characteristic signature of an A-chain in this language. Importantly, I show that it cannot. §4 concludes.

3.1 A necessary excursus into the syntax of embedded clauses

It is often[2] true that A-movement cannot extract nominals from finite clauses.[3] Consider the classic English contrast in (5).

(5) a. Maria$_i$ seems [__$_i$ to be a genius].

b. * Maria$_i$ seems [that __$_i$ is a genius].

In (5), we see that Subject-to-Subject Raising in English applies freely out of nonfinite

[2]But not universal. See the discussion in 2.3.1 in chapter 2

[3]This work was presented at the 22nd Workshop on the Structure and Constituency in Languages of the Americas at the University of British Columbia on April 22nd, 2017. See Ostrove (2017a). I acknowledge and thank Ivy Sichel for her guidance and interest in this work.

clauses, but may not apply out of finite CPs. Therefore, in our search for A-movement in SMP Mixtec, it would help to understand the syntax of embedded clauses.

SMP Mixtec, like many other predicate-initial languages, lacks non-finite verbal morphology (see Myhill 1985, and the discussion in Macaulay 2005 of Chalcatongo Mixtec). Instead, embedded clauses in which the verb is obligatorily marked irrealis take the place of non-finite clauses in other languages. Recall from chapter 1 that TAM morphology is primarily tonal.

(6) a. Ká'an ñá [{ kèba'a, kéba'a, keba'a } ñá].
think.PRES she win.PAST win.PRES win.IRR she
'She thinks she { won, is winning, will win }.'

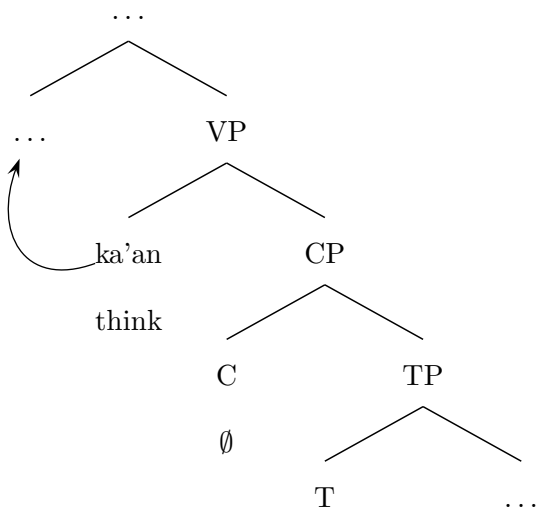
b. Kôni ñá [{ *kèba'a, *kéba'a, ✓ keba'a } ñá].
want.PRES she win.PAST win.PRES win.IRR she
'She wants to win.'

c. Kìxă ñá [{ *kèba'a, *kéba'a, ✓ keba'a } ñá].
start.PAST she win.PAST win.PRES win.IRR she
'She started to win.'

In(6a), the embedded clause may surface with any of the three TAM categories in this language. This corresponds, as the translation indicates, to a finite embedded clause. In(6b-c) under *koni* 'want' and *kixă* 'start/begin,' the embedded clause looks largely the same as the finite embedded clause in (6a): we have an embedded subject, and an embedded verb with finite TAM morphology. Crucially, though, the verb must be marked irrealis.

This leads to an interesting question of analysis. For example, it seems clear that the finite embedded clause in (6a) should be analyzed as in (7).

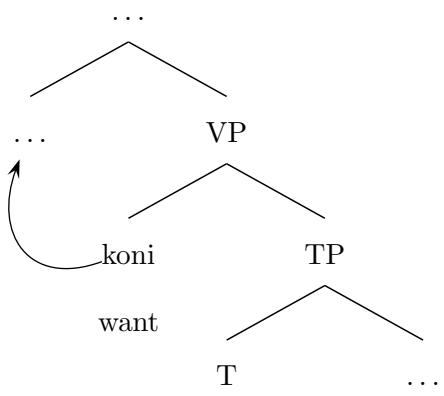
(7)



In (7), the clause embedded under *ka'an* 'think' is analyzed largely as it would be in English. This seems right, given that the two mean the same thing, and the tense value of the lower predicate is independent of the higher, as reflected by the morphology.

How should we analyze the embedded clauses in (6b-c) is not obvious. It could be that all embedded clauses are CPs. Alternatively, it could be that the morphological defectiveness of (6b-c) implies a structural difference, as in (8), in which the embedded clause is sub-phasal.

(8)



The transparency implied by the sub-phasal boundary in (8) might be used to

explain the obligatory irrealis marking observed in (6b-c).

I will argue that, in fact, embedded irrealis clauses like those in (6b-c) come in two varieties, depending on whether Obligatory Control occurs into them. For predicates which require Obligatory Control, (8) is the only available structure. Interestingly, not all predicates which require Control in English do so in SMP Mixtec. For these predicates when Obligatory Control does not occur, I propose that the structure in (7) is most appropriate. That is, SMP Mixtec patterns with Balkan languages like Bulgarian and Modern Greek, which also show a split in the syntax of embedded subjunctives. See Terzi (1992), Dobrovie-Sorin (1994), Krapova (1998, 2001), Landau (2004), *inter alia*.

To make this case, we will consider the classical interpretive tests for Obligatory Control: the availability of *de re* interpretations in §3.1.1, the possibility of strict readings under ellipsis in §3.1.2, and whether local binding is obligatory in §3.1.3. To preview the discussion, we will see that embedded irrealis clauses come in two varieties. The first require local binding and are interpreted like PRO structures in English. These occur under predicates like *kixă* 'begin/start.' The second do not require local binding, and their subjects are instead interpreted like regular pronouns. These occur under predicates like *koni* 'want.' I will refer to these as control-type irrealis clauses and *pro*-type irrealis clauses, respectively.[4] (9) provides a list of predicates that select each type of irrealis clause.

(9) a. **Control-type irrealis**: *kixă* 'start/begin,' *santsi'i* 'finish/end,' *nantŏso* 'forget'

b. ***pro*-type irrealis**: *koni* 'want,' *kachi* 'promise,' *nakwatu* 'pray'

c. **Finite embedded clause**: *ka'an* 'think,' *ka'a* 'tell,' *kachi* 'say,' *kundàà ini* 'wonder, realize,' *xini* 'know'

[4]Control-type and *pro*-type irrealis clauses match up perfectly with C-subjunctinves and F-subjunctives in the Balkan literature. I find this terminology a bit more transparent.

3.1.1 Availability of *de re* interpretations

In well-studied languages with Obligatory Control structures, we frequently see a restricted range of interpretations associated with PRO versus a regular pronoun. Consider the contrast between (10) and (11).

(10) ***De re* context:** Maria has had a successful career as a professional author for many years. One day, she is cleaning out some old desk drawers when she stumbles across a manuscript without an author listed. She decides to read it though, and she enjoys it very much. By the time she finishes, she has no idea who the author is. Unbeknownst to her, Maria was in fact the author of the manuscript! She says out loud: "The person who wrote this is a great author!"

a. Did Maria$_i$ claim [PRO$_i$ to be a great author]? **No**. (**x** *de re*)

b. Did Maria$_i$ say [that she$_i$ was a great author]? **Yes**. (✓ *de re*)

(11) ***De se* context**: Maria has had a successful career as a professional author for many years. One day, she is cleaning out some old desk drawers when she stumbles across a manuscript without an author listed. She decides to read it though, and she enjoys it very much. By the time she finishes, Maria realizes that she herself was in fact the author of the manuscript! She says out loud: "I am a great author!"

a. Did Maria$_i$ claim [PRO$_i$ to be a great author]? **Yes**. (✓ *de se*)

b. Did Maria$_i$ say [that she$_i$ was a great author]? **Yes**. (✓ *de se*)

In the (a) examples of (10-11), we see that PRO requires an interpretation involving knowledge of self, i.e., a *de se* interpretation. In contrast, in the (b) examples, we see that a regular pronoun supports both a *de re* and a *de se* reading. The inability to bear a *de re* interpretation is a widely reported property of Obligatory Control structures cross-linguisticallly, and serves as a common diagnostic for Obligatory Control. See

Landau (2013) for discussion.

In SMP Mixtec, pronouns in embedded finite clauses are also compatible with both *de re* and *de se* interpretations. This is shown in (12-13).[5]

(12) ***De re* context**: Maria is a professional writer who has written many books and papers in the course of her career. One day, she is cleaning out a desk drawer and she finds a manuscript with no author listed. It has an interesting title, though, so she sits down to read it. Unbeknownst to Maria, she is in fact the author of the paper! She says out loud:

a. Ba'á tsyáa nà yó'o!
well write.PRES they this
'This person writes well!'

b. Á ntàà yá yó'o? Káchi ñá Maria$_i$ [ba'á tsyáa ñá$_i$]?
Q true it this say.PAST M. well write.PRES she
'Is this true? Maria said that she writes well.'

c. Ahan.
'Yes.' (✓ *de re*)

(13) ***De se* context**: Maria has had a successful career as a professional author for many years. One day, she is cleaning out some old desk drawers when she stumbles across a manuscript without an author listed. She decides to read it though, and she enjoys it very much. By the time she finishes, Maria realizes that she herself was in fact the author of the manuscript! She says out loud:

a. Ba'á tsyáa ì!
well write.PRES I
'This (person) writes well!'

b. Á ntàà yá yó'o? Káchi ñá Maria$_i$ [ba'á tsyáa ñá$_i$]?
Q true it this say.PAST M. well write.PRES she
'Is this true? Maria said that she writes well.'

[5]This elicitation methodology was inspired by Pearson (2013, 2015).

c. Ahan.

'Yes.' (✓ *de se*)

In (12) and (13), we see that regular pronouns can support both *de re* and *de se* interpretations, just as in English. With this in mind, recall the form of Control-type and *pro*-type irrealis clauses, reproduced in (14a) and (14c) respectively.

(14) a. Nàntŏso ñá Juana [nakatsya **ñá** míí tsyàà].
forget.PAST she J. wash.IRR she the clothes
'Juana forgot to wash the clothes.'

b. * Nàntŏso ñá Juana [nakatsya __ míí tsyàà].

c. Kôni rà Sergio [kuxi chá ga **rà** ndùchi].
want.PRES he S. eat.IRR very more he bean
'Sergio wants to eat more beans.'

d. * Kôni rà Sergio [kuxi chá ga __ ndùchi].

In (14), we are reminded that both Control-type and *pro*-type irrealis clauses obligatorily have overt pronominal subjects. Therefore, we might expect these elements to have the same range of interpretations as the pronoun in (12-13).

Interestingly, this prediction only partially pans out. First, the pronominal subject of the *pro*-type subjunctive in (14c) is compatible with both a *de re* and a *de se* interpretation, just like a regular pronoun. This is shown in (15-16).[6]

[6] *Ndúxa* 'should,' generally supports bouletic contexts. (i) and (ii) were inspired by several contexts presented in Hacquard (2011). Many thanks to Pranav Anand for pointing out this potential confound to me.

(1) Yô'o xa **ndúxa** ku'u kusi -k ú bítsí.
you PERF should go.IRR sleep.IRR -INTR you now
'You should go to sleep now.' (*Context: A father is very stern, and says to his child:*)

(2) Kwa'ă **ndúxa** kachûn rà Juán áto kò kôni ka rà kachûn rà náxihnu rà 45.
much should work.IRR he J. if NEG want.PRES EMPH he work.IRR he aged.IRR he 45
'Juan should work hard if he doesn't want to work when he's 45.' (*Context: Juan is young and wants to retire at the age of 45.*)

(15) ***De re* context with *pro*-type irrealis**: Sergio goes to a party one evening. Later, he is looking through some pictures taken at the party, and he sees someone in the background who looks noticeably skinny. This person's face is obscured though. What Sergio does not know is he is looking at a picture of himself! Not knowing the identity of the thin individual, he says out loud:

a. Ndúxa kuxi chá ga ñà kan ndúchi.
should eat.IRR very more they that beans
'That person should eat more beans.'

b. Á ntàà yá yo'o? Káchi rà Sergio$_i$ [kôni rà$_i$ [kuxi chá ga rà ndúchi]]?
Q true it this say.PAST he S. want.PRES he eat.IRR very more he beans
'Is this true? Sergio said that he wants to eat more bean.'

c. Ahan.
'Yes.' (✓ *de re*)

(16) ***De se* context with Control-type irrealis:** Context: Sergio goes to a party one evening. Later, he is looking through some pictures taken at the party, and he sees someone in the background who looks noticeably skinny. This person's face is obscured. Despite this, he recognizes by the clothes the thin person is wearing that he is looking at a picture of himself. He says out loud:

a. Ndúxa kuxi chá ga ì ndúchi.
should eat.IRR very more I beans
'I should eat more beans.'

b. Á ntàà yá yo'o? Káchi rà Sergio$_i$ [kôni rà$_i$ [kuxi chá ga rà nduchi]]?
Q true it this say.PAST he S. want.PRES he eat.IRR very more he beans
'Is this true? Sergio said that he wants to eat more bean.'

c. Ahan.
'Yes.' (✓ *de se*)

In (15-16), we see that the pronominal subject of the embedded *pro*-type irrealis clause is interpreted like a regular pronoun. This is not totally surprising, given that we see a morphologically overt pronoun within the irrealis clause.

More interestingly, the overt pronominal subject of Control-type irrealis clauses is incompatible with a *de re* interpretation. Like PRO in English, it forces a *de se* interpretation. This is demonstrated in (17-18).

(17) ***De re* context with Control-type irrealis**: Juana is the manager of a large used clothing store with only a few employees. They keep a tight schedule in order to ensure that everything gets done in a day. One of the chores that needs to be done every day is to wash new donations in a washing machine in the back of the building. One morning, she sees that someone forgot to wash the new donations from the day before. She is mad about the mistake, and goes back to look at the security cameras to see who is at fault. She sees someone walk right past the pile of obviously dirty clothes at the end of the day! She cannot see the person's face, so she doesn't know that, in fact, she herself made the mistake! She says out loud, angerly:

a. Nàntǒso nà kan [nakatsya nà míí tsyàà]!
forget.PAST they that wash.IRREAL they the clothes
'That person forgot to wash the clothes!'

b. Á ntàà yá yo'o? Káchi ñá Juana$_i$ [nàntǒso ñá$_i$ [nakatsya ñá$_i$ míí tsyàà]]?
Q true it this say.PAST she J. forgot.PAST she wash.IRREAL she the clothes
'Is this true? Juana said that she forgot to wash the clothes.'

c. Ú'un.
'No.' (**x** *de re*)

(18) ***De se* context with Control-type irrealis**: Juana is the manager of a large used clothing store with only a few employees. They keep a tight schedule in

order to ensure that everything gets done in a day. One of the chores that needs to be done every day is to wash new donations in a washing machine in the back of the building. One morning, she sees that someone forgot to wash the new donations from the day before. She is mad about the mistake, and goes back to look at the security cameras to see who is at fault. She sees someone walk right past the pile of obviously dirty clothes at the end of the day! She recognizes from the video that the culprit is herself. She says out loud:

a. Nàntŏs ì [nakats ì míí tsyàà]!
forget.PAST I wash.IRREAL I the clothes
'I forgot to wash the cloths!'

b. Á ntàà yá yo'o? Káchi ñá Juana$_i$ [nàntŏso ñá$_i$ [nakatsya ñá$_i$ míí tsyàà]]?
Q true it this say.PAST she J. forgot.PAST she wash.IRREAL she the clothes
'Is this true? Juana said that she forgot to wash the clothes.'

c. Ahan.
'Yes.' (✓ *de se*)

In (17), we see that the pronominal subject of a Control-type irrealis clause patterns exactly like PRO in (10a) in being incompatible with a *de re* interpretation. Likewise, (18) shows that a *de se* context is licit, again, just as with PRO.

Therefore, we have an important interpretative distinction between *pro*-type and Control-type irrealis clauses. As their names suggest, the pronominal subject of the *pro*-type irrealis is compatible with the range of interpretations expected of a regular pronoun, while Control-type irrealis clauses are compatible, like PRO, only with *de se* readings.

3.1.2 Availability of strict interpretation under ellipsis

Another commonly observed difference between pronouns and PRO in Obligatory Control constructions is their behavior under verb phrase ellipsis, or VPE. Consider the contrast in (19).

(19) a. Maria wants PRO to win the race, and Julio does [want PRO to win the race] too.
(✓sloppy, **x** strict)

b. Diego thinks that he is a genius, and Julio does [think that he is a genius] too.
(✓sloppy, ✓strict)

(19a) can only mean that Julio wants for himself to win the race too, i.e., the so-called "sloppy" interpretation. This contrasts with the interpretation of the regular pronoun within the ellipsis site in (19b), which can mean either that Julio thinks he himself is a genius, or that Julio thinks Diego is a genius too. Both the sloppy and the "strict" interpretations are available in (19b).

This yields a prediction. If the pronominal subject of a Control-type irrealis clause really behaves like PRO, then it should only be compatible with a sloppy interpretation under ellipsis. In contrast, if the pronominal subject of a *pro*-type irrealis clause really is interpreted like a regular pronoun, it should support either a strict or a sloppy reading.

SMP Mixtec provides the opportunity to test this prediction because it has VPE. This is shown in (20).

(20) Kuxi rà Macario kolo, [să =ti ísi'i rà ba].
eat.IRREAL he M. turkey like.this =also wife his EMPH
'Macario will eat turkey, and his wife will too.'

As seen in (20), VPE in SMP Mixtec involves the initial verb being replaced by *sǎ*, which speakers usually translate into Spanish as *entonces* 'then' or *así* 'like this.' The verbal clitic *=ti* 'also' cliticizes to *sǎ*, which is natural given that *sǎ* takes the place of the verb. Finally, the emphatic marker *ba* occurs finally. It is unclear if *ba* modifies the nominal that survives ellipsis or the entire second conjunct, and for our purposes it does not matter.

Pronouns in embedded finite clauses in SMP Mixtec behave like English pronouns in supporting both strict and sloppy interpretations. This is shown in (21).

(21) Ká'an rà Diego lísto ba'ǎ rà, sǎ =ti rà Julio ba.
think.PRES he D. clever quite he like.this =also he J. EMPH
'Diego thinks that he is quite clever, and so does Julio.' (✓sloppy, ✓strict)

Having established the two types of ambiguity of (21), we can ask if irrealis clauses show a similar ambiguity. Consider (22).

(22) Kôni ñá Maria$_i$ [keba'a ñá$_i$], [sǎ =ti rà Julio ba].
want.PRES she M. win.IRR she like.this =also he R. EMPH
'Maria wants to win, and Julio does too.' (✓*strict*, ✓*sloppy*)

The pronominal subject of the irrealis clause embedded under *koni* 'want' in (22) is compatible with both a strict and a sloppy reading under ellipsis.

In contrast, the pronominal subject of Control-type irrealis clauses patterns precisely with PRO in English. This is shown in (23).

(23) Nàntǒso rà Juân [nakatsya rà míí tsyàà], [sǎ =ti ñá Maria
forget.PAST he J. wash.IRR he the clothes like.this =also she M.
ba].
EMPH
'Juan forgot to wash the clothes, and Maria did too.' (**x**, ✓SLOPPY)

(23), unlike (22), is compatible with only a sloppy interpretation. In this way, we see another parallel between the subjects of Control-type irrealis clauses and PRO, and between the subjects of *pro*-type irrealis clauses and regular pronouns.

3.1.3 Necessity of local binding

The third characteristic of PRO in Obligatory Control constructions that we will examine here is local binding. In English, PRO requires a local binder, where an overt pronoun does not. This is demonstrated in (24).

(24) a. John$_i$ claimed [PRO$_{i,\ *j}$ to be a great writer].

b. John$_i$ said [that he$_{i,\ j}$ is a great writer].

If the parallel established so far between the pronominal subject of a Control-type irrealis clause and PRO is on the right track, then we predict that Control-type pronominal subject must be locally bound as well. In contrast, if the pronominal subject of a *pro*-type irrealis clause is interpreted as a regular pronoun, then we predict that it need not be locally bound.

This prediction largely pans out, although the actual empirical landscape is more complex and interesting. Let us first consider the behavior of regular embedded pronominal subjects. Just as in the English in (24b), regular pronouns across clause boundaries may, or may not, be coreferent.

(25) Ká'an rà Julio$_i$ [lísto ba'ă rà$_{i,\ j}$].
think.PRES he J. clever quite he
'Julio$_i$ thinks that he$_{i,\ j}$ is quite clever.'

In (25), we see that regular subjects need not be locally bound. Therefore, if the subjects of *pro*-type irrealis clauses are identical to regular pronouns, we expect the same range of interpretation. This turns out to be the case, but not in a simple way. First, consider (26).

(26) Kôni ñá Juana$_i$ [keba'a ñá$_{i,\ *j}$].
want.PRES she J. win.IRR she
'Juana wants to win.'

In (26), we see that the pronominal subject of a *pro*-type irrealis clause must be bound by a local binder, in this case the matrix subject *ñá Juana*. In this way, these cases seem to behave more like the Obligatory Control configuration in (24a) than (24b).

There is a complication, though. Pronominal subjects of *pro*-type irrealis clauses may have a non-local binder, like a pronoun and unlike PRO. In this case, however, the irrealis clause must be overtly marked with the complementizer *ná*.

(27) Kôni ñá Juana$_i$ [ná keba'a ñá$_{*i,\ j}$].
want.PRES she J. NÁ win.IRR she
'Juana wants her to win.'

In (27), we see that the pronominal subject of a *pro*-type irrealis clause introduced by the complementizer *ná* must be interpreted as non-locally bound.[7] This leads to the following characterization in (28).

(28) Overt complementizers block Obligatory Control configurations.

The characterization in (28) is strikingly similar to common instances of Chomsky's (1981) 'Avoid Pronoun Principle.' These are shown in (29).

(29) Spanish

a. Juan$_i$ quiere [PRO$_{i,\ *j}$ irse].
J. wants PRO to.leave
'Juan wants to leave.'

b. Juan$_i$ quiere [que *pro*$_{*i,\ j}$ se vaya].
J. wants C *pro* leaves.SUBJ
'Juan wants him to leave.'

[7]The connect to switch reference systems is tempting. I hesitate to make this connection strongly here for typological reasons. All Mesoamerican languages are traditionally described as lacking switch reference systems. In fact, this property is considered a core aspect of the Mesoamerican Sprachbund, of which all Mixtec languages are a part (Campbell et al., 1986). More work would be necessary to strengthen the connection between the SMP Mixtec pattern in (26-27) and switch reference systems, or demonstrate further differences.

In (29), we see that in Spanish, there is nothing inherent about the complement of *querer* 'want' that requires an Obligatory Control interpretation. Rather, the Obligatory Control interpretation emerges when no overt CP layer is present as in (29a). Likewise, the Obligatory Control interpretation is necessarily blocked when an overt CP layer occurs in the lower clause. I take this as an instance of fundamentally the same phenomenon in (28).

But since we have seen so far that Obligatory Control interpretations are required with Control-type irrealis clauses, we now expect that such clauses will be ungrammatical with an overt complementizer, and therefore a free interpretation of the embedded pronominal subject. This turns out to be the case. Consider (30).

(30) a. Kìxǎ rà Macario$_i$ [taxa'a rà$_{i,\ *j}$].
start.PAST he M. dance.IRR he
'Macario started to dance.'

b. * Kìxǎ rà Macario$_i$ [ná taxa'a rà$_j$].
start.PAST he M. NÁ dance.IRR he
Intended: 'Macario$_i$ started him$_j$ to dance.'

c. Nàntǒso ñá Maria$_i$ [nakatsya ñá$_{i,\ *j}$ míí tsyàà].
forget.PAST she M. wash.IRR she the clothes
'Maria forgot to wash the clothes.'

d. * Nàntǒso ñá Maria$_i$ [ná nakatsya ñá$_j$ míí tsyàà].
forget.PAST she M. NÁ wash.IRR she the clothes
Intended: 'Maria$_i$ forgot that she$_j$ washed the clothes.'

In (30), we see that Control-type irrealis clauses, which display all of the other core interpretative properties of Obligatory Control, are incompatible with an overt CP layer.

To summarize, complement irrealis clauses come in two types. The first are clauses whose subjects exhibit the interpretative properties of Obligatory Control, such as being necessarily *de se* and supporting only "sloppy" readings under ellipsis. These

"Control-type" irrealis clauses also disallow overt complementizers and the unbound subject interpretation linked with the appears of the complementizer.

The second type are irrealis clauses whose subjects do not exhibit the interpretative properties of Obligatory Control in that they allow *de re* interpretations and "strict" readings under ellipsis. These "*pro*-type" irrealis clauses also allow overt complementizers and the unbound subject interpretation which the presence of the CP-layer seems to make possible.

3.1.4 Wrapping up embedded clauses

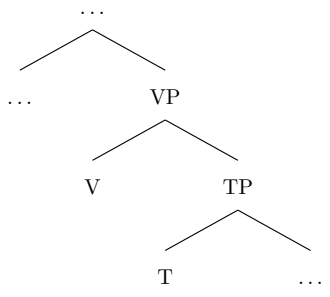
Expanding our view a little, we now see that embedded irrealis clauses can be divided into two kinds depending on their interpretive effects. These are summarized in (31).

(31)

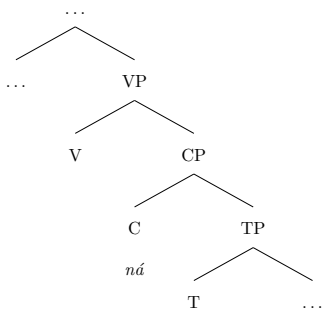
		de re available?	strict reading under VPE available?	Non-OC Available?
a.	Regular pronouns	✓	✓	✓
b.	PRO	**x**	**x**	**x**
c.	Pronoun in PRO-type irr.	✓	✓	**x**/✓
d.	Pronoun in Control-type irr.	**x**	**x**	**x**

Based on the discussion above, I propose the structures for embedded irrealis clauses in (32-33).

(32) With Obligatory Control



(33) Without local binding



(32-33) captures the core empirical claim of this investigation into embedded irrealis clauses: where local binding occurs, resulting in an Obligatory Control configuration, there is no CP layer in the embedded clause. Likewise, where Obligatory Control does not occur, there is a CP layer.

Now, this does not derive the difference between *pro*- and Control-type irrealis clauses, nor the distinction between (32) and the proposed structure for true finite embedded clauses in this language, presented above in (7). Rather, this distinction does not seem to be at C, but rather at T. As Krapova (1998) and Landau (2000, 2004) observe, the core difference between these two varieties of embedded irrealis, or more broadly embedded subjunctive, clauses is the interpretation of tense within the embedded clause. Consider the *pro*-type irrealis in (34a) in contrast with the Control-type in (34b).

(34) a. Kòni rà doktór$_i$ [tsii rà$_i$ rà Macario **itsyààn**].
want.PAST doctor catch.IRREAL he he M. tomorrow
'The doctor wanted to catch Macario tomorrow.'

b. * Kìxă míí leso$_i$ [taxa'a rí$_i$ **itsyààn**].
start.PAST the rabbit dance.IRREAL it.AML tomorrow
Intended: 'The rabbit started to dance tomorrow.'

In both of the examples in (34), the matrix clause is past tense, while both embedded clauses have the adjunct *itsyààn* 'tomorrow.' Clearly, *itsyààn* 'tomorrow' is not compatible with the past tense referent introduced by the matrix clause. Why, then, is (34a) grammatical? Following the works cited above, I take this to indicate that the tense of *pro*-type irrealis clauses is not isomorphic to the tense of the matrix clause. Interestingly, (34b) shows the opposite for Control-type irrealis clauses. Here, no independent tense referent is available, failing to license *itsyààn* 'tomorrow.' In other words, the tense of the Control-type irrealis clause is anaphorically dependent on the tense of the matrix clause.

These results are interesting in their own right for their implications for the theory of Obligatory Control. For our purposes here, though, they will principally be important when we come to consider the locality restrictions on A-movement in SMP Mixtec in §3.2.1 below. We turn to the A-movement in question in the section which follows.

3.2 The A-properties of quantifier fronting

Recall from (4) above that SMP Mixtec has a process which raises quantificational expressions to the left of the verb. (4) is reproduced as (35).

(35) a. Íntoso mií sâ nda'ă mií ítŭn kân.
be.on.PRES.PL the bird in the tree that
'The birds are in that tree.'

b. [Nda'ă iin ítŭn]$^{\exists}$ íntoso [ntskû sâ]$^{\forall}$ __.
in one tree be.on.PRES.PL all bird
'All the birds are in a tree.' ($\exists > \forall$)

c. [Ntskû sâ]$^{\forall}$ íntoso __ [nda'ă iin ítŭn]$^{\exists}$
all bird be.on.PRES.PL in one tree
'All the birds are in a tree.' ($\forall > \exists$)

(35b-c) demonstrate quantifier fronting. Here, a quantified nominal is displaced relative to its base position in (35a) to the left of the verb. As can also be seen in (35b-c), the quantificational expression which undergoes fronting takes high scope relative to any other quantifiers in the sentence.

A variety of quantification elements may be fronted, regardless of whether other quantifiers are present in the rest of the clause. Some further examples of quantifier fronting are shown in (36).

(36) a. Ntsi'bĭ$_i$ kí'i ñá __$_i$.
egg pick.up.PRES she

'She's gathering (some indefinite quantity of) eggs.'

b. Ní iin nà$_i$ ǐ sókŏn __$_i$.
even one them NEG tall
'No one is tall.' (It is not the case that even one of them is tall.)

c. iin tsǐna$_i$ ntá'yi __$_i$.
one dog cry.PRES
'A dog is crying.'

d. Kwa'ă líbro$_i$ kúú __$_i$ ñà ntsí'ì.
many book COP.PRES it purple
'Many of the books are purple.'

e. Kòhmǐ líbro$_i$ nàkààba __$_i$.
four book fall.PAST
'Four books fell.'

In (36), we see a variety of quantification elements undergoing fronting. These range from bare indefinites in (36a) to negative indefinites in (36b), as well as nominal phrases with numerals, as in (35e).

Failing to front a quantifier nominal expression is usually judged as grammatical, but it seems to be obligatory for others. In all cases, quantifier fronting is strongly preferred. But consider (37).

(37) a. # Kí'i ñá ntsi'bǐ.
egg pick.up.PRES she
Intended: 'She's gathering (some indefinite quantity of) eggs.'
Good as 'She's picking up an (known quantity, most naturally one) egg.'

b. ?? Ǐ sókŏn ní iin nà .
NEG tall even one them
'No one is tall.'

c. Ntá'yi iin tsǐna.
cry.PRES a dog
'A (non-specific) dog is crying.'

d. ?? Ñà ntsí'ì kúú kwa'ă míí líbro.
it purple COP.PRES many the book
'Many of the books are purple.'

e. Ñà$_i$ nàkààba kòhmĭ líbro$_i$.
it fall.PAST four book
'Four (specific) books fell.'

In (37b, d) we see that failure to front negative indefinites or nominals quantified by *kwa'ă* 'many' is rather degraded, but not fronting the indefinites in (37a, c), or the nominal phrase with a number in (37e), is judged as alright, although the interpretation may change.

In addition to a movement analysis of quantifier fronting, another possibility presents itself: base-generation followed by binding of a null pronoun. These two analyses are sketched in (38).

(38) a. Quantifier fronting as movement:
QP$_i$ ... $\_\__i$

b. Quantifier fronting as base generation:
QP$_i$ ... *pro*$_i$

(38a) presents the movement analysis that I adopt here, while (38b) shows an alternative analysis that does not involve movement. Let us see several arguments for why the movement analysis in (38a) is the best fit for this language.

First, if base generation were involved, it is unclear why this process would be restricted to quantified expressions. Consider the contrast between (39a), which is repeated from (36a), and (39b).

(39) a. Nsti'bĭ$_i$ kí'i ñá $\_\__i$.
egg pick.up.PRES she
'She's gathering (some indefinite quantity of) eggs.'

b. *? Míí ntsi'bǐ kí'i ñá $\_\__i$.
the egg pick.up.PRES she
Intended: 'She's picking up the egg.'

In (39a), we see that only the bare indefinite can undergo quantifier fronting; the definite description in (39b) cannot. If base generation were involved, it is unclear why this restriction would hold.

Second, SMP Mixtec actually has a process[8] that seems to employ the syntax in (38b) but shows a different syntactic profile. I call this process 'topic left dislocation.' This is shown in (40).

(40) Topic left dislocation

a. Ntskû rà tsyahá$_i$, kíchûn nà$_i$.
all he man work.PRES they
'All the men, they are working.'

b. Kwa'ă líbro$_i$ kíí yá$_i$ ñà kwă'à.
many book COP.PRES it it red
'Many books, they are red.'

c. Míí ntsi'bǐ$_i$, kí'i ñá yá$_i$.
the egg pick.up.PRES she it
'The egg, she is picking it up.'

In (40), we see that topic left dislocation involves a nominal expression to the left of the clause, with an overt pronoun in the base position. This hanging left topic is also optionally separated from the remainder of the clause by an intonational break, indicated with a comma throughout.

Importantly, topic left dislocation does not seem to involve movement. This can be shown in several ways. First, we can see in comparing the grammatical (40c) to the ungrammatical (39b) that topic left dislocation does not impose the same restrictions

[8]Two, in fact. See §4.2.2.

that quantifier fronting does on the kinds of elements that can be involved. This is consistent with base generation, but not, as we discussed just before, with movement.

Second, topic left dislocation does not display the same boundedness restrictions that quantifier fronting does (see §3.2.1 below). Where quantifier fronting is clause bounded, (41) is entirely grammatical.

(41) a. **Ntsi'i míí rà tsyahá**$_i$ xa nì ka'an rà Juan xí'in ná sí'i
all the he man PERF PAST promise he J. with they.FEM woman
[kachûn **rà**$_i$ itsyààn].
work.IRR he tomorrow
'All of the men, Juan promised the women that they would work tomorrow.'

b. **Ntsi'i míí** rà tsyahá, kúú sihĭ ini ñá Maria [chi
all the he man COP.PRES happy inside she M. because
kachûn **rà** itsyààn].
work.IRR he tomorrow
'All of the men, Maria is happy because they will work tomorrow.'

(41a) shows that topic left dislocation may form a dependency across a finite clause boundary. This contrasts, as we will see shortly in §3.2.1, with quantifier fronting. (41b) shows such a dependency may also reach into an island, specifically a because-clause island. See §4 for more discussion of islands in this language. I conclude that topic left dislocation does not involve movement, but rather base generation and binding of a pronoun in argument position.

Compare the properties of quantifier fronting and topic left dislocation, summarized in (42).

(42) Quantifier fronting versus topic left dislocation

		Cross clause boundaries/islands?	Overt pronoun?	Restriction to quantifiers?
a.	Quantifier fronting	**x**	**x**	✓
b.	Topic left dislocation	✓	✓	**x**

If we wanted to analyze quantifier fronting as involving base generation followed by binding of a null pronoun, we would need to explain why quantifier fronting and topic

left dislocation do not demonstrate the same characteristics. This contrast provides strong evidence that quantifier fronting truly involves movement.

Having concluded that quantifier fronting involves movement, I show in the rest of this section that it has the properties of A-movement.

3.2.1 Quantifier fronting is clause bounded

As we discussed in §3.1, A-movement processes are frequently[9] bounded by CP boundaries. This contrasts with A'-movement, which often appears not to be clause bounded. (43) presents these data for wh-movement and relativization briefly, and see §4 for an extensive discussion of A'-movement processes in this language.

(43) a. Yó$_i$ ká'an míí doktór [kúu __$_i$]?
who think.PRES the doctor be.sick.PRES
'Who does the doctor think is sick?'

b. Ná$_i$ ntá'yi ná bálí$_i$ [ká'an rà mástro [nàkààba __$_i$]].
they.FEM cry.PRES they.FEM little.PL think.PRES he teacher fall.PAST

'The girls they the teacher thinks fell are crying.'

In (43), we see that A'-movement processes seem to span cross clause boundaries. Importantly, quantifier fronting may not. Recall from §3.1 that embedded finite clauses and embedded irrealis clauses where Obligatory Control does not occur have a CP boundary. With this in mind, consider (44).

(44) a. Ndàtŭ'un ñá Maria xí' ìn [**kwa'ă kwê'e kárro**$_i$ nàkatsya rà Pedro
chat.PAST she M. with me many very car wash.PAST he P.
__$_i$ koni].
yesterday
'Maria told me that Pedro washed many cars yesterday.'

[9]But not universally. See §2.2.3.1 of chapter 2

b. * **Kwa'ă kwê'e kárro**$_i$ ndàtŭ'un ñá Maria xí' ìn [nàkatsya rà
many very car chat.PAST she M. with me wash.PAST he
Pedro __$_i$ koni].
P. yesterday
Intended: 'Maria told me that Pedro washed many cars yesterday.'

c. Ká'an rà doktór [**ní iin nà**$_i$ kò kú'u __$_i$].
think.PRES he doctor not.even one them NEG be.sick.PRES
'The doctor thinks that no one is sick.'

d. * **Ní iin nà**$_i$ ká'an rà doktór [kò kú'u __$_i$].
not.even one them think.PRES he doctor NEG be.sick.PRES
Intended: 'The doctor thinks that no one is sick.'

e. Kôni ñá Maria [**kwa'ă kwê'e kárta**$_i$ ná tsyaa kò'òba ñá __$_i$
want.PRES she M. many very letter NÁ write.IRR brother her
].

'Maria wants her brother to write many letters.'

f. * **Kwa'ă kwê'e kárta**$_i$ kôni ñá Maria [ná tsyaa kò'òba ñá
many very letter want.PRES she M. NÁ write.IRR brother her
__$_i$].

Intended: 'Maria wants her brother to write many letters.'

In (44), we see that quantifier movement is restricted to the clause in which the quantified nominal originates, and may not cross CP boundaries.

Quantifier movement can, however, apply out of embedded TPs when no CP layer is present. As we saw in §3.1, embedded clauses lack a CP layer when Obligatory Control occurs in either *pro*- or Control-type irrealis clauses. This is shown in (45).

(45) a. **U'un tsyaká**$_i$ kôni nà tsyahá$_j$ [tsiin nà$_j$ __$_i$].
five fish want.PRES they man catch.IRR they
'The men want to catch five fish.'

b. **Kwa'ă kárro**$_i$ kìxă rà Pedro$_j$ [nakatsya rà$_j$ __$_i$].
many car start.PAST he P. wash.IRR he
'Pedro started to wash many cars.'

In (45a), we see that quantifier fronting may occur out of a *pro*-type irrealis clause involving Obligatory Control, and (45b) shows that quantifier fronting may also occur out of Control-type irrealis clauses, where we saw in §3.1 Obligatory Control must occur.

With this, we see that quantifier fronting is only permitted out of an embedded clause when that clause is a bare TP. This is consistent with the hypothesis that quantifier fronting is A-movement, but would be inconsistent with attested A'-movement processes in this language.

3.2.2 Quantifier fronting expands variable-binding possibilities

Importantly for our purposes, quantifier fronting does not reconstruct for binding, but does allow for variable-binding configurations that we impossible in its absence. To see this, first consider the examples in (46), which provide a baseline to show that Condition C of the binding theory works similarly in SMP Mixtec works as it does in

English.[10]

[10]I do not test Condition A here for two reasons. First, reflexives are constructed on the basis of *míí* 'the' followed by a pronoun. This is shown in (i).

(1) a. Ini yùtátá xìnì ñá Maria$_i$ míí ñá$_i$.
in mirror see.PAST she M. the she
'Maria saw herself in the mirror.'

b. Ini yùtátá xìnì míí ñá$_{*i,\ j}$ ñá Maria$_i$.
inside mirror see.PAST the she she M.
Bad as: 'Herself$_i$ saw Maria$_i$ in the mirror,' good as: '*She*$_j$ saw Maria$_i$ in the mirror.'

As can be seen in (ib), reflexives in SMP Mixtec also support logophoric interpretations. This logophoricity is further demonstrated by the inability of inanimate *míí*-marked phrases to serve as transitive subjects (Charnavel and Sportiche 2016). This is shown in (ii).

(2) a. Káchi tútu [ba'á ká'abi (*míí) á].
say.PRES paper good read.PRES the it
'The book says that it (*itself) reads well.'

b. Nù yá'a míí kárro ínchichi iin *letrero* ra kách á [ñà$_i$ bă'a (*míí)
where pass.PRES the car stand.PRES.SG a sign and say.PRES it it good.NEG the
yá$_i$].
it
'Where the traffic (the cars) pass, there stands a sign and it says that it (*itself) is broken.'

This confounding logophoricity means that it would be difficult to interpret any results of anaphor binding in these contexts.

Second, SMP Mixtec has a reciprocal anaphor, *tá'àn*, but its syntax makes it difficult to evaluate binding relations in the relevant contexts. This is because *tá'àn*, as well as its containing constituent, obligatorily incorporates into the verb and reduces its transitivity by one.

(3) a. Kàñ é míí kárro.
hit.PAST we.INCL the car
'We hit the car.'

b. Yé$_i$ kàni tá'àn$_i$ -k yé$_i$ __$_i$.
we.INCL hit.PAST each.other -INTR we.INCL
'We hit each other.'

c. Tàx é ita nda'ă míí ná sí'i.
give.PAST we.INCL flower to the they.FEM woman
'We gave flowers to the women.'

d. Tàxi nda'ă tá'$_i$ én ita __$_i$.
give.PAST to each.other we.INCL flower
'We gave each other flowers.'

e. Kôñ é se'e ñá Maria.
like.PRES we.INCL child she M.
'We like Maria's child.'

f. Yé$_i$ kôni se'e tá'àn$_j$ -k yé$_i$ __$_j$.
we.INCL like.PRES child each.other -INTR we.INCL
'We like each others children.'

(46) a. Kôni ñá$_{i,\ *j}$ se'e ñá Maria$_j$.
love.PRES she child she M.
'She$_i$ loves Maria$_{*i,\ j}$'s children.'

b. Kôni ñá Maria$_i$ se'e ñá$_{i,\ j}$.
love.PRES she M. child her
'Maria$_i$ loves her$_{i,\ j}$ child.'

In (46a), we see that Condition C in SMP Mixtec behaves identically to English. Here, an R-expression in direct object position may not be bound by a pronoun in subject position. Reversing the command relations, as in (46b), restores well-formedness.

Importantly, we see that quantifier fronting expands variable-binding possibilities and fails to induce Weak Crossover. Consider (47).

(47) a. [Ntskû se'e ñá Maria$_i$]$_j$ kôni ñá$_i$ __$_j$.
all child she M. love.PRES she
'She$_i$ loves all of Maria$_i$'s children.'

b. [Ntskû míí ñá sí'i$_i$]$_k$ kôni se'e ná$_i$ __$_k$.
all the they.FEM woman love.PRES child their.FEM
'Their$_i$ children love all the women$_i$.'

In (47a), we first see that quantifier fronting calls off the Condition C violation that would be assessed if the quantifier were interpreted in its original position (that is, there is no reconstruction for Condition C). From this, it seems clear that quantifier fronting does not obligatorily reconstruct for binding.

In this language, only direct objects may incorporate. Therefore, we can conclude little from the ungrammaticality of (iv,a), as it has the further confound that subjects generally may not incorporate. Regardless, (iv,b) is grammatical, although it us unclear what this means for our theory.

(4) a. * Ntskû míí ná sí'i$_i$ kôni se'e tá'àn$_j$ __$_j$ __$_i$.
all the they.FEM women love.PRES child each.other
Intended: 'Each other's children love all the women.'

b. * Ntskû míí ná sí'isub$j$ kôni se'e tá'àn$_i$ __$_j$ __$_i$.
all the they.FEM woman their love.PRES child each.other
'All of the women love each other's children.'

Now consider (47b). Here, we see that fronting a quantified object R-expression over a coreferent pronominal possessor in the subject does not induce Weak Crossover. As we will see in §4, A'-movement consistently induces Weak Crossover, much as in English. Some of these data are presented in (48) for comparison, though see §4 for a much more extensive discussion.

(48) a. *? Yó$_i$ kôni se'e nà$_i$ __$_i$?
who love.PRES child their
Intended: 'Who$_i$ does their$_i$ child love?'

b. *? Ñá$_i$ ntá'yi míí ñá lo'o$_i$ [ká'a nána ñá$_i$ [kú'u __$_i$]
she cry.PRES the she little think.PRES mother her be.sick.PRES
].

Intended: 'The little girl who her mother thinks is sick is crying.'

In (48), we see that A'-movement consistently induces Weak Crossover. This is demonstrated for wh-movement and relativization in (48a) and (48b) respectively. Importantly for our purposes, in (47b) we see that quantifier fronting does not induce Weak Crossover. Rather, in (47b) we see that quantifier fronting feeds variable binding in a way that is not available in English. This failure to induce Weak Crossover, coupled with the ability to affect variable binding, provide a strong argument that quantifier fronting is A-movement, not A'-movement.

Finally, quantifier fronting allows for variable-binding possibilities that are impossible in its absence. Consider the contrast between (49a) and (49b).

(49) a. * Ntàsaba'ă xito'o tún$_i$ ntskû kárro$_i$.
repair.PAST owner it.WOOD every car
Intended: 'Its$_i$ owner repaired every$_i$ car.'

b. Ntskû kárro$_i$ xa ntàsaba'ă xito'o tún$_i$ __$_i$.
every car PERF repair.PAST owner it.WOOD
'Its$_i$ owner repaired every car$_i$.'

In (49a), we see an unsurprising pattern in which variable-binding without c-command is severely degraded in SMP Mixtec. See Wasow (1972), Lasnik (1976), and Reinhart (1983), among others, though see Barker (2012). Importantly, in (49b) we see that when overt quantifier fronting of the object over the subject occurs, variable-binding becomes possible. In this way, we see that quantifier fronting allows for variable-binding configurations that would be impossible in its absence.

With this, we see that quantifier fronting demonstrates many unique hallmarks of A-movement, such as the ability to alter variable binding relations, affect Condition C of the binding theory, and the failure to induce Weak Crossover. Therefore, from the perspective of variable binding, quantifier movement patterns strongly with A-movement.

3.2.3 Quantifier fronting can extract objects of prepositons and possessors

Another distinguishing feature that distinguishes quantifier fronting from A'-movement processes in SMP Mixtec is the ability to extract objects of prepositions and possessors.[11] First, consider the behavior of A'-movement. A'-movement processes systematically prohibit extraction of possessors or objects of prepositions, shown in (50).[12]

[11] See chapter 4, particularly §4.3.3.4 for more details. The discussion here suffices for comparative purposes, particularly when we consider the behavior of pronoun doubling in §3.3.4.

[12] I only show unaccusative subjects and transitive direct objects because unergative and transitive subjects are islands. See §4.1.3. Quantifier fronting, which as we will see shortly can extract possessors and objects of prepositions from internal argument position, cannot do so from external argument position. This is shown in (i) and (ii), for possessor extraction from external argument DPs and subject PPs respectively.

(1) a. * Ntskû nà$_i$ kárakono se'e __$_i$.
all them run.PRES child
Intended: 'Everyone's children are running.'

b. * Ntskû míí ná sí'i nákatsya ìhì __$_i$ tsyàà.
all the they woman wash.PRES husband clothes
Intended: 'The husbands of all the women are washing clothes.'

(50) No wh-extraction of possessors

a. Rí$_i$ nì xi'i tsǐna sâna rà Juan.
it.AML PAST die dog POSS.AML he J.
'Juan's dog died.'

b. * Yó$_i$ nì xi'i tsǐna sâna __$_i$?
who PAST die dog POSS.AML
Intended: 'Whose dog died?'

c. [Yó$_i$ tsǐna sâna __$_i$]$_k$ nì xi'i __$_k$?
who dog POSS.AML PAST die
'Whose dog died?'

d. Ntasaba'a rà Macario se'e ñá Maria.
heal.IRR he M. child she M.
'Macario will heal Maria's child.'

e. * Yó$_i$ ntasaba'a rà Macario se'e __$_i$?
who heal.IRR he M. child
Intended: 'Whose child will Macario heal?'

f. [Yó$_i$ se'e __$_i$]$_k$ ntasaba'a rà Macario __$_k$?
who child heal.IRR he M.
'Whose child will Macario heal?'

In (50b, e), we see that simple wh-movement may not extract possessors while stranding the possessum. Rather, SMP Mixtec requires a construction referred to as pied-piping with inversion, a common phenomenon in Mesoamerican languages (Smith-Stark 1988, Black 1994, Aissen 1996, Eberhardt 1999, Broadwell et al. 2006, Coon 2009, and Cable 2012, 2013, inter alia). The inversion can be seen in comparing the base examples without wh-expressions, like (50a, d) to the pied-piping with inversion cases in (50c, f). In the base examples, we see that possessors obligatory follow possessa. In contrast, when pied-piping with inversion takes place, the wh-possessor must proceed the possessum.

c. * Ntskû míí rà tsyahá$_i$ kúú sihǐ ini ísi'i __$_i$.
all the he man COP.PRES happy inside wife
Intended: 'All of them men's wives are happy.'

In this sense, we see an inversion from the regular order of possessor and possessum. See chapter 5 for much more detail on pied-piping with inversion.

In addition to possessors, objects of prepositions likewise cannot be extracted by wh-movement.[13] Rather, pied-piping with inversion must be used.[14]

(51) No wh-extraction of objects of prepositions

a. Nì ka'an rà Juan xí'in ñá Maria.
PAST talk he J. with she M.
'Juan talked with Maria.'

b. * Yó$_i$ nì ka'an rà Juan xí'in __$_i$?
who PAST talk he J. with
Intended: 'Who did Juan talk with.'

c. [Yó$_i$ xí'in __$_i$]$_k$ nì ka'an rà Juan __$_k$?
who with PAST speak he J.
'Who with did Juan speak?'

d. Rà$_i$ kíxì rà Macario$_i$ xihĭn ísi'i rà.
he sleep.PRES he M. beside wife his
'Macario is sleeping beside his wife.'

e. * Yó$_i$ kíxì rà Macario xihĭn __$_i$?
who sleep.PRES he M. beside
Intended: 'Who is Macario sleeping beside?'

f. [Yó$_i$ xihĭn __$_i$]$_k$ kíxì rà Macario __$_k$?
who beside sleep.PRES he M.
'Who is Macario sleeping beside?'

In (51), we see a pattern identical to that in (50) in that a wh-object of a preposition cannot undergo independent A'-movement and strand the preposition. Rather, pied-piping of the whole prepositional phrase must occur, along with inversion.

[13] As in many Oto-Manguean languages, lexical items which correspond to prepositions in other languages are in fact derived from body parts. The category of these items is an open mystery that I will not address here. Instead, I take the maximally conservative view by treating true body parts and their spatial uses separately.

[14] An alternative construction, which likewise does not involve A'-extraction of the complement of the preposition, may also be used. See the appendix.

The inability to A'-extract a possessor or the object of a preposition, along with the presence of pied-piping with inversion, seems to be one of the signatures of A'-movement in this language, as it is in many other Mesoamerican languages. With this in mind, consider the behavior of quantifier fronting in (52) and (53).

(52) Quantifier fronting may extract possessors

a. Ndàkinúû láncha ña'a míí rà tsyahá.
sink.PAST boat POSS the he man
'The man's boat sank.'

b. U'un rà tsyahá$_i$ ndàkinúû láncha ña'a __$_i$
five he man sink.PRES boat POSS
'The five men's boat sank.'

c. * [U'un rà tsyahá$_i$ lácha ña'a __$_i$]$_k$ ndàkinúû __$_k$.
five he man boat POSS sink.PAST
Intended: 'The five men's boat sank.'

d. Nàkatsya ñá Maria tsyàà míí nà bálí.
wash.PAST she M. clothes the they little.PL
'Maria washed the children's clothes.'

e. Ntsi'i nà bálí$_i$ nàkatsya ñá Maria tsyàà __$_i$.
all they little.PL wash.PAST she M. clothes
'Maria washed all the children's clothes.'

f. * [Ntsi'i nà bálí$_i$ tsyàà __$_i$]$_k$ nàkatsya ñá Maria __$_k$.
all they little.PL clothes wash.PAST she M.
Intended: 'Maria washed all the children's clothes.'

(53) Quantifier fronting may extract objects of prepositions

a. Nì ka'an ñá Maria xí'in rà Juan.
PAST speak she M. with he J.
'Maria spoke with Juan.'

b. Ní iin nà$_i$ kò ní ka'an ñá Maria xí'in __$_i$.
not.even one them NEG PAST speak she M. with
'Maria spoke with no one.'

c. * [Ní iin nà$_i$ xí'in __$_i$]$_k$ kò ní ka'an ñá Maria __$_k$.
not.even one them with NEG PAST speak she M.
Intended: 'Maria spoke with no one.'

d. Tàxi rà Sergio xu'ú nda'ă ñá Juana.
give.PAST he S. money to she J.
'Sergio gave money to Juana.'

e. Kwa'ă ñà yibì$_i$ tàxi rà Sergio xu'ú nda'ă __$_i$.
many it people give.PAST he S. money to
'Sergio gave money to many people.'

f. * [Kwa'ă ñà yibì$_i$ nda'ă __$_i$]$_k$ tàxi rà Sergio xu'ú __$_k$.
many it people to give.PAST he S. money
Intended: 'Sergio gave money to many people.'

In (52) and (53), we see that quantifier fronting shows rather different behavior from wh-movement. Unlike wh-movement, quantifier fronting is totally incompatible with pied-piping with inversion, as seen in (52c, f) and (53c, f). Instead, quantifier fronting obligatorily[15] strands either the preposition or the possessor.

Now, if we continue from the premise that wh-movement is A'-movement while quantifier fronting is A-movement, we are lead to an interesting conclusion about the availability of possessum and preposition stranding, as well as pied-piping with inversion. This is stated in (54).

(54) A'-movement prohibits, but A-movement requires, possessum/preposition stranding.

[15]Quantifier fronting is not mandatory in some of these examples. This is dependent on the target scope. For instance, in (52b), the quantified phrase *u'un rà tsyahá* 'five men' may remain in situ, but it must take low scope. Specifically, (52b) must have the interpretation in which *u'un* 'five' scopes over the existential inherent to *láncha* 'boat.' In other words, (52b) means that there are five boats, one for each man. In contrast, in (i) the existential must scope over *u'un* 'five,' yielding an interpretation in which there is one boat which the five men collectively share.

(1) Ndàkinúû láncha ña'a u'un rà tsyahá.
sink.PAST boat POSS five he man
'The boat of the five men sank.'

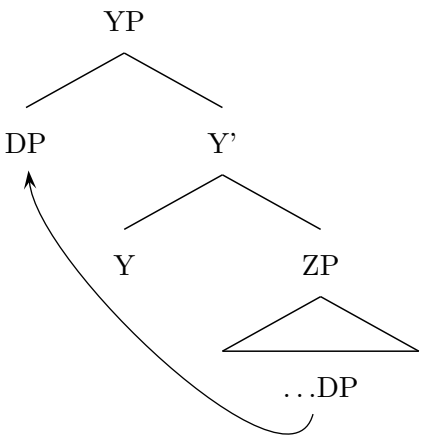
(54) provides a strong and useful diagnostic. Specifically, if A-movement is involved, then we expect to see extractability of possessors and objects of prepositions. Unfortunately, I am unaware of any other potential A-movement processes in this language against which we could test (54), but with this much in place, let us return to pronoun doubling to see if we can find any evidence of an A-chain.

3.3 Pronoun doubling does not form an A-chain

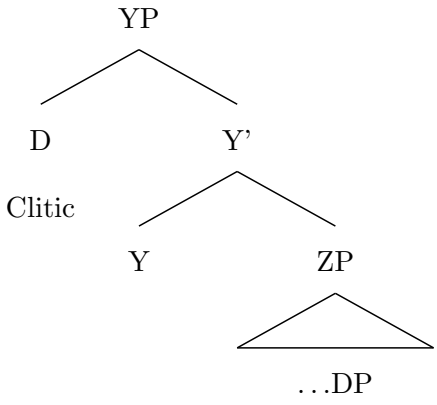
Recall why we might expect to see the properties of an A-chain in pronoun doubling constructions. As established in §2 of chapter 2, I analyze clitic doubling here as involving phrasal displacement of the doubled constituent, followed by reduction of the higher copy to a clitic. This Move-and-Reduce account of clitic doubling is presented in (55).

(55) Where Spec,YP is an A-position

a.



b.



In the system shown in (55), clitic doubling is a subvariety of multiple spell out under A-movement. The higher position is spelled out as the clitic, while the full doubled nominal phrase is exponed in the lower position.

If we want to analyze pronoun doubling as clitic doubling, then we should

be able to diagnose either A-movement or A'-movement. So far in this chapter, we have investigated the various properties of A-movement in SMP Mixtec, in particular quantifier fronting. Given this background, we will expect (56).

(56) If pronoun doubling is derived through A-movement, then:

a. Pronoun doubling should be clause bounded (c.f. §3.2.1)

b. Pronoun doubling should expand variable-binding possibilities and ameliorate Condition C violations (c.f. §3.2.2)

c. Pronoun doubling should extract objects of P and possessors (c.f. §3.2.3)

Below, I will demonstrate that only (56a) holds.

3.3.1 Pronoun doubling is clause bounded

Like quantifier fronting, pronoun doubling is clause bounded. To see this, first consider (57) and (58).

(57) a. Tát ì ndatu'un rà Juán xí' ìn áto ba'á ñá Maria xí'a nì
wait.PRES I chat.IRR he J. with me if good she M. after PAST
tahan.
quake
'I hope that Juan tells me that Maria is well after the earthquake.'

b. **Rà**$_i$ xa ndàtu'un rà Juán$_i$ xí' ìn [íyo ba'á ñá Maria], so
he PERF said he J. with me be.PRES.SG good she M but
xi̱n ì xa'á ñá Julia.
know.NEG I about she J.
'Juan already told me that Maria is fine, but I don't know about Julia.'

c. * **Ñá**$_i$ xa ndàtu'un rà Juan xí' ìn [íyo ba'á ñá Maria$_i$]...

(58) a. ☒ Xa káchi iin ñá sí'i xí' ìn [baxí rà Roberto
PERF say.PAST one she woman with me come.IRR he R.
itsyààn], so xi̱n ì yó ñá.
tomorrow but know.PRES.NEG I who she

'A woman has already told me that Roberto will come tomorrow, but I don't know who she was.'

b. * **Rà**$_i$ xa káchi iin ñá sí'i xí' ìn [baxí rà Roberto$_i$
he PERF say.PAST one she woman with me come.IRR he R.
itsyààn], so xǐn ì yó ñá.
tomorrow but know.PRES.NEG I who she

Let us first consider (57). Here, (57a) sets up a context in which the embedded subject *ñá María* is highly topical. Therefore, we might expect pronoun doubling to be able to target this argument, per the general observation that pronoun doubling tracks topics (see the discussion in chapter 1, and all of chapter 5). Interestingly, in (57c), we see that pronoun doubling may not reach into the embedded clause. Instead, only the matrix subject in (57b) may be doubled. Therefore, I conclude that pronoun doubling cannot reach into embedded CPs.

(58) suggests the same conclusion. In the matrix clause of the grammatical (58a), we have two arguments: a non-specific indefinite subject *iin ñá sí'i* 'a woman' and a prepositional phrase *xí'ìn* 'with me.' As discussed in chapter 1, and more extensively in chapter 5, non-specific indefinites may never undergo pronoun doubling, and as we will see shortly in §3.3 of this chapter, objects of prepositions may never undergo pronoun doubling either. In other words, we have no nominals in the matrix clause that could be the targets of pronoun doubling, but we do have an embedded subject *rà Roberto*, that could. In particular, the embedded subject, as a specific, referential nominal, makes an excellent possible target for pronoun doubling. This means that if pronoun doubling could reach into embedded CPs, then we expect (58b) to be grammatical. It is not. This again points to the conclusion that pronoun doubling cannot penetrate embedded CPs.

But as we established in §1 and §2.1 of this chapter, SMP Mixtec makes a systematic distinction between embedded CPs and embedded TPs in irrealis clauses. Interestingly, pronoun doubling is prohibited into embedded CP irrealis clauses, but may

freely occur into embedded bare TPs. This is shown in (59) and (60-61) for embedded irrealis clauses with and without a CP layer, respectively.

(59) No pronoun doubling out of embedded irrealis with CP layer (without Obligatory Control)

a. Ná kúndàà ini Roberto níhìn kwê ñá Maria, koni rà ná
SUBJ know.PRES inside R. skinny very she M. want.IRR he SUBJ
kaxi ñá chocolate .
eat.IRR she chocolate
'Should Roberto know how skinny Maria is, he'd want her to eat chocolate.'

b. **Rà**$_i$ xa kóni rà$_i$ [ná kaxi ñá Maria chocolate].
he PERF want.PRES he SUBJ eat.IRR she M. chocolate
'He already wants Maria to eat chocolate!'

c. * **Ñá**$_i$ xa kóni rà [ná kaxi ñá Maria$_i$ chocolate].
it PERF want.PRES he SUBJ eat.IRR she M. chocolate

d. * **Ñà**$_i$ xa kóni rà [ná kaxi ñá Maria chocolate$_i$].
it PERF want.PRES he SUBJ eat.IRR she M. chocolate

(60) Pronoun doubling out of embedded irrealis without CP layer (with Obligatory Control)

a. Tibi nákana míí Mario nù chíñú, nantŏso rà [nakatsya rà
early leave.PRES.SG the M. where work forget.IRR he wash.IRR he
tsyàà].
clothes
'If Mario leaves work early, he'll forget to wash the clothes.'

b. **Rà**$_i$ xa nàntŏso míí Mario$_i$ [nakatsya rà míí tsyàà].
he PERF forget.PAST the M. wash.IRR he the clothes
'Mario has already forgotten to wash the clothes!'

c. **Ñà**$_i$ xa nàntŏso míí Mario [nakatsya rà míí tsyàà$_i$].
he PERF forget.PAST the M. wash.IRR he the clothes
'Mario has already forgotten to wash the clothes!'

(61) Pronoun doubling out of embedded irrealis without CP layer (with Obligatory Control)

a. Ná konì míí Marco míí chocolate, koni rà kaxi rà ñà.
SUBJ see.IRR the M. the chocolate want.IRR he eat.IRR he it
'Should Marcos see the chocolate, he'll want to eat it.'

b. **Ñà**$_i$ xa kóni rà [kaxi rà míí chocolate$_i$]!
it PERF want.PRES he eat.IRR he the chocolate
'He already wants to eat the chocolate!'

In comparing (59) to (60-61), we see that the language makes a sharp distinction between the availability of pronoun doubling out of embedded CPs versus embedded bare TPs. In (59), just as we saw in (57-58), we see that pronoun doubling may not occur out of embedded CPs. But in (60-61), we see that pronoun doubling readily occurs out of embedded bare TPs. Indeed, pronoun doubling from the embedded TP is often judged as better than matrix doubling in these contexts. For instance, while both sentences in (60) are grammatical, (60c) is judged as more natural than (60b).

From this, we can straightforwardly conclude that pronoun doubling is clause bounded. Therefore, we see that pronoun doubling shares this core property with quantifier movement. At the same time, this by itself is not enough to draw any kind of meaningful parallel between pronoun doubling and A-movement processes of the sort that a Move-and-Reduce analysis of clitic doubling requires. In particular, recall from chapter 2 that both φ-agreement systems and some A-movement processes never crosses CP boundaries. Therefore, the diagnostic of clause boundedness is only usefil if a process is not clause bounded. As such, we do not yet have enough reason to establish a firm connection between clause boundedness and A-movement to the exclusion of φ-agreement.

Given that clause boundedness by itself is inconclusive, let us examine the results of other diagnostics.

3.3.2 Pronoun doubling does not affect binding

To begin, recall from §2.1 of chapter 2 that clitic doubling can have consequences for anaphoric relations of various kinds in various languages. In this discussion, I will focus on the discussion of Modern Greek from chapter 2, with all data coming from Alexiadou and Anagnostopoulou (1997).

The first way in which clitic doubling may affect anaphoric relations has to do with the Bound Anaphora Condition, namely that pronounds interpreted as bound variables must be c-commanded by the quantifiers which bind them (Reinhart, 1983). Consider the Modern Greek in (62).

(62) * O Petros epestrepse [tu idioktiti tu$_i$]$_j$ [to kathe aftokinito]$_i$
the P.NOM returned.3SG the owner.GEN his the every car.ACC
xtes to vradi.
yesterday the night
Intended: 'Petros returned its$_i$ owner every car$_i$ last night.'

This example fails because the accusative quantifier *to kathe aftokinito* 'every car' does not c-command the pronoun within the indirect object, and so it cannot bind it. Crucially, clitic doubling the accusative DP, as in (63), repairs the violation.

(63) O Petros **to**$_i$ epestrepse [tu idioktiti tu$_i$]$_j$ [to kathe aftokinito]$_i$
the P.NOM it.ACC returned.3SG the owner.GEN his the every car.ACC
xtes to vradi.
yesterday the night
'Petros returned its$_i$ owner every car$_i$ last night.'

As we saw in §2.2.2.1, the contrast between (62) and (63) is understandable in the context of our Move-and-Reduce analysis of clitic doubling. In the context of such an analysis, the accusative quantifier first raises across the indirect object, which contains the crucial pronoun. This is shown in (64).

(64) O Petros [*to kathe aftokinito*$_i$] epestrepse [to idioktiti tu$_i$] [to kathe aftokinito].
the every car the owner.GEN his

From this more prominent position, the quantifier *to kathe aftokinito* 'every car' commands, and can therefore bind, the pronoun within the indirect object. The reduction of the higher copy to a clitic and full spell out of the lower copy both take place in the post-syntactic derivation, and will be irrelevant for, and invisible to, the mechanisms of scope and binding. In the context of such an analysis, then, the well-formedness of (63) falls out as a standard instance of the expansion of binding possibilities made available by movement to A-positions.

Importantly for our purposes, pronoun doubling in SMP Mixtec behaves differently from clitic doubling in Greek. First, recall that variable-binding in SMP Mixtec is subject to the same c-command requirement as Modern Greek. This is shown in (65), repeated from (49).

(65) a. Ntàsaba'ă ntskû rà tsyahá$_i$ kárro rà$_i$.
repair.PAST every he man car his
'Every man$_i$ repaired his$_i$ car.'

b. * Ntàsaba'ă xito'o tún$_i$ ntskû kárro$_i$.
repair.PAST owner it.WOOD every car
Intended: 'Its$_i$ owner repaired every$_i$ car.'

In (65a), we see that a quantified subject may bind a variable in the object, but (65b) demonstrates that the inverse relation in which the object binds a variable in the subject is not possible.

Now, if pronoun doubling were derived by A-movement, as proposed for the Modern Greek in (63-64), then we would predict that pronoun doubling the object in (65b) should make variable-binding possible. This is because the pre-Reduce structure of (65b) would be as in (66), comparable to (64).

(66) *Ntskû kárro*$_i$ xa ntàsaba'ă xito'o tún$_i$ ntskû kárro.
every car PERF repair.PAST owner it.WOOD every car

Crucially, this prediction does not pan out: pronoun doubling does not allow for novel variable-binding possibilities. This is shown in (67).

(67) * Tún$_i$ xa ntàsaba'ă xito'o tún$_i$ ntskû kárro$_i$.
it.WOOD PERF repair.PAST owner it.WOOD every car
Intended: 'Its$_i$ owner repaired every car$_i$.'

In (67), we see that pronoun doubling the quantified object over the variable in the subject does not license variable-binding. This would be surprising if pronoun doubling were derived by A-movement, especially when one recalls that overt quantifier movement of the object over the subject, as in (68), repeated from (49b), does license variable binding.

(68) Ntskû kárro$_i$ xa ntàsaba'ă xito'o tún$_i$ __$_i$.
every car PERF repair.PAST owner it.WOOD
'Its$_i$ owner repaired every car$_i$.'

In this way, we see that pronoun doubling cannot be reduced to a form of A-movement, as pronoun doubling does not allow the same expansion of variable binding configurations as A-movement.

In addition to its failure to license novel variable binding relations, pronoun doubling also fails to call off otherwise licit variable-binding configurations. First, consider the behavior of clitic doubling in Modern Greek as reported in Alexiadou and Anagnostopoulou (1997), demonstrated in (69).

(69) a. Sistisa [kathe gineka]$_i$ [ston melondiko andra tis$_i$
introduced.1SG every woman.ACC to.the future husband.DAT hers
]$_j$.

'I introduced every woman to her future husband.'

b. * **Tu**$_j$ sistisa [kathe gineka]$_i$ [ston melondiko
to.him.DAT introduced.1SG every woman.ACC to.the future
andra tis$_i$]$_j$.
husband.DAT hers
Intended: 'I introduced every woman to her future husband.'

In (69a), we see that without clitic doubling, the quantified direct object DP *kathe gineka* 'every woman' may bind a variable in the indirect object. Crucially, in (69b), we see that clitic doubling the indirect object *ston melondiko andra tis* 'to her future husband' renders this variable-binding impossible.

Alexiadou and Anagnostopoulou (1997) use the ungrammaticality of (69b) to argue for a Move-and-Reduce analysis of clitic doubling in Greek. This is because the ill-formedness of (69b) follows from the logic of this sort of analysis because it will produce the pre-Reduction in (70).

(70) [*Ston melondiko andra tis*$_i$] sistisa [kathe gineka]$_i$ [ston melondiko
to.the future husband her every woman
andra tis$_i$].

In (70), the quantifier no longer c-commands the variable. This derives why clitic doubling in this configuration calls of the otherwise licit variable-binding relation.

Importantly, pronoun doubling cannot affect variable binding in this same way. This is shown in (71).

(71) ***Context***: I am a dance teacher, and all of my students are excellent dancers. We are putting on a performance, and I know that all of the parents will love it.

a. Ná konì ntskû táta$_i$ se'e rà$_i$, kuu sijǐ kwê'e ini rà.
SUBJ see.IRR all father child his COP.IRR happy very inside him
'When every father$_i$ sees his$_i$ child, he will be very happy.'

b. Rà$_i$ xa xìnì ntskû táta$_i$ se'e rà$_i$!
he PERF see.PAST all father child his
'All the fathers$_i$ have already seen their$_i$ children!'

c. Ñà$_i$ xa xìnì ntskû táta se'e$_i$ rà!
it PERF see.PAST all father child his
'All the fathers$_i$ have already seen their$_i$ children!'

In (71) we see that pronoun doubling the object *se'e rà* 'his child,' which contains a pronoun bound by the quantifier *ntskû táta* 'every father' in subject position. Crucially, in (71c), we see that doubling the object over its binding quantifier causes no problem, despite that under a Move-and-Reduce analysis, there would be a higher instance of the object containing the bound pronoun in pre-verbal position, outside of the c-command domain of the subject quantifier. Despite this, (71c) is still grammatical.

In this way, we see that pronoun doubling systematically fails to show the effects on variable-binding observed in true clitic doubling systems. This casts serious doubt on the possibility of analyzing pronoun doubling with the same Move-and-Reduce technology necessary for clitic doubling in languages like Modern Greek.

The last variable-binding relation worth considering is the behavior of cataphora. Recall from earlier discussions of Amharic in chapter 2 that clitic doubling ameliorates otherwise marked cataphoric relations in that language.

(72) Amharic object doubling affects variable-binding (Kramer, 2014)

a. Ṯigist$_i$ tämari -wa$_i$ -n ayy -ätʃtʃ.
T. student -her -ACC see -3SG.FEM
'Ṯigist$_i$ saw her$_i$ student.'

b. ?* Tämari -wa$_i$ Ṯigist$_i$ -in ayy -ä.
student -her T. -ACC see -3SG.MASC
Intended: 'Her$_i$ student saw Ṯigist$_i$.'

c. Tämari -wa$_i$ Ṯigist$_i$ -in ayy **-at**$_i$.
student -her T. -ACC see -3SG.MASC
'Her$_i$ student saw Ṯigist$_i$.'

In (72b), we see that cataphora is judged as severely degraded in Amharic. Interestingly, although unexplainedly, clitic doubling improves this otherwise marked cataphoric

structure.

Despite the ungrammaticality of cataphora being unexplained, this relation is worth investigating for two reasons. First, similar relations to cataphora are widely reported in other clitic doubling systems, notably in Rioplatense Spanish, as reported by Suñer (1988). Consider (73), reproduced from chapter 2.

(73) Rioplatense Spanish (Suñer 1988, pg. 421)

a. Todos$_i$ quieren a su$_i$ madre.
everyone likes their mother
'Everyone likes their mother.'

b. *?Su$_i$ madre quiere a todos$_i$.
their mother likes everyone
Intended: 'Their$_i$ mother likes everyone$_i$.'

c. Su$_i$ madre **los$_i$** quiere a todos$_i$.
their mother them likes everyone
'Their$_i$ mother likes everyone$_i$.'

(73) demonstrates an identical pattern to the Amharic in (72). Therefore, this amelioration of otherwise marked cataphoric constructions seems to be a signature property in many clitic doubling systems.

Interestingly, cataphora is also judged as severely degraded in SMP Mixtec. This is shown in (74).

(74) a. Xìnì nána míí Julio$_i$ rà$_i$.
see.PAST mother the J. him
'Julio$_i$'s mother saw him$_i$.'

b. * Xìnì nána rà$_i$ míí Julio.
see.PAST mother his the J.
Intended: 'His$_i$ mother saw Julio$_i$.'

Although it is, at present, somewhat unclear why (74b) is not possible, this judgment is robustly replicated amongst speakers, and seems to be a property of binding in this

language.

Now, if pronoun doubling were comparable to clitic doubling in Amharic and Rioplatense Spanish, then we would expect pronoun doubling to ameliorate the marked cataphoric structure in (74b). Interestingly, this is not the case, as shown in (75).

(75) a. Ná konì nána míí Julio$_i$ xí'i rà$_i$ ndxíxi, ñá saha ñá.
NÁ see.IRR mother the J. drink.PRES he alcohol she get.mad.IRR she
'If Julio's mother sees him drinking alcohol, she'll get mad.'

b. Ñá$_i$ xa nàni'i [mí'i nána míí Julio$_j$]$_i$ rà$_j$.
she PERF find.PAST the mother the J. him
'Julio's mother already found him!'

c. Rà$_j$ xa nàni'i [mí'i nána míí Julio$_j$]$_i$ rà$_j$.
he PERF find.PAST the mother the J. him
'Julio's mother already found him!'

d. * Ñá$_i$ xa nàni'i [mí'i nána rà$_j$]$_i$ míí Julio$_j$.
she PERF find.PAST the mother his the J.
Intended: 'His$_j$ mother already found Julio$_j$.'

e. * Rà$_j$ xa nàni'i [mí'i nána rà$_j$]$_i$ míí Julio$_j$.
she PERF find.PAST the mother his the J.
Intended: 'His$_j$ mother already found Julio$_j$.'

(75a) sets up a context in which both Julio and his mother are topical. But (75e) remains impossible, even though under a Move-and-Reduce analysis of pronoun doubling the object *míí Julio* in this example would both precede and command the pronoun within the subject. This is illustrated in (76).

(76) *Míí Julio$_i$* xa nàni'i [míí nána rà$_i$] míí Julio.
the J. PERF find.PAST the mother his the J.
Intended: 'His$_i$ mother already found Julio$_i$.'

If (76) were an accurate representation of how (75e) were derived, then we would predict it to be grammatical in the same way that (77) is. As is clear from (75e), this is clearly not the case.

(77) Nàni'i míí Julio$_i$ nána rà$_i$.
find.PAST the J. mother his
'Julio$_i$ found his$_i$ mother.'

In this subsection, we have seen that pronoun doubling systematic fails to show any of the effects on variable-binding observed in clitic doubling systems. This makes conflating pronoun doubling with A-movement not a viable option, as it was for clitic doubling systems.

3.3.3 Pronoun doubling may not target objects of P and possessors

Recall from §3.2.3 of this chapter that quantifier fronting, which I argued to be an A-movement process, may extract possessors and objects of prepositions. These data are repeated in (78) from (50-51).

(78) a. U'un rà tsyahá$_i$ ndàkinúû láncha ña'a __$_i$.
five he man sink.PAST boat POSS
'The five men's boat sank.'

b. Ní iin nà$_i$ kò ní ka'an ñá Maria xí'in __$_i$.
not.even one them NEG PAST speak she M. with
'Maria spoke with no one.'

It was suggested above that the ability to extract from possessors and objects of prepositions is a signature property of A-movement in this language. Therefore, we have the expectation that if pronoun doubling were derived through A-movement, then we should see extraction of possessors and objects of prepositions here as well. This turns out to not be the case.

Let us first consider PP objects. (79-83) demonstrate the five primary prepositions that occur as the heads of object PPs: *nuhŭ* 'on, of (lit. face),' *xí'in* 'with (lit. side),' and *sátá* 'on (lit. back),' *xa'á* 'about, under (lit. foot)' and *nda'ă* 'to (lit. hand)'.[16]

[16]Note that prepositional phrases headed by *nuhŭ*, literally 'face,' show a range of acceptability. In

(79) a. **Rà**$_i$ íyǐ'bi rà$_i$ [nuhŭ nána rà].
he afraid.PRES he of mother his
'He's afraid of his mother.'

b. * **Ñá**$_i$ íyǐ'bi rà [nuhŭ nána$_i$ rà].
she afraid.PRES he of mother his

(80) a. **Ñá**$_i$ xa kà'a ñá Juana$_i$ [xí'in rà Mârco].
she PERF talk.PAST she J. with he M.
'Juana already talked with Marcos.'

b. * **Rà**$_i$ xa kà'a ñá Juana [xí'in rà Mârco$_i$].
she PERF talk.PAST she J. with he M.

(81) a. **Rí**$_i$ ntsíku míí tsǐna$_i$ [sátá míí rà lo'o].
it.AML follow.PRES the dog behind the he little
'The dog is following the boy.'

b. * **Rà**$_i$ ntsíku míí tsǐna [sátá míí rà lo'o$_i$].
it.AML follow.PRES the dog behind the he little

(82) a. **Rà**$_i$ ká'a rà mástro$_i$ [xa'á míí libro].
he talk.PRES he teacher about the book
'The teacher is talking about the book.'

---

particular, if the object of *nuhŭ* is an animal, it may be fully grammatical.

(1) a. **Rí**$_i$ xa kàyu'u míí rà lo'o [nuhŭ tsǐna$_i$ sâna rà].
it.AML PERF shout.PAST the he little on dog POSS.ANI his
'The boy has already shouted at his dog.'

b. **Rí**$_i$ í'yǐbi rà lo'o [nuhŭ míí kwáyi$_i$].
it.AML afraid.PRES he little of the horse
'The boy is afraid of the horse.'

As Iara Mantenuto (p.c.) pointed out, this may be an artifact of *nuhŭ* literally meaning 'face.' In many Mixtec languages, including SMP Mixtec, round objects like faces, *yohó* 'moon,' and others, are frequently referenced with animal pronouns. Therefore, in these cases, the doubled pronoun *rí* 'it.AML' may be cross-referencing *nuhŭ* 'face' itself, rather than the DP embedded within it. At the same time, it is not clear why this agreement with P would only be available if the nominal within PP is an animal. Contrast (i) with (ii).

(2) * **Rí**$_i$ xa kàyu'u míí rà lo'o [nuhŭ$_i$ míí kò'obi rà].
it.AML PERF shout.PAST the he little face the sister his
Intended: 'The boy already shouted at his sister.'

I leave a complete analysis of this 'animal amelioration effect' to future research.

b. * **Ñà**$_i$ ká'a rà mástro [xa'á míí libro$_i$].
it talk.PRES he teacher about the book

(83) a. **Rà**$_i$ xa tàxi rà doktór$_i$ mánta [nda'ă nà yátá].
he PERF give.PAST he doctor blanket to they old
'The doctor already gave a blanket to the elders.'

b. **Ñà**$_i$ xa tàxi rà doktór mánta$_i$ [nda'ă nà yátá].
it PERF give.PAST he doctor blanket to they old
'The doctor already gave the blanket to the elders.'

c. * **Nà**$_i$ xa tàxi rà doktór mánta [nda'ă nà yátá$_i$].
they PERF give.PAST he doctor blanket to they old
Intended: 'The doctor already gave the blanket to the elders.'

As is clear from (79-83), pronoun doubling may not cross-reference a nominal within a PP in object position. This is true both for the direct objects in (79-82) and the indirect object in (83).

SMP Mixtec also allows PP subjects. This is SMP Mixtec's reflex of the common cross-Mixtec pattern in which experiencer subjects occur within a PP headed by *ini* 'inside'.[17] As in the case of object PPs, pronoun doubling cannot cross-reference a nominal within a subject experiencer PP.

(84) a. ☒ Kusijǐ [in ì].
happy inside me

[17]Two cross-Mixtec notes are in order. First, as DiCanio (p.c.) points out, *ini* in other Mixtec languages means 'heart.' In SMP Mixtec, *ánima* is the word for 'heart.' Speakers intuit that *ini* has no body part correspondent.

Second, Mixteca Alta varieties often have these experiencer PPs as the subjects of transitives. This is demonstrated in (ia) for Chalcatongo Mixtec from Macaulay (2005). Mixteca Baja varieties, like SMP Mixtec, seem to never allow this. The SMP Mixtec translation of the Chalcatongo Mixtec in (ia) is provided in (ib).

(1) a. Nì- naa [inì čáá] ndo'ò.
COMP- lose insides man basket

'The man forgot his basket.' (Chalcatongo Mixtec, Macaulay 2005)

b. Nàntŏso rà tsyahá [xa'ă míí ndò'ò rà].
forget.PAST he man about the basket his
'The man forgot his basket.' (San Martín Peras Mixtec)

'I am happy.'

b. * Yù'u$_i$ kusijǐ [in ì$_i$].
I happy inside I

(85) a. ☒ Ntsí'i [ini ñá Juana].
worried inside she J.
'Juana is quite worried.'

b. * Ñá$_i$ ntsí'i [ini ñá Juana$_i$].
she worried inside she J.

(86) a. ☒ Íyo ba'á [ini tát ì].
is good inside father my
'My father feels well.'

b. * Rà$_i$ íyo ba'á [ini tát$_i$ ì].
he is good inside father my

In each of the pairs in (84-86), we see that pronoun doubling may not occur with a DP embedded within a PP subject. This contrasts with other experiencer subjects that are not embedded within a PP. Some of these are demonstrated in (87).

(87) a. **Rà**$_i$ sáha rà Juan.
he mad.PRES he J.
'Juan is mad.'

b. **Ñá**$_i$ kùnaha ñá Isabela$_i$.
she tired.PAST she I.
'Isabela was tired.'

In comparing the grammaticality of (87) with the ungrammaticality of the (b) examples in (84-86), we see that experiencers may generally undergo pronoun doubling, but only if they are not embedded within a PP.

Interestingly, it is possible to pronoun an experiencer subject embedded within *ini* 'inside' under very specific circumstances. Specifically, if *ini* incorporates into the verb, then doubling the subject becomes obligatory once again.

(88) a. Yù'u$_i$ kusijǐ ini -k ì$_i$.
I happy inside INTR I
'I am happy.'

b. * ⊠ Kusijǐ ini -k ì.

In (88a), we see that *ini* 'inside' has been incorporated into the verb because of its placement to the left of the intransitive marker *-k*.[18] Once this incorporation takes place, pronoun doubling again demonstrates its usual obligatory patterning for a local person subjects like the first person pronoun (see chapter 5). This effect of incorporation acts as the exception that proves the rule: pronoun doubling is impossible for nominals embedded within PP.

Now consider the behavior of possessors. As it turns out, pronoun doubling is categorically banned from cross-referencing a possessor. The suite of data in (89-92) demonstrates that pronoun doubling of possessors in any position is impossible .

(89) No doubling of possessor in unaccusative subject

a. Áto rí nì xi'i tsǐna sâna rà Julio, íyo ntsí'ì ba'á rà.
if it.AML PAST die dog POSS he J. is sad quite he
'If Julio's dog dies, he will be quite sad.'

b. **Rí$_i$** xa nì xi'i tsǐna$_i$ sâna rà.
it.AML PERF PAST die dog POSS he
'His dog already died!'

c. * **Rà$_i$** xa nì xi'i tsǐna sâna rà$_i$.
he PERF PAST die dog POSS he

(90) No doubling possessor in unergative subject

[18]Recall from the discussion in chapter 2 that the intransitive marker *-k* is rare in the data throughout for routine phonological reasons. SMP Mixtec, like most other Mixtec languages, prohibits both codas and onset consonant clusters. Therefore, it only occurs if it can be syllabified as an onset. Now, most vowel initial words like *ini* actually begin with a glottal stop: [ʔini]. Therefore, *-k* cannot be syllabified in the onset of this word, as an elicit cluster [kʔ] would result. Additionally, *-k* cannot be syllabified as a coda, as codas as universally banned in this language. As such, *-k* occurs only before three vowel-initial pronouns: *ì* [ì] 'I,' *ú* [ú] 'you,' and *é* [é] 'we (INCL).'

a. Áto ná ndakojo táta ñá Juana na'a, kusijǐ kwê ini ñá.
if SUBJ arise.IRR father she J. early happy very inside she
'If Juana's father gets up early, she will be very happy.'

b. **Rà**$_i$ xa ndàkojo táta$_i$ ñá Juana!
he PERF arise.IRR father she J.
'Juana's father has already gotten up!'

c. * **Ñá**$_i$ xa ndàkojo táta ñá Juana$_i$!
he PERF arise.IRR father she J.

(91) No doubling possessor in transitive object

a. Ntsìbì ñan ì, rà Mario ra rà Juán, rà xa ndànda'á rà.
both brother my he M. and he J. he PERF marry.PAST he
'Both of my brothers, Mario and Juan, are married.'

b. **Ñá**$_i$ xa xín ì ísi'i$_i$ rà Mario, so tá kon ì ísi'i rà Juán.
she PERF see.PRES I wife he M. but IRR.NEG see I wife he J.
'I already know Mario's wife, but I still haven't met Juan's wife.'

c. * **Rà**$_i$ xa xín ì ísi'i rà Mario$_i$ ...
he PERF see.PRES I wife he M.

(92) No doubling possessor in transitive subject

a. **Ñá**$_i$ xa nùhmǐ ísi'i$_i$ rà Macario tsǐna ló'ó.
she PERF hug.PAST wife he M. dog little
'Macario's wife already hugged the puppy.'

b. **Rí**$_i$ xa nùhmǐ ísi'i rà Macario tsǐna$_i$ ló'ó.
it.AML PERF hug.PAST wife he M. dog little
'Macario's wife already hugged the puppy.'

c. * **Rà**$_i$ xa nùhmǐ ísi'i rà Macario$_{\text{I}}$ tsǐna ló'ó.
he PERF hug.PAST wife he M. dog little
Intended: 'Macario's wife already hugged the puppy.'

The (a) examples of (89-92) set up a context in which the possessor and the possessum are highly topical. Despite this, in the (c) examples show we see that doubling a possessor is never grammatical.

Therefore, one of the distinctive hallmarks of A-movement in this language, the ability to extract possessors and objects of prepositions, may not be replicated in pronoun doubling constructions. I take this as good circumstantial evidence that pronoun doubling cannot be reduced to a camouflaged form of A-movement, as in Move-and-Reduce analyses of clitic doubling.

3.4 Where do we go from here?

In this chapter, we considered what evidence could be used to support an analysis in which pronoun doubling could be seen as a form of clitic doubling derived by A-movement of the doubled argument, followed by reduction of the higher copy to a clitic. See Uriagereka (1995), Sportiche (1996), Alexiadou and Anagnostopoulou (1997), Suñer (2000), Anagnostopoulou (2003), Preminger (2009), Harizanov (2014b,a), Kramer (2014), and Baker and Kramer (2018) for more details on this kind of analysis.

In this chapter we saw that the evidence suggests that this hypothesis is incorrect. Consider where this leaves us in our investigation. We know that pronoun doubling is a process that replicates the φ-features of an argument in a clause. Furthermore, we know that there are two (and a half) ways that we can derive multiple exponence of φ-feature bundles. These are summarized in (93) from §1 and §2.

(93) Multiple exponence of φ-feature bundles can be derived by:

a. Providing a value to some node's unvalued φ-features (φ-agreement).

b. Spell-out of multiple copies in a movement chain (clitic doubling).

i. This movement chain could be formed by A-movement.

ii. This movement chain could be formed by A'-movement.

In this chapter, (95bi) has been eliminated as a possibility. But it could still be the case that pronoun doubling is derived through movement to an A'-position, as

in (95bii). To investigate this possibility, we will need a reasonable overview of the landscape of A'-movement in SMP Mixtec. The chapter which follows aims to provide that necessary background.

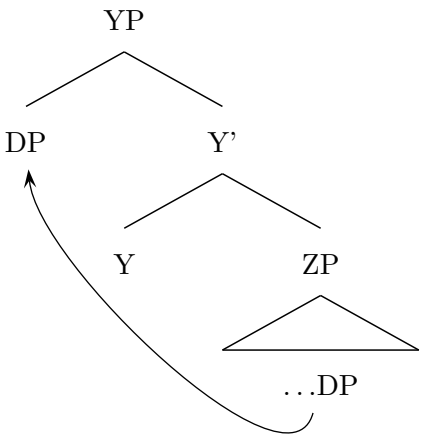
Chapter 4

Diagnosing A'-movement in San Martín Peras Mixtec

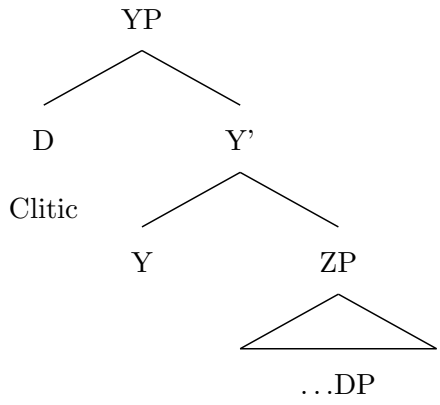
In this chapter, we will investigate the possibility that pronoun doubling is derived via A'-Move-and-Reduce. The mechanisms of this account would be fundamentally the same as Move-and-Reduce presented in §2 of chapter 2, although the movement chain that leads to the reduced higher copy would be A'-movement, rather than A-movement. This is shown in (1).

(1) Where Spec,YP is an A'-position

a.



b.



This sort of A'-chain reduction is, in fact, attested. Consider wh-clitic doubling in Illasi and Monnese, both Northern Italian dialects analyzed in Poletto and Pollock (2004).

(2) Wh-clitic doubling in Illasi (Verona)

a. **S'**$_i$ a -lo fato **che**$_i$?
what.CL has -he done what
'What has he done?'

b. **Ndo**$_i$ e -lo ndat **endoe**$_i$?
where.CL is -he gone where
'Where has he gone?'

(3) Wh-clitic doubling in Monnese (Brescia)

a. **Ch'**$_i$ et fat **què**$_i$?
what.CL have -you done what
'What have you done?'

b. **Ngo**$_i$ fet majà **ngont**$_i$?
where.CL do-you eat where
'Where do you eat?'

In both of the languages in (2-3), the head of an A'-movement dependency is realized with a wh-clitic, while the full wh-phrase is pronounced in its first Merge posiiton. Therefore, the machinery to apply a Move-and-Reduce account of clitic doubling to A'-movement must, in principle, be available in the theory of grammar. Clearly, then, this path is worth exploring.

To explore this possibility, we would want to ask if the relation between the pronoun and its double exhibits the standard array of properties associated with A'-dependencies. Unfortunately, we cannot simply apply the diagnostics familiar from work on other languages. There are several reasons for this. First, SMP Mixtec lacks one of the signature properties of A'-movement dependencies: parasitic gaps (Engdahl 1983 and a great deal of subsequent work).

(4) No parastic gaps in SMP Mixtec

a. Ntsyâ ntsìka$_i$ kòñ ú t_i [tá xàx ú rí$_i$]?
which banana like.PAST you while eat.PAST you it.AML
'Which banana$_i$ did you enjoy t_i while eating it$_i$?'

b. * Ntsyâ ntsìka$_i$ kòñ ú t_i [tá xàx ú ☒$_i$]?
Intended: 'Which banana$_i$ did you enjoy t_i while eating ☒$_i$?'

Second, in chapters 1 and 3 we saw that pronoun doubling is sensitive to discourse properties like referentiality and topicality. Therefore, if pronoun doubling were derived by some kind of A'-movement, this movement might be similar to Topicalization in English. Importantly, Topicalization famously does not exhibit Weak Crossover effects (Lasnik and Stowell, 1991).

(5) Topicalization does not induce Weak Crossover (Lasnik and Stowell 1991)

a. This bookt_i, I expect [its$_i$ author] to buy t_i.

b. John$_i$, I believe [his$_i$ mother] loves t_i.

Therefore, if pronoun doubling constructions failed to display Weak Crossover (as it in fact does), we could not necessarily use this as evidence against an A'-movement derivation. Finally, as we saw in §3, pronoun doubling is strictly clause-bounded. Therefore, we cannot consider questions of unboundedness and islandhood, also key properties of A'-movement cross-linguistically.

In order to diagnose A'-movement in pronoun doubling constructions, then, we must turn to language-internal identifiers of A'-movement. This will allow us to compare the profile of pronoun doubling to that of A'-movement. Should the two align, then we have reason to propose an analysis like (1). But if they show irreconcilable differences, then clearly we should not attempt to reduce one to the other.

Happily, SMP Mixtec provides such an identifier: a prohibition on A'-extraction of external arguments. §4.1 examines this restriction in detail and draws parallels be-

tween it and other cross-linguistic phenomena like syntactic ergativity and split intransitivity. With this much in place, §4.2 and §4.3 examine A'-dependencies in two constructions: restrictive relative clauses and wh-questions, with a special focus on how each finds a way to circumvent the restriction against A'-extraction of external arguments. The fact that such strategies are needed and in place demonstrates that the restriction holds.

We can then ask if a similar restriction holds of pronoun doubling. To anticipate, it does not, as shown in §4.4. This lack of parallelism strongly suggests that A'-movement has no role in the derivation of pronoun doubling. We must therefore seek an alternative, and identifying that alternative is the work of §5. Meanwhile, the business of establishing that conclusion will have as a useful by-product an initial overview of the A'-binding constructions of SMP Mixtec, one which can, hopefully, form the basis for more detailed investigation in the future.

4.1 Syntactic "ergativity" in SMP Mixtec and beyond

We are now investigating the thesis in (6).

(6) **A'-extraction restriction in SMP Mixtec**: External arguments may not undergo A'-extraction.

Such restrictions are not unknown. (6) bears a strong family resemblance to restrictions well-known from ergative languages. We consider these apparent parallels in §4.1.1. With this in place, §4.1.2 discusses why SMP Mixtec is not properly classified as a syntactically ergative language, but rather as a split intransitive language.

4.1.1 Background on syntactic ergativity

Most contemporary work on ergativity assumes that ergativity is properly conceived of in morphological terms, as in (7).

(7) **Morphological ergativity**: The phenomenon in which transitive subjects receive distinct morphological marking.

In languages which display morphological ergativity, the distinct morphological marking for transitive subjects is referred to as ergative case if the marking is on the noun, or "ergative agreement" in the case of head-marking languages (Nichols 1986).

In a strictly ergative system, nouns which don't receive ergative case or trigger ergative agreement are said to receive absolutive case, or trigger absolutive agreement. Consider nominal marking in Tongan, which displays a classical ergative pattern. Note that Tongan sentences alternate between VOS and VSO order, and these alternations do not affect ergative case assignment.

(8) Morphological ergativity in Tongan (Anderson 1976)

a. Na'e lea **'a** etalavou.
PAST speak ABS young.man
'The young man spoke.'

b. Na'e alu **'a** Tevita ki Fisi.
PAST go ABS David to Fiji
'David went to Fiji.'

c. Na'e tamate'i **'a** Kolaiate **'e** Tevita.
PAST kill ABS Goliath ERG David
'David killed Goliath.'

d. Na'e ma'u **'e** Siale **'a** e me'a'ofa.
PAST receive ERG Charlie ABS the gift
'Charlie received the gift.'

In (8a-b) we see that intransitive subjects receive absolutive case. In (8c-d) we see that transitive subjects, both agentive and non-agentive, are assigned ergative case.

Much early and and contemporary work on ergative languages takes ergativity to be "skin-deep" in that ergative subjects behave syntactically like nominative subjects in languages such as English. Notably, Anderson (1976) proposes that ergative

case assignment is a fundamentally post-syntactic process, foreshadowing contemporary analyses of ergative case in terms of the concept of 'dependent' case (Marantz 1991, Baker 2015, inter alia). But in addition to the unique morphological marking of transitive subjects characteristic of morphological ergativity, there is an additional syntactic complexity to be considered: many ergative langauges also treat transitive subjects different with respect to extraction processes. An early treatment is Dixon (1979), whose observations about Dyirbal, a morphophologically ergative Pama-Nyugnan language of Northeast Australia, we quickly review here.

Dixon observes that ergative subjects in Dyirbal may not be relativized.

(9) a. Nguma -$\emptyset_i$ [$\_\__i$ yabu -ngu bura -ngu] dunggara -nyu.
father -ABS mother -ERG see -REL cry -PAST
'Father, who was seen by mother, was crying.'

b. * Nguma -ngu$_i$ [yabu -$\emptyset$ $\_\__i$ bura -ngu] dunggara -nyu.
father -ERG mother -ABS see -REL cry -PAST
Intended: 'Father, who saw mother, was crying.'

(9a) demonstrates that the gap inside the relative clause may correspond to the absolutive argument, but (9b) shows that the gap cannot be in the ergative position.

Dyirbal offers a 'way out' - a syntactic repair which allows expression of the intended meaning in (9b). As the gap inside the relative clause may not be ergative, Dyirbal requires that the relative clause undergo antipassivization. (10a) illustrates the antipassive constructions, and (10b) shows its use in permitting relativization of what would otherwise have been an ergative argument.

(10) a. Nguma -$\emptyset$ bural -**nga** -nyu yabu -gu.
father -ABS see -ANTIPASS -PAST mother -DAT
'Father saw mother (antipassive).'

b. Nguma -$\emptyset_i$ [$\_\__i$ bural -**nga** -**ngu** yabu -gu] dunggara -nyu.
father -ABS see -ANTIPASS -REL mother -DAT cry -PAST
'Father, who saw mother, was crying.'

Interestingly, this turns out to be a common pattern among morphologically ergative languages cross-linguistically. For instance, West Greenlandic demonstrates an exactly similar pattern: ergatives may not be relativized, so the relative clause must be antipassivized. In (11), we see that the gap inside a relative clause may correspond to an absolutive argument, but not an ergative-marked argument.

(11) West Greenlandic ban on ergative relatives (Bittner 1994, by way of Deal 2016)

a. miiqqa -t [__ABS sila -mi pinnguar -tu -t]
child -PL.ABS outdoors -LOC play -REL.INTRANS -PL
'the children who are playing outdoors'

b. miiqqa -t [Juuna -p __ABS paari -sa -i]
child -PL.ABS J. -ERG look.after -REL.TRANS -3SG/PL
'the children that Juuna is looking after'

c. * angut [__ERG aallaat tigu -sima -sa -a]
man.ABS gun.ABS take -PRF -TRANS -3SG/SG
Intended: 'the man who took the gun'

As in the Dyirbal in (10), we see that extraction of 'the logical subject' in West Greenlandic is made possible by applying the antipassive in the relative clause.

(12) West Greenlandic antipassive as a repair for ergative extraction

a. Juuna -p miiqqa -t paari -v -a -i.
J. -ERG child -PL.ABS look.after -IND -TRANS -3SG/PL
'Juuna is looking after the children.'

b. Juuna miiqqa -nik paar **-si** -v **-u** -q.
J. child -PL.INSTR look.after -ANTIPASS -IND -INTRANS -3SG
'Juuna is looking after the children.'

c. angut [__ABS aallaam -mik tigu **-si** -sima **-su** -q
man.ABS gun -INSTR take -ANTIPASS -PRF -REL.INTRANS -3SG
]

'the man who took the gun'

Given this, Polinsky (2016) defines this 'syntactic ergative' as in (13).

(13) **Syntactic ergativity revised**: the inaccessibility of ergative arguments to A'-movement with a gap in the base position, as contrasted with the accessibility of absolutive arguments to such movements. (Polinsky 2016)

Antipassivization is not the only attested repair strategy to circumvent (13), as not all ergative languages have an antipassive. For instance, Tongan, which lacks an antipassive, requires a resumptive pronoun with ergative relative clauses (data from Polinsky 2016).

(14) Tongan transitive

a. 'Oku fakamolemole'i 'e Mele 'a e kaiha'a.
PRES forgive ERG M. ABS DET thief
'Mele forgives the thief.'

b. * e ta'ahine$_i$ ['oku fakamolemole'i __$_i$ 'a e kaiha'a]
DET girl PRES forgive ABS DET thief
Intended: 'the girl who forgives the thief'

c. e ta'ahine$_i$ ['oku **ne**$_i$ fakamolemole'i 'a e kaiha'a]
DET girl PRES she forgive ABS DET thief
'the girl who (she) forgives the thief'

d. e kaiha'a$_i$ ['oku fakamolemole'i 'e Mele __$_i$]
DET thief PRES forgive ERG M.
'the thief that Mele forgives.'

e. * e kaiha'a$_i$ ['oku fakamolemole'i 'e Mele **'a ia**]
DET thief PRES forgive ERG M. ABS him
Intended: 'the thief that Mele forgives (him).'

(15) Tongan intransitive

a. 'Oku tangi 'a Mele.
PRES cry ABS M.
'Mele is crying.'

b. e ta'ahine$_i$ ['oku tangi $\_\__i$]
DET girl PRES cry
'the girl who is crying'

c. e ta'ahine ['oku (***ne**) tangi (* 'a ia)]
DET girl PRES she cry ABS she
Intended: 'the girl who (she) is crying'

(14-15) demonstrate a fundamentally similar pattern to Dyirbal and West Greenlandic. In Tongan we see that, just like Dyirbal and West Greenlandic, extraction of the ergative argument is impossible. But Tongan also circumvents the restriction in (13) not by requiring a voice manipulation, but by allowing and requiring a resumptive pronoun in the ergative position within the relative clause.

4.1.2 An extraction restriction in SMP Mixtec

Despite the fact that SMP Mixtec is not a morphologically ergative language, it exhibits a pattern of restriction and repair that is similar in many respects to those just described for ergative languages. As in Dyirbal, West Greenlandic, and Tongan, a gap inside a relative clause can never correspond to a transitive subject, but extraction of intransitive subjects and internal arguments is not restricted.

(16) a. Kàni rà Mârco ñá Maria.
hit.PAST he M. she M.
'Mârco hit Maria.'

b. Ñá$_i$ xákù ñá lo'o$_i$ [kàni rà Mârco $\_\__i$].
she cry.PRES she little hit.PAST he M.
'The girl who Mârco hit is crying.'

c. * Rà$_i$ xáku rà lo'o$_i$ [kàni $\_\__i$ ñá Maria].
he laugh.PRES he little hit she Maria
Intended: 'The boy who hit Maria is laughing.'

d. Rí$_i$ yata xkuu míí chútun$_i$ [nì xi'i $\_\__i$].
it.AML old COP.PAST the cat PAST die

'The cat that died was old.'

(16a) is the basic transitive, demonstrating VSO order. (16b) and (16d) show that the gap in a relative clause can be the direct object of the transitive or the intransitive subject of *xi'i* 'die.' (16c) shows that the gap inside the relative may not correspond to the transitive subject, just as in Dyirbal, West Greenlandic, and Tongan. Just as in Tongan, the repair strategy for (16c) is to use a resumptive pronoun, as shown in (17).[1] (17) provides the grammatical alternative to (16c).

(17) Rà$_i$ xáku rà lo'o$_i$ [kàni **rà**$_i$ ñá Maria].
he laugh.PRES he little hit he she Maria
'The boy who hit Maria is laughing.'

Also like Tongan, resumptive pronouns are banned when a gap is otherwise available. Compare the ungrammatical (18a) and (18b) to the grammatical ones in (16b) and (16d).

(18) a. * Ñá$_i$ xákù ñá lo'o$_i$ [kàni rà Mârco **ñá**$_i$].
she cry.PRES she little hit.PAST he M. she
Intended: 'The girl who Mârco hit is crying.'

b. * Rí$_i$ yata xkuu míí chútun$_i$ [nì xi'i **rí**$_i$].
it.AML old COP.PAST the cat PAST die it.AML
Intended: 'The cat that died was old.'

As we shall see, the pattern in SMP Mixtec is more complex than this cursory demonstration can capture. Instead, this language is best classified as not a syntactically ergative language.

[1]There is a strong preference to front the resumptive pronoun to the beginning of the relative clause. Similar preferences for fronting resumptives have been reported in Hebrew as well (Doron, 1982). Throughout, I put the resumptive in the canonical position of the transitive subject, after the verb, for ease of reading.

4.1.3 Split-intransitivity in SMP Mixtec and beyond

While there are surely interesting parallels between the extraction restiction in SMP Mixtec and those found in ergative languages, there are also two important reasons to not conflate the two effects.

First, of course, SMP Mixtec lacks any morphological case or head-marking. Therefore, it is hard to consider this language ergative in any of the relevant senses. See Deal (2016) for an important discussion of this point. Second, the presentation in (16), though accurate, is incomplete. In (16d), I demonstrated that a gap must be used to relativize an intransitive subject. Crucially, though, not all intransitive subjects behave identically in this regard. Consider the relative clauses in (19) and (20).

(19) Intransitive with obligatory gap in relative clause

a. Nì ka' ìn xí'in míí rà lo'o$_i$ [ínchichi __$_i$ xíȟǐn be'e].
PAST talk I with the he little stand.PRES.SG beside house
'I talked with the boy that is standing next to the house.'

b. * Nì ka' ìn xí'in míí rà lo'o$_i$ [ínchichi **rà**$_i$ xíȟǐn be'e].
PAST talk I with the he little stand.PRES.SG he beside house
Intended: 'I talked with the boy that (he) is standing next to the house.'

c. Tún$_i$ tààba ndú míí ítŭn [nàkààba __$_i$ ichǐ yo'o].
it.WOOD remove.PAST we.EXCL the tree fall.PAST path this
'We removed the tree that fell in the road.'

d. * Tún$_i$ tààba ndú míí ítŭn [nàkààba **tún**$_i$ ichǐ yo'o].
it.WOOD remove.PAST we.EXCL the tree fall.PAST it.WOOD path this
Intended: 'We removed the tree that (it) fell in the road.'

(20) Intransitive with obligatory resumptive in relative clause

a. Kusihǐ kwê ini míí rà lo'o$_i$ [sákwa'a **rà**$_i$].
happy very in the he little study.PRES he
'The boy who (he) is studying is very happy.'

b. * Kusihǐ kwê ini míí rà lo'o$_i$ [sákwa'a $\_\__i$].
happy very in the he little study.PRES
Intended: 'The boy who is studying is very happy.'

c. Ndákwa ndààba míí doktór$_i$ [nì iyǐ'bi **rà**$_i$].
suddenly jump.PAST the doctor PAST be.scared he
'The doctor who (he) got a scare jumped suddenly.'

d. * Ndákwa ndààba míí doktór$_i$ [nì yǐ'bi $\_\__i$].
suddenly jump.PAST the doctor PAST be.scared
Intended: 'The doctor who got a scare jumped suddenly.'

In comparing (19) to (20), we see that intransitive predicates come in two varieties. In (19), the intransitive predicates, *ínchichi* 'be standing' and *nakààba* 'fall[2],' behave like the intransitive predicate *xi'i* 'die' in (16d) in requiring a gap if their sole argument is relativized. By contrast, the predicates in (20), *sakwa'a* 'study' and *iyǐ'bi* 'be frightened,' behave differently - if their sole argument is relativized, a gap may not occur. Instead, as with the transitive subjects in (16), relativization by resumption is the only option.

In this way, we see that intransitive predicates must be split into two categories. For the first, exemplified in (19), the intransitive argument behaves identically to a transitive direct object with respect to relative clause extraction: both require a gap. For the second, seen in (20), the intransitive argument patterns instead with the transitive subject in resisting extraction and requiring resumption under relativization.

(21) Obligatory behavior of SMP Mixtec argument type under relativization

Transitive object	Intransitive subject	Transitive subject
GAP	GAP or *pro*	*pro*

The pattern in (21) is a reflex of the well-attested phenomenon of general phenomenon is well-attested phenomenon of split-intransitivity.

[2] *Nakààba* 'fall' is often produced as *kanakààba*. As far as I have been able to tell, there is no syntactic or semantic difference between the two pronunciations, and both are used throughout this dissertation.

(22) **Split intransitivity**: A system in which some intransitive subjects are treated like agents and some intransitive subjects are treated like direct objects (Merlan 1985, cited through Mithun 1991).

Split intransitive systems have been reported in a variety of genetically and geographically diverse languages, from Bats (Northeast Caucasian, Anderson 1976), Dakota (Siouan, Anderson 1976, Mithun 1991), Central Pomo (Pomoan, Mithun 1991), Choctaw (Heath 1977, Davies 1986), Cherokee (Iroquoian, Scancarelli 1987), Ika (Chibchan, Frank 1990), Chol (Mayan, Vázquez Álvarez and Maldonado 2013), Colloquial Guaraní (Mithun 1991), Lhasa Tibeatan (Sino-Tibetan, DeLancey 1985), Saweru (Papuan, Donohue 2001), and Acehnese (Austronesian, Durie 1985). Additionally, split intransitive systems are commonly attested in other Oto-Manguean languages, most notably Chocho (Mock 1982) and the Para-Mixtecan language Amuzgo (Smith and Tapia 2002).

In all split intransitive languages, the division of intransitive predicates into two sub-categories is determined on semantic grounds (DeLancey 1981, 1985; van Valin 1990; Mithun 1991, inter alia). For SMP Mixtec, (23) presents a non-exhaustive list of intransitive predicates whose arguments pattern with transitive subjects and transitive objects respectively.

(23) List of intransitive predicates whose sole argument patterns:

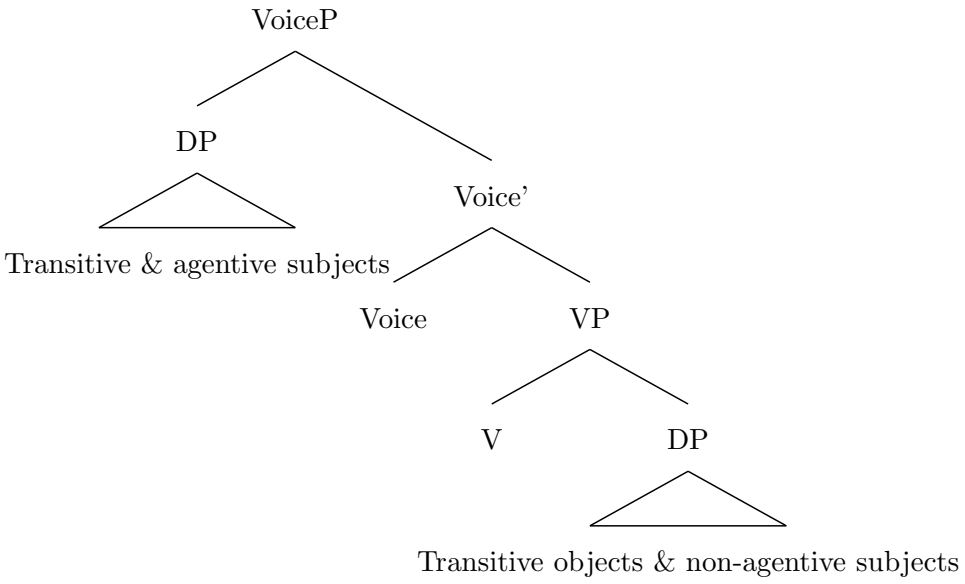
a. **With transitive subjects** (show A'-extraction restriction)
taxa'a 'dance,' *kichiñu* 'work,' *karakono* 'run,' *ixutsya* 'swim,' *xika* 'walk,' *xa'a* 'go,' *naki'i* 'depart,' *ya'a* 'pass' *ntachí* 'fly,' *xita*, 'sing' *ndaaba* 'jump,' *sasiki* 'play,' *xixi* 'eat,' *xi'ì* 'drink,' *ndakoho* 'arise,' *ndakuchichi* 'stand up (eventive),' *kuntu'u* 'sit down (eventive),' *saha* 'get mad (eventive),' *kixì* 'sleep' *kaxă* 'snore,' *kixa kú'ù* 'be pacient,' *ndàkŭ* 'be strong,' *chichí* 'be ripened,' *iyi'̌bi* 'be frightened'

b. **With transitive direct objects** (do not show A'-extraction restriction)

kanakààba ''fall, *xi'i* 'die,' *ku'u* 'be sick,' *kokon* 'be hungry,' *kunaha* 'get tired (eventive),' *inchichi* 'be standing (stative),' *intu'u* 'be sitting (stative),' *ndakinúú* 'sink,' *xa'anu* 'get big (eventive),' *xi'ita* 'grow (as in a plant),' *kaki'i ini* 'have hiccups,' *nduxáán* 'vomit,' *tsiìn xí'in nda'ă kwá'â* 'be righthanded,' *bixí xini* 'feel cold,' *i'ní xini* 'feel hot'

We will not attempt a detailed semantic discussion here, but the relevant distinction seems to be clear enough - it is based on what Mithun (1991) calls control over the event, volitionality, and performativity. Notably, this split is virtually identical to the division of intransitive predicates in Central Pomo (Mithun 1991). The most natural way to model this pattern syntactically is to adopt the Unaccusativity Hypothesis (Perlmutter 1978 as implemented by Burzio 1986). Updating those proposals in the conventional way, we will assume the syntax in (24).

(24)



Transitive objects and non-agentive subjects now emerge as similar (both are complements of V in initial representations), as we have a basis for understanding why some

intransitive subjects pattern with transitive subjects while others pattern with transitive objects. I will use the standard terms for the two classes of predicates, calling the intransitive predicates in (23a) unergatives and those in (23b) unaccusative.

There is an additional reason to adopt the syntax in (24). This comes from the observation that external arguments are islands for quantifier fronting, but internal arguments are not. This is shown briefly in (25).

(25) a. * Ntskû míí ná sí'i nákatsya ìhì $\_\__i$ tsyàà.
all the they woman wash.PRES husband clothes
Intended: 'The husbands of all the women are washing clothes.'

b. Ntsi'i nà bálí$_i$ nàkatsya ñá Maria tsyàà $\_\__i$.
all they little.PL wash.PAST she M. clothes
'Maria washed all the children's clothes.'

Crucially, unergative and unaccusative verbs divide with respect to this effect: subjects of unergative verbs are islands for quantifier fronting, while subjects of unaccusative verbs are not. This is shown in (26).

(26) a. * Ntskû nà$_i$ kárakono se'e $\_\__i$.
all them run.PRES child
Intended: 'Everyone's children are running.'

b. * Ní iin nà$_i$ kò kíchiñu [ìhì $\_\__i$].
not.even one them NEG work.PRES husband
Intended: 'No one's husband is working.'

c. Ní iin nà$_i$ kò ní xi'i tsĭna sâna $\_\__i$.
not.even one them NEG PAST die dog POSS.ANI
'No one's dog died.'

d. Ntskû ná sí'i$_i$ kú'u [se'e $\_\__i$].
all they.FEM woman be.sick.PRES child
'All the women's children are sick.'

These observations provide support for the syntactic proposal summarized in (24), especially if it is true that complements are transparent for outward movement in

a way that specifiers are not (i.e., a generalized version of Huang's 1982 condition on extraction domains). This proposal, in turn, allows us to state the extraction condition as in (27).

(27) **A'-extraction restriction in SMP Mixtec**: External arguments may not undergo A'-extraction.

In this way, we see that the A'-extraction restriction in SMP Mixtec, while tantalizingly similar to the more familiar ergativity condition on extraction, is actually rather different. While syntactic ergativity seems to be concerned with morphological case (see Deal 2016 especially), the A'-extraction restriction in SMP Mixtec is purely structural.

It remains for us to probe how this restriction plays out in two familiar domains for A'-movement, relative clauses and constituent questions, and then to return to our central theme, the nature of pronoun doubling. This investigation will bolster the claim that the A'-extraction restriction in (27) is a reliable language-internal diagnostic tool for identifying A'-movement.

4.2 Relative clauses

As we saw in (16-20), there are two types of relative clauses, depending on which argument is relativized.

(28) a. Rà$_i$ yata kwê'e míí tat$_i$ é [káxă **rà**$_i$].
he old very the father our.INCL snore.PRES he
'The man (respectfully, our father) who (he) is snoring is very old.'

b. * Rà$_i$ yata kwê'e míí tat$_i$ é [káxă __$_i$].

c. Ñá$_i$ ntá'yi míí ñá lo'o$_i$ [tà'abi **ñá**$_i$ yùtátá].
she cry.PRES the she little break.PAST she mirror
'The girl who (she) broke the mirror is crying.'

d. * Ñá$_i$ ntá'yi míí ñá lo'o$_i$ [tà'abi __$_i$ yùtátá].

In (28), we see that external argument relative clauses, built either from unergative subjects as in (28a-b) or transitive subjects like (28c-d), require resumption. The pronouns found in such cases are identical in form to regular pronouns, and are glossed throughout as such.[3]

In contrast to these constructions, internal argument relatives require a gap.

(29) a. * Ñà$_i$ xàxi ndú míí ya'ă [xì'ita **ñà**$_i$ yé'e yó'o].
it eat.PAST we.EXCL the chile grow.PAST it garden this
Intended: 'We ate the chiles that (they) grew in this garden.'

b. Ñà$_i$ xàxi ndú míí ya'ă [xì'ita __$_i$ yé'e yó'o].
'We ate the chiles that grew in this garden.'

c. * Tún$_i$ nàka'mi é míí itŭn$_i$ [xà'antsya míí Juan **tún**$_i$].
it.WOOD burn.PAST we.INCL the tree cut.PAST the J. it.WOOD
Intended: 'We burned the tree that Juan cut (it) down.'

d. Tún$_i$ nàka'mi é míí itŭn$_i$ [xà'antsya míí Juan __$_i$].
'We burned the tree that Juan cut down.'

The internal argument relatives in (29), whether they be unaccusative subject relatives like (29a-b) or transitive object relatives in (29c-d), differ from the external argument relatives in (28) in requiring a gap.

In this section, we will systematically review the properties of these two relative clause types, and ultimately defend a fairly standard pair of conclusions about them.

(30) a. External argument relative clauses, i.e., those with obligatory resumption, do not involve any movement. Instead, the pronoun is base-generated and bound by a null operator in Spec,CP.

[3]See McCloskey (To appear) for more on the isomorphism in form between regular and resumptive pronouns.

b. Internal argument relative clauses, i.e., those with obligatory gaps, are derived by A'-movement of a null operator to Spec,CP of the relative clause.

The hypotheses in (30) are not particularly noteworthy from the viewpoint of syntactic theory or typology, but they are important for our discussion here in confirming that the extraction condition of (27) regulates A'-movement and not a binding relation.

§4.2.1 begins by showing the A'-movement properties of internal argument relatives, while §4.2.2 shows that external argument relatives (i.e., those involving resumption) show no movement properties.

4.2.1 The A'-properties of internal argument relatives with gaps

4.2.1.1 Gaps are sensitive to islands

Movement, of course, should imply island-sensitivity (Ross, 1967). Here I document that non-resumptive relatives show such sensitivity.

I have identified three types of islands in SMP Mixtec. The first are what I refer to as so-that-clause islands, which are a variety of adjunct clause. In SMP Mixtec, these are not introduced by any distinct morphology. Instead, the verb obligatorily occurs in the irrealis.

(31) So-that clauses

a. Ndzyàà Santa Cruz nì xa'a tát é yó'o xí'in nà bálí [
towards S. C. PAST go father our.INCL this with they little.PL
xiki'i nà tsyàà nà].
buy.IRR they clothes their
'This man went to Santa Cruz with the children so that they could buy new clothes.'

b. *Veneno* tàhààn míí rà tsyahá kân ini tskwî [kùù ntsi'i míí
poison put.PAST.PL the he man that in water die.IRR all the
tsyaká].
fish

'That man put poison in the water so that all the fish would die.'

The gap inside of a relative clause may not occur within a so-that clause.

(32) So-that clauses are islands

a. * Míí tsyàà$_i$ [nì xa'a míí táté yó'o Santa Cruz xí'in nà bálí [xiki'i nà __$_i$]], ndibi kwê na' á.
the clothes PAST go the man this S. C. with they little buy.IRR they pretty very appear it
Intended: 'The clothes [that the man went to Santa Cruz with the children [so that they could buy __]] is very pretty.'

b. *? Nă'anu kwê'e míí tsyaká$_i$ [tàhààn míí rà tsyahá *veneno* ini tskwî [kùù __$_i$]].
big.PL very the fish put.PAST.PL the he man poison in water die.IRR
Intended: 'The fish [that the man put poison in the water [so that __ would die]] are very big.'

Consider a second variety of adjunct clause island - clauses introduced by *chi* 'because.'

(33) Because-clauses

a. Nì xa'a rà Mârco nù yá'bi [chi kôni ísi'i rà kôñù].
PAST go he M. where market because like.PRES wife his meat
'Marcos went to the market because his wife likes meat.'

b. Ntsí'i kwê'e ini rà Julio [chi nì xi'i míí chútun].
sad very inside he J. because PAST die the cat
'Julio is very sad because the cat died.'

Just like the so-that clauses in (32), because-clauses are also islands for relativization.

(34) Because-clauses are islands

a. * Ba'ă xáxi míí kôñù$_i$ [nì xa'a rà Mârco nù yá'bi [chi
good eat.PRES the meat PAST go he M. where market because
kôni ísi'i rà __$_i$]].
like.PRES wife his
Intended: 'The meat that Marcos went to the store because his wife likes tastes good.'

b. * Rí$_i$ yata xkuu míí chútun [ntsí'i kwê'e ini rà Julio [chi
it.AML old COP.PAST the cat sad very inside he J. because
nì xi'i __$_i$]].
PAST die
Intended: 'The cat that Julio is very sad because died was old.'

Embedded wh-questions constitute a final island type. Their form is illustrated in (35).

(35) Embedded wh-question

a. Xǐn ì [yó nà kàni ñá Maria].
know.NEG.PRES I who they hit.PAST she M.
'I don't know who hit Maria.'

b. Xǐn ì [nachúún nì xi'i tsǐna].
know.NEG.PRES I why PAST die dog
'I don't know why the dog died.'

There will be more to say about the syntax of such clauses in §4.3. For now, it is enough to observe that a relative clause gap may not occur within an embedded wh-question.

(36) Embedded wh-questions are islands

a. * Ñá$_i$ ntá'yi ñá lo'o$_i$ [xǐn ì [yó nà kàni __$_i$]].
she cry.PRES she little know.PRES.NEG I who they hit.PAST
Intended: 'The girl who I don't know who hit is crying.'

b. * Rí$_i$ yata xkuu tsǐna$_i$ [xǐn ì [nachúún nì xi'i __$_i$
it.AML old COP.PAST dog know.NEG.PRES I why PAST die
]].

Intended: 'The dog that I don't know why died was old.'

I take this island sensitivity as a strong indication that relative clauses with gaps are derived via movement.

4.2.1.2 Gaps induce Crossover

Since Postal (1971), Wasow (1972) and Chomsky (1981), it has been observed that Crossover effects provide a diagnostic for A'-movement. Crossover phenomena fall into two categories: Strong Crossover and Weak Crossover. Let us begin by examining Strong Crossover.

Strong Crossover arises when an A'-bound trace is also bound from an A-position within the domain of the moved phrase. It is therefore useful in distinguishing A-bound from A'-bound positions.

(37) a. * Who$_i$ does she$_i$ consider $\_\__i$ a genius?

b. * the woman$_i$ [who$_i$ she$_i$ considers $\_\__i$ a genius]

c. Ellen$_i$ seems to herself$_i$ $\_\__i$ to be a genius.

In (37a-b), we see that gap would be bound both by the wh-phrase 'who' in A'-position, as well as the subject 'she' in an A-position. This leads to ungrammaticality. In contrast, in (37c), the gap is bound only by the subject in A-position. As such, the sentence does not trigger Strong Crossover.

If gaps in SMP Mixtec relative clauses are derived by A'-movement like the English in (40b), then we expect the same Strong Crossover violation. This turns out to be the case.

(38) a. * Kŭú sihĭ ini míí rà tsyahá$_i$ [ká'a rà$_i$ kú'u
COP.PRES.NEG happy inside the he man think.PRES he sick.PRES
$\_\__i$].

Intended: 'The man who he$_i$ thinks $\_\__i$ is sick is not happy.'

b. * Ñá$_i$ ntá'yi míí ñá$_i$ lo'o [káchi ñá [xàxi'i'i míí tsǐna __$_i$]].
she cry.PRES the she little say.PAST she bite.PAST the dog
Intended: 'The girl who she$_i$ said the dog bit __$_i$ is crying.'

Now let us consider Weak Crossover. Consider the English sentences in (39).

(39) a. * Who$_j$ does his$_j$ mother love __$_j$?

b. *? the guy$_j$ [who$_j$ his$_j$ mother loves __$_j$]

c. Ellen$_{it}$ seems to her$_i$ friends __$_i$ to be stressed out.

In (39), we see that A'-bound traces may not enter into anaphoric relations with pronouns which are immediately commanded by the moved phrase. In contrast, in (39c), we see that A-bound traces may.

Here, too, relative clause gaps reveal themselves to be derived by way of A'-movement because they trigger Weak Crossover, just like the A'-movement processes in (39a-b).

(40) a. *? Ntsi'ì míí nà bálí$_i$ [ká'an nána nà$_i$ [kàni míí ñá
every the they little.PL think.PRES mother their hit.PAST the she
Maria t_i]] ntá'yi nà.
M. cry.PRES they
Intended: 'All the boys$_i$ that [their$_i$ mother thinks [Maria hit t_i]] is crying.'

b. *? Ntsi'ì míí rà bálí$_i$ [ká'an maestra rà$_i$ [íta t_i
all the he little.PL think.PRES teacher his$_i$ stand.PRES.PL
xihǐn be'e]] nǐ kixì ba'á rà.
beside house PAST.NEG sleep well he
Intended: 'All the boys$_i$ that [his$_i$ teacher thinks [t_i are standing beside the house]] didn't sleep well.'

In sum, then, we have seen that internal-argument relative clauses (i.e., those involving binding of a gap) demonstrate several A'-movement properties, such as sensi-

tivity to islands, Strong Crossover, and Weak Crossover. With this background, let us consider the properties of external-argument relative clauses.

4.2.2 The properties of external-arguments relatives with resumptive pronouns

We have already seen that relative clauses in which an external argument position is relativized forbid gaps and require resumptive pronouns. The crucial data is presented again in (41.)

(41) a. Rí$_i$ xà'ani ndú míí chú'u$_i$ [xàxi **rí**$_i$ nuhmà náhña].
it.AML kill.PAST we.EXCL the goat eat.PAST it.AML sprout chayote
'We killed the goat that (it) ate the chayote sprouts.'

b. * Rí$_i$ xà'ani ndú míí chú'u$_i$ [xàxi __$_i$ nuhmà náhña].

c. Nuhŭ mésa kanùù martíyo$_i$ [nàkààba **ñà**$_i$ sátá kitsì lo'o].
on table be.on.PRES.SG hammer fall.PAST it on animal little
'The hammer that (it) squished (fell on) the bug (the little animal) is on the table.'

d. * Nuhŭ mésa kanùù martíyo$_i$ [nàkààba __$_i$ sátá kitsì lo'o].

e. I'iní kwê'e xíni tát$_i$ é [kíchiñu **rà**$_i$ yúkŭ kân].
hot very feel.PRES father our.INCL work.PRES he forest that
'The man (respectfully: our father) who (he) works in the fields feels very hot.'

f. * I'iní kwê'e xíni tát$_i$ é [kíchiñu __$_i$ yúkŭ kân].

If we assume that such relatives involve binding of the pronoun rather than movement (a conclusion already suggested by the fact that such structures provide ways of circumventing a restriction on movement), then they should contrast with the filler-gap relatives in showing non-movement properties. This section shows that this expectation pans out.

4.2.2.1 Resumptive pronouns are island insensitive

The clearest piece of evidence that resumptive constructions in SMP Mixtec do not involve movement is that they may freely occur inside of islands. This is similar to other resumptive constructions cross-linguistically (see McCloskey To appear for an overview). Consider the same set of islands that we investigated in §4.2.1.1. First are so-that clause islands in (42-44). Recall that these adjunct clauses are not introduced by any overt morpheme, but rather can be spotted because of the obligatory irrealis marking on the predicate. The second are because-clause islands in (45-46), and the third are wh-islands in (47-48).

(42) Transitive subject resumptive in so-that clause island

a. Kíchûn ntsya'a rà Juan ná kuxi ísi'i rà kǒñù ntsikihi.
work.PRES hard he J. SUBJ eat.IRR wife his meat everyday
'Juan works hard so that his wife can eat meat every day.'

b. Kwátsyá ntsya'a ñá sí'i$_i$ [kíchûn rà Juan [ná kuxi **ñá**$_i$ kǒñù ntsikihi]].
happy very she woman work.PRES he J. SUBJ eat.IRR she meat everyday
'The woman [who Juan works [so that **she** can eat meat everyday]] is very happy.'

(43) Unergative subject resumptive in so-that clause island

a. Nì'i ñá maestra libro kìxàà ñá [sakwa'a nà bálí].
carry.PAST she teacher book come.PAST she study.IRR they little.PL
'The teacher brought (came carrying) book so that the children would study.'

b. Kúú sihi̋ ini nà bálí$_i$ [nì'i ñá maestra libro kìxàà ñá [sakwa'a **nà**$_i$]].
COP.PRES happy in they little.PL carry.PAST she teacher book come.PAST she study.IRR they
'The children [that the teacher brought books [so that **they** could study]] are very happy.'

(44) Unergative subject resumptive in so-that clause island

a. Xìta ñá iin yâà nuhŭ se'e lo'o ñá [ná kusi yá].
sing.PAST she a song to child little her SUBJ sleep.IRR it
'The woman sang a song to her baby so that it would fall asleep.'

b. Míí yá lo'o$_i$ [xìta ñá míí yâà [ná kusi yá$_i$]], Mârco
the it little sing.PAST she the song SUBJ sleep.IRR it M.
náni rà.
called.PRES he
'The baby$_i$ [who the woman sang the song [so that **it** would fall asleep]],
he is called Marcos.'

(45) Transitive subject resumptive in because-clause island

a. Ntsí'i kwê'e ini ñá maestra [chi kàni rà Julio ñá Maria].
sad very in she teacher because hit.PAST he J. she M.
'The teacher is very sad because Julio hit Maria.'

b. Rà$_i$ xáku míí rà lo'o$_i$ [ntsí'i kwê'e ini ñá maestra [chi
he laugh.PRES the he little sad very in she teacher because
kàni **rà**$_i$ ñá Maria]].
hit.PAST he she M.
'The boy [who the teacher is very sad [because **he** hit Maria]] is laughing.'

(46) Unergative subject resumptive in because-clause island

a. Kwátsyá kwê'e ñá Maria [chi kíchûn rà Juan bítsí].
happy very she M. because work.PRES he J. now
'Maria is very happy because Juan is working now.'

b. Míí rà tsyahá$_i$ [kwátsyá kwê'e ñá Maria [chi kíchûn **rà**$_i$]],
the he man happy very she M. because work.PRES he
Juan náni rà.
J. called.PRES he
'The man [who Maria is happy [because **he** is working]] is named Juan.'

(47) Transive subject resumptive in wh-island

a. Xĭn ì yó kàni rà Sergio.
know.PRES.NEG I who hit.PAST he S.
'I don't know who Sergio hit.'

b. Rà$_i$ xáku míí rà lo'o$_i$ [xǐn ì [yó kàni **rà**$_i$]].
he laugh.PRES the he little know.PRES.NEG I who hit.PAST he
'The boy who I don't know who he hit is laughing.'

(48) Unergative subject resumptive in wh-island

a. Kúndàà in ì [ntsyáchi kíchûn míí Juan].
wonder.PRES in I where work.PRES the J.
'I wonder where Juan is working.'

b. Juan náni míí rà tsyahá [kúndàà in ì [ntsyáchi kíchûn **rà**$_i$]].
J. called.PRES the he man wonder.PRES in I where work.PRES he
'The man who I wonder where he is working is named Juan.'

(42-48) again support the conclusion that resumptive pronoun relatives do not involve movement. It is important to bear in mind in this context that resumption is not a general strategy to rescue island violations. This is because, as we have already established, relativization of internal arguments is incompatible with resumption, requiring a gap instead. (49) is repeated from (18). See §4.2.1 as well.

(49) Intransitive with obligatory gap in relative clause

a. Nì ka' ìn xí'in míí rà lo'o$_i$ [ínchichi __$_i$ xíhǐn be'e].
PAST talk I with the he little stand.PRES.SG beside house
'I talked with the boy that is standing next to the house.'

b. * Nì ka' ìn xí'in míí rà lo'o$_i$ [ínchichi **rà**$_i$ xíhǐn be'e].
PAST talk I with the he little stand.PRES.SG he beside house
Intended: 'I talked with the boy that (he) is standing next to the house.'

c. Tún$_i$ tààba ndú míí ítǔn [nàkààba __$_i$ ichǐ yo'o].
it.WOOD remove.PAST we.EXCL the tree fall.PAST path this
'We removed the tree that fell in the road.'

d. * Tún$_i$ tààba ndú míí ítǔn [nàkààba **tún**$_i$ ichǐ yo'o].
it.WOOD remove.PAST we.EXCL the tree fall.PAST it.WOOD path this

Intended: ‘We removed the tree that (it) fell in the road.’

4.2.2.2 Resumptive pronouns do not induce Weak Crossover

Relative clauses involving resumption also show no sensitivity to Weak Crossover.

(50) No WCO with resumptive pronoun in transitive subject position

a. Ñá sí’i$_i$ [ká’an ñá Maria [ndanda’ă ñá$_i$ xí’in rà Juân]], míí
she woman think.PRES she M. marry.IRR she with he J. the
yá ndàà yá kôni ñá rà Mârco.
it truth it love.PRES she he M.
‘The woman$_i$ [that Maria thinks [(she$_i$) will marry Juan]] actually loves Marcos.’

b. Ñá sí’i$_i$ [ká’an táta ñá$_i$ [ndanda’ă ñá$_i$ xí’in rà Juân]], míí
she woman think.PRES father her marry.IRR she with he J. the
yá ndàà yá kôni ñá rà Mârco.
it truth it love.PRES she he M.
‘The woman$_i$ [that her$_i$ father thinks [(she$_i$) will marry Juan]] actually loves Marcos.’

(51) No WCO with resumptive pronoun in unergative subject position

a. Ntsi’ì míí ná bálí$_i$ [ká’an rà maestro [ba’á táxa’a
all the they.FEM little.PL think.PRES he teacher well dance.PRES
ná$_i$]] kúú sihĭ ini ná.
they.FEM$_i$ COP.PRES happy inside they.FEM
‘All the girls$_i$ [that the teacher thinks [(they$_i$) dance well]] are happy.’

b. Ntsi’ì míí ná bálí$_i$ [ká’an táta ná$_i$ [ba’á
all the they.FEM little.PL think.PRES father their.FEM well
táxa’a ná$_i$]] kúú sihĭ ini ná.
dance.PRES they.FEM$_i$ COP.PRES happy inside they.FEM
‘All the girls$_i$ [that their$_i$ father thinks [(they$_i$) dance well]] are happy.’

(52) No WCO with resumptive pronoun in unergative subject position

a. Rà tsyahá$_i$ [ká'an míí Maria [kìchiñu ba'á rà$_i$ míí kwîya yó'o]]
he man think.PRES the M. work.PAST well he the year this
xìi rà iin be'e.
buy.PAST he a house
'The man$_i$ [that Maria thinks [(he$_i$) worked hard this year]] bought a house.'

b. Rà tsyahá$_i$ [ká'an nána rà$_i$ [kìchiñu ba'á rà$_i$ míí kwîya yó'o]]
he man think.PRES mother his work.PAST well he the year this
xìi rà iin be'e.
buy.PAST he a house
'The man$_i$ [that his$_i$ mother thinks [(he$_i$) worked hard this year]] bought a house.'

I take this as further indication that movement is not involved in these resumptive constructions.[4]

4.2.3 Interim summary: The syntax of relative clauses

In this section, we have seen two patterns.

(53) a. The availability of A'-movement in relative clauses depends on which argument is relativized.

i. If an internal argument is relativized, then the language uses A'-movement.

[4]Strong Crossover is much more difficult to test because in Irish, Hebrew, and Palestinian Arabic (Shlonsky 1992) and Lebanese Arabic (Aoun et al. 2001), resumption induces Strong Crossover. See McCloskey (1990) and Shlonsky (1992) for extensive discussion of this effect.

Interestingly, SMP Mixtec upholds the cross-linguistic trend of resumptive pronouns triggering Strong Crossover. In these examples, I employ the most common naturally-occurring epithet, which is formed by modifying a pronoun with *kân* 'that,' yielding an interpretation similar to 'the guy' or 'the gal.'

(1) a. * Yo'o míí tát$_i$ é [ká'an rà$_i$ kân [ba'á kíchiñu rà$_i$]].
the the father our.INCL think.PRES he that well work.PRES he
Intended: 'That's the man$_i$ that the guy$_i$ thinks he$_i$ works well.'

b. * Ñá$_i$ ntá'yi míí ñá$_i$ lo'o [káchi ñá kân [tà'abi ñá$_i$ yùtátá]].
she cry.PRES the she little say.PAST she that break.PAST she mirror
Intended: 'The little girl$_i$ who the gal$_i$ said she$_i$ broke the mirror is crying.'

That said, this Strong Crossover effect is not useful in diagnosing movement, or the lack thereof.

ii. If an external argument is relativized, then A'-movement is no longer available.

b. External argument relatives require a resumptive construction, which I argued above does not involve any movement, much less A'-movement.

With this in mind, let us recall the A'-extraction restriction introduces in (27) and repeated in (54).

(54) **A'-extraction restriction in SMP Mixtec**: External arguments may not undergo A'-extraction.

(54), then, seems to provide us with what we have been seeking: a reliable, language-internal diagnostic for identifying A'-movement. To demonstrate that the restriction is general and applies to more A'-movement constructions than just relativization, we need to broaden the discussion. We do this in the following sections by examining constituent questions.

4.3 Constituent Questions

Just as the syntactic profile shown by relative clauses depends on which argument is relativized, so too do we see a split in constituent questions. The split is, again, dependent on whether the relevant gap is interpreted as an external or as an internal argument. Once again, we see the effects of the ban on extraction of external arguments. However, the repair used to circumvent the restriction in this case is different.

For internal arguments, we see a construction that looks much like English, demonstrated in (55).

(55) a. Nă$_i$ xì'i míí tsǐna __$_i$?
what drink.PAST the dog
'What did the dog drink?'

b. Yó$_i$ kàni rà Julio __?
who hit.PAST he J.
'Who did Julio hit?'

c. Yó$_i$ íchichi __$_i$ nuhŭ barco?
who stand.PRES.SG facing boat
'Who is standing facing the boat?'

d. Ntsyâ míí ná sí'i$_i$ kú'u __$_i$.
which the they.FEM woman sick.PRES
'Which of the women are sick?'

In (55), we see that wh-phrases that correspond to internal arguments front to the beginning of the clause, very much like English. I will show below that the syntax of these constructions shows further parallels to English, and is derived through A'-movement of the wh-phrase to Spec,CP.

In contrast, and as in relative clauses, gaps may not occupy positions which correspond to external arguments. This is demonstrated by the ungrammaticality of (56). In particular, compare (56a) to (55b).

(56) a. * Yó$_i$ kàni __ rà Julio?
who hit.PAST he J.
Intended: 'Who hit Julio?'

b. * Yó$_i$ káchûn __$_i$?
who work.PRES
Intended: 'Who is working?'

c. * Ntsyâ míí nà tsyahá yó'o ba'á kíchiñu __?
what the they man this well work.PRES
Intended: 'Which of these men is working well?'

In (56a) and (56b) we see that gaps are impossible in the subject positions of unergative verbs and of transitive verbs respectively. For this reason, questions like (55b) are unambiguous, unlike in some other VSO languages (McCloskey, 1977). Additionally,

observe by the ungrammaticality of (56c) that D-linking the wh-phrase does not improve this extraction (see Pesetsky 1987).[5]

So far we see a pattern exactly analogous to the one we observed for relative clauses. However, when we observe the grammatical versions of (56) we see that they involve a different construction containing two pieces, which I refer to respectively as the cleft and the pseudo-relative.

(57) a. [Yó (kúú) nà] [kàni __ rà Julio]?
who COP.PRES they hit.PAST he J.
'Who are they that hit Julio?'

b. [Yó (kúú) nà] [káchûn __]?
who COP.PRES they work.PRES
'Who are they that are working?'

c. [Ntsyâ nà (kúú) míí nà tsyahá yó'o] [ba'á kíchiñu __]?
which they COP.PRES the they man this well work.PRES
'Which of them are these men that are working well?'

In (57), we see that to form an interrogative dependency on an external argument,

[5]Occassionally, speakers judge unergative wh-questions like (56b) as grammatical, although with a strongly different interpretation. Consider the contrast in (i).

(1) a. Yó$_i$ káxă __$_i$?
who snore.PRES
'Who snores?'

b. [Yó kúú nà] [káxă __]?
who COP.PRES they snore.PRES
'Who is snoring?'

In (ia), we see that wh-movement of an unergative subject is possible if it induces a thetic, quasi-generic judgment. The speaker who provided this judgment intuited that (ia) would be grammatical in a context in which we are doctors doing demographic research on snorers. In this way, *yó* 'who' would not refer to a specific individual. Rather, the meaning of (ia) could be paraphrased as 'Who are such that it is the case that they snore?' In contrast, (ib) induces a strongly categorical judgment. This question would be most appropriate if we can hear someone snoring, and you want to know the identity of the snorer.

While these semantic effects are interesting, I will not discuss them further. This is for two reasons. First, I do not have a clear semantic generalization for the sorts of contexts in which (ia) would be licit or for the set of unergative predicates that can support this interpretation. This is largely because I do not have enough tokens of sentences like (ia) being judged grammatical. Second, this interpretation is not available for transitive subjects.

two pieces are necessary. The first is the cleft, which is the first bracketed portion of each sentence in (57). The cleft consists of the interrogative phrase, a copula, and a "subject," which limits the domain of the interrogative phrase.

The second piece is what I will call the pseudo-relative - the second bracketed portion of the sentences in (57). Superficially, pseudo-relatives look like relative clauses, but we can be sure that these are not true relatives in SMP Mixtec, because if they were, a resumptive should be required, as we saw in §4.2. In contrast, resumptives in pseudo-relatives are prohibited, as shown in (58).

(58) a. * [Yó (kúú) nà] [kàni **nà** rà Julio]?
who COP.PRES they hit.PAST they he J.
Intended 'Who are they that (they) hit Julio?'

b. * [Yó (kúú) nà] [káchûn **nà**]?
who COP.PRES they work.PRES they
Intended: 'Who are they that (they) are working?'

c. * [Ntsyâ nà (kúú) míí nà tsyahá yó'o] [ba'á kíchiñu **nà**]?
which they COP.PRES the they man this well work.PRES they
Intended: 'Which of them are these men that (they) are working well?'

There is much work still to be done in investigating the complexities of this syntactic pattern. I will not attempt to do that work here, since it is not centrally relevant for my principal concern, which is to show that the external argument restriction also holds of interrogative wh-movement, and is therefore a reliable diagnostic for A'-movement in this language. My primary goal for now, then, will be to demonstrate that internal argument constituent questions such as those in (55) show the expected properties of A'-movement.

Questions then arise about the apparent subject gap in the pseudo-relative. If this gap is derived by movement, like the gap in internal argument relative clauses in §4.2, is the relevant movement A- or A'-movement? The answers to all of these questions remain unclear, but in Appendix A I gather a body of evidence which suggest that the

subject gap, whatever its exact character, does not show the signature properties of A'-movement, either the general properties or the language-internal properties. If this is correct, the question of its interaction with the external argument restriction is moot.

4.3.1 The properties of internal-argument constituent questions

Internal argument constituent questions were initially demonstrated in (55), and (59) provides further examples.

(59) a. Nă chì'i Julia __?
what sow.PAST J.
'What did Julia sow?'

b. Nă xá'anu __ yé'è?
what grow.PRES garden
'What is growing in the garden?'

c. Ntsyâ míí rí kitsì ntá'yi __?
which the it.AML animal cry.PRES
'Which of the animals is crying?'

In (59) we see that internal argument extraction generally involves a gap in object position, with the wh-phrase fronted to the left edge of the clause. Before delving into the more detailed syntax of constructions like (80), we can observe several properties which are consistent with wh-movement in English. First, like English, wh-movement is obligatory.

(60) Wh-movement is obligatory

a. * Chì'i Julia **nă**?
sow.PAST J. what
Intended: 'What did Julia sow?'

b. * Xá'anu **nă** yé'è?
grow.PRES what garden
Intended: 'What is growing in the garden?'

c. * Ntá'yi **ntsyâ míí rí** **kitsì**?
cry.PRES which the it.AML animal
Intended: 'Which of the animals are crying?'

Second, wh-movement is unbounded.

(61) Wh-movement is unbounded

a. Nă$_i$ ká' ún [kôni nà [kuxi nà __$_i$ itsyààn]]?
what think.PRES you want.PRES they eat.IRR they tomorrow
'What do you think they will want to eat tomorrow?'

b. Yó$_i$ ká'an ñá Maria [índu'u __$_i$ xihi̵n be'e]?
who say.PAST she M. sit.PRES beside house
'Who did Maria say is sitting beside the house?'

In (61), we see that wh-movement may cross clause boundaries, indicated that this process is unbounded.

4.3.1.1 Internal-argument wh-constructions are island sensitive

Internal argument wh-questions are island sensitive. This is demonstrated in (62-64) for the three islands we have been examining throughout: so-that clause islands, because-clause islands and wh-islands.

(62) So-that clause island

a. Kíchûn ntsya'a rà Juan [ná kuxi ísi'i rà kŏñù ntsikihi].
work.PRES hard he J. SUBJ eat.IRR wife his meat everyday
'Juan works hard so that his wife can eat meat every day.'

b. * Nă$_i$ kíchûn ntsya'a rà Juan [ná kuxi ísi'i rà __$_i$]?
what work.PRES hard he J. SUBJ eat.IRR wife his
Intended: 'What does Juan work hard so his wife can eat __?' (so-that clause island)

(63) because-clause island

a. Ntsí'i kwê'e ini rà Julio [chi nì xi'i chútun sâna rà].
sad very inside he J. because PAST die cat POSS his
'Julio is very sad because his cat died.'

b. * Yó$_i$ ntsí'i kwê'e ini rà Julio [chi nì xi'i __$_i$]?
who sad very in he J. because PAST die
Intended: 'Who is Julio very sad because __ died?' (because-clause island)

(64) Wh-island

a. Xǐn ì [náchûn xìxi míí Diego xità bìhkŏ].
know.NEG.PRES I why eat.PAST the D. tortilla party
'I don't know why Diego ate tortillas at the party.'

b. * Nă$_i$ xǐñ ú [náchûn xìxi míí Diego __ bìhkŏ]?
what know.NEG.PRES you what eat.PAST the D. party
Intended: 'What do you not know why Diego ate __ at the party?' (Embedded wh-question island)

Island sensitivity, as in the case of relative clauses, provides strong evidence that movement is involved in deriving internal argument wh-questions.

4.3.1.2 Wh-movement induces Crossover

If wh-movement is a form of A'-movement, we expect to find both Strong and Weak Crossover effects in this construction. We do, as shown in (65) and (66) respectively.

(65) Wh-movement shows Strong Crossover

a. Ká'an nà$_i$ [kú'u nà$_i$].
think.PRES they sick.PRES they
'They think that they are sick.'

b. * Yó$_i$ ká'an ná$_i$ [kú'u __]?
who think.PRES they sick.PRES
Intended: 'Who$_i$ do they$_i$ think is sick?'

(66) Wh-movement shows Weak Crossover

a. * Yó$_i$ kôni nána nà$_i$ __$_i$?
who love.PRES mother their
Intended: 'Who$_i$ does his$_i$ mother love __$_i$?'

b. * Yó$_i$ ká'an nána nà$_i$ [kàni rà Julio __$_i$]?
who think.PRES mother their hit.PAST he J.
Intended: 'Who$_i$ does their$_i$ mother think Julio hit __$_i$?'

c. * Yó$_i$ ká'an nána nà$_i$ [nàkaaba __$_i$]?
who think.PRES mother their fall.PAST
Intended: 'Who$_i$ does their$_i$ mother think __$_i$ fell?'

In (65) we see that a pronominal subject in the matrix clause can easily be coreferent with the unaccusative subject of *kú'u* 'be sick (PRES)' in the embedded clause. But if the embedded subject is an interrogative phrase that undergoes wh-movement to the matrix clause, the two may no longer enter into an anaphoric relation. This is a Strong Crossover effect, and a good indicator of A'-movement.

Likewise, in (66) we see a classical Weak Crossover pattern in which a wh-phrase may not be coreferent with a pronoun that doe snot c-command its base position. The presence of the Weak Crossover effect is further evidence that wh-movement in this language is A'-movement.

4.3.2 Wrapping up wh-constructions

In this section, we have seen that wh-movement of internal arguments is, certainly and unsurprisingly, A'-movement. Crucially, as we saw in §4.3, this A'-movement strategy is not available to external arguments. As such, when we are sure that we are dealing with A'-movement in SMP Mixtec, we observe that it is subject to the external argument extraction restriction. That being so, we can use the restriction as a probe for the role of A'-movement in a given syntactic configuration. We now return to our central theme, the nature of pronoun doubling, with this useful result in hand.

4.4 Returning to pronoun doubling

Throughout all of the complexities of this chapter, one connection emerges with considerable clarity: that between A'-movement and the external argument restriction, as restated in (67).

(67) **A'-extraction restriction in SMP Mixtec**: External arguments may not undergo A'-extraction.

On that basis, we are left with a simple prediction. If pronoun doubling is derived with a step of A'-movement, then external arguments should be prohibited from being the targets of pronoun doubling. As is already abundantly clear, this prediction is false. Pronoun doubling readily, and often preferentially as we will see in §5, targets external arguments. Some of the relevant data are replicated in (68).

As we have seen dozens of times throughout this dissertation, the prediction in (150) is categorically and robustly false. Pronoun doubling readily, and often preferentially as we will see in §5, targets external arguments. A subset of these data are replicated in (151).

(68) Pronoun doubling may target external arguments

a. **Rà**$_i$ xa kàni rà Julio$_i$ ñá Maria!
he PERF hit.PAST he J. she M.
'Julio already hit Maria!'

b. **Rí**$_i$ xa nì xă'nu míí chútun$_i$ míí *báso.*
it.AML PERF PAST break the cat the cup
'The cat already broke the cup.'

c. **Ñá**$_i$ xa xà'antsya míí Juana$_i$ ítŭn.
she PERF cut.PAST the J. tree
'Juana already cut down the tree.'

d. **Nà**$_i$ sákwa'a nà bálí$_i$.
they study.PRES they little.PL

'The children are studying.'

e. **Rà** káxă rà lá'nu$_i$.
he snore.PRES he old.man
'The old man is snoring.'

f. **Rí**$_i$ kárakono míí kwáyi$_i$.
it.AML run.PRES the horse
'The horse is running.'

This simple observation strongly suggests that we cannot derive pronoun doubling with a Move-and-Reduce analysis of clitic doubling in whidh the prerequisite movement is A'-movement. This, coupled with the discussion from §3 that pronoun doubling cannot be analyzed in terms of A-movement, leaves us with one available avenue of analysis: φ-agreement.

Chapter 5

Pronoun doubling as topic agreement

From the start of this dissertation, we observed that pronoun doubling in SMP Mixtec is a process that doubled the φ-features of an argument with a preverbal morpheme that is morphophonologically identical to a regular pronoun. As such, we considered the two (and a half) main possibilities in the literature for deriving this sort of replication of φ-features. These options are repeated in (1).

(1) Multiple exponence of φ-feature bundles can be derived by:

  a. Providing a value to some node's unvalued φ-features (φ-agreement).

  b. Spell-out of multiple copies in a movement chain (clitic doubling).

    i. This movement chain could be formed by A-movement.

    ii. This movement chain could be formed by A'-movement.

In chapters §3 and §4, we saw several arguments against each of the options in (1b). With respect to A-movement, chapter 3 demonstrated that pronoun doubling has no effect on anaphoric relations, neither expanding binding possibilities nor showing reconstruction effects. In this way, pronoun doubling contrasts with A-movement in SMP Mixtec (as represented by quantifier fronting), with well-studied A-movements in other languages, and with clitic doubling in many languages, such as Rioplatense

Spanish (Suñer 1988), Modern Greek (Alexiadou and Anagnostopoulou 1997), Bulgarian (Harizanov 2014a,b), and Amharic (Kramer 2014). In addition, we have seen that SMP Mixtec, like Amharic, seems to forbid structures in which pronouns precede their antecedents (i.e., cataphora). This is illustrated again in (2).

(2) a. * **Rà**$_j$ xa nàni'i [míí nána rà$_j$]$_i$ míí Julio$_j$.
he PERF find.PAST the mother his the J.
Intended: 'His$_j$ mother already found Julio$_j$.'

b. **Tún**$_j$ xa ntàsaba'á [xito'o$_i$ [kárru$_j$]] tún$_j$.
it.WOOD PERF repair.PAST owner car it.WOOD
'The owner of the car has already repaired it.'

Unlike clitic doubling in Amharic, however, pronoun doubling in SMP Mixtec cannot repair this kind of violation, whatever its origin, as we see again in (2a). Likewise, in (2b) we see that pronoun doubling fails to induce a Condition C violation. Under the assumption that this effect on binding is a core diagnostic for A-chains, and by extension clitic doubling (see Preminger 2009), I take the lack of an effect on binding as evidence that no A-chain is involved.

At the same time, there is no a priori requirement that the movement chain in clitic doubling be an A-chain (see Poletto and Pollock 2004). In §4, we saw that no evidence for A'-movement could be found either. First, let us consider (2b) more closely. If this sentence were derived by A'-movement, we would expect that it should produce a Weak Crossover violation, just as A'-movement in this language usually does. This prediction is not met: (2b) is fully grammatical and natural, indicating no Weak Crossover occurs. This sheds initial doubt that pronoun doubling could be derived through A'-movement. At the same time, there is a tight connection between pronoun doubling and topicalization, as we have seen throughout and as we shall see in much closer detail shortly. As Lasnik and Stowell (1991) observe, Topicalization in English systematically fails to induce Weak Crossover. Therefore, pronoun doubling might not

necessarily induce Weak Crossover.

Stronger evidence that no A'-movement is involved in these constructions comes from the observation that external arguments in this language never undergo A'-extraction with a gap. Instead, alternative constructions must be used, yielding a pattern that is rather reminiscent of syntactic ergativity (see Dixon 1979, Deal 2015b, and Polinsky 2016 for overviews). This pattern is briefly sketched in (3), though see §4 for a much more detailed discussion.

(3) a. Yó$_i$ kàni ñá Maria $\_\_{}_i$?
who hit.PAST she M.
'Who did Maria hit?'

b. * Yó$_i$ kàni $\_\_{}_i$ ñá Maria?
who hit.PAST she M.
Intended: 'Who hit Maria?'

c. Rí$_i$ xàxi ndú míí tsyuhŭ$_i$ [xà'ani míí tsĭna $\_\_{}_i$].
it.AML eat.PAST we the turkey.hen kill.PAST the dog
'We ate the turkey hen that the dog killed.'

d. * Ñá$_i$ ntá'yi míí ñá lo'o$_i$ [tà'abi $\_\_{}_i$ yùtátá].
she cry.PRES the she little break.PAST mirror
Intended: 'The girl who broke the mirror is crying.'

In (3a) and (3c), we see that internal arguments grammatically undergo A'-extraction with a gap, specifically wh-movement and relativization respectively. Contrast these grammatical sentences with those in (3b) and (3d), which demonstrate that external arguments may not undergo the same A'-extraction as internal arguments can. Even stronger, we saw in §4 that external arguments may not undergo A'-extraction at all.

Importantly for our purposes, if pronoun doubling were derived through A'-movement, then we expect to see the effect of this same A'-extraction restriction. Specifically, we predict that it should be ungrammatical to pronoun double an external argument. This turns out to be, emphatically, not the case. External argument doubling is

extremely common in this language, as (4) demonstrates.

(4) a. **Rà**$_i$ kàni rà Julio ñá Maria.
he hit.PAST he J. she M.
'Julio hit Maria.'

b. **Ñá**$_i$ tà'abi ñá Juana$_i$ yùtátá.
she break.PAST she J. mirror
'Juana broke the mirror.'

With this, it becomes clear that attempting to derive pronoun doubling through clitic doubling would be an unintuitive task at best, and entirely incompatible with the data at worst. Therefore, let us consider the alternative from (1): pronoun doubling is derived by φ-agreement.

In this chapter, I will show that a φ-agreement account is indeed the best fit for these data. That said, pronoun doubling demonstrates a slew of properties that are typologically rare, if not entirely unattested in φ-agreement systems. For instance, pronoun doubling may target subjects or objects, as we have seen throughout.[1]

[1]Svenonius (p.c.) suggests that examples like (5b) in which agreement targets the object may involve some sort of syntactic passivization. This matters because it would mean that pronoun doubling would always target subjects.

SMP Mixtec does not have a true passive, as discussed briefly in §3. This language does have a method to decrease transitivity and reduce the valency of a predicate by one: noun incorporation. The most common kind of incorporation that all speakers accept and produce is object incorporation, demonstrated in (i).

(1) a. Ndzyáa ì lánchi.
watch.PRES I sheep
'I am watching the sheep.'

b. Yù'u ndzyáa lánchi -k ì.
I watch.PRES sheep -INTR I
'I am sheep-watching.'

In addition to object incorporation, some speakers seem to allow agent incorporation, although the syntax of this is far from clear. This is shown in (ii).

(2) a. Rà xá'antsya rà Juan naranja.
he cut.PRES he J. orange
'Juan is cutting oranges.'

b. Rà xá'antsya naranja kúú rà Juan.
he cut.PRES orange INTR he J.
'Juan is orange-cutting.'

(5) a. **Rà**$_i$ xá'antsya rà Juan$_i$ chìkí.
he cut.PRES he J. tuna
'Juan is cutting *tunas*.'

b. **Rí**$_i$ xá'antsya rà Juan chìkí$_i$.
it.AML cut.PRES he J. tuna
'Juan is cutting *tunas*.'

In (5a), we see that the subject undergoes pronoun doubling, while in (5b), the object is targeted. Speakers are emphatic and clear in their judgment that the two sentences in (5) are truth-conditionally equivalent and refer to the same event. Rather, the usual judgment on the distinction between (5a) and (5b) is that the choice between them *depende en la conversación* 'depends on the conversation.' Below, I will argue that the difference between them can be summarized as in (6).

(6) φ-agreement in SMP Mixtec targets topics.

The goal of this chapter is to demonstrate that (6) is the best characterization of these data and to develop a theory of topic agreement.

Naturally, the idea of topic agreement has two necessary components. First, it must be demonstrated that pronoun doubling is a form a φ-agreement. In §5.1 I will review several arguments that support analyzing pronoun doubling as φ-agreement, beyond the negative evidence presented above. Second, it must be shown that this φ-agreement process targets topics. §5.2 reviews the relevant literature on topics and demonstrates that topichood indeed fits the distribution of pronoun doubling.

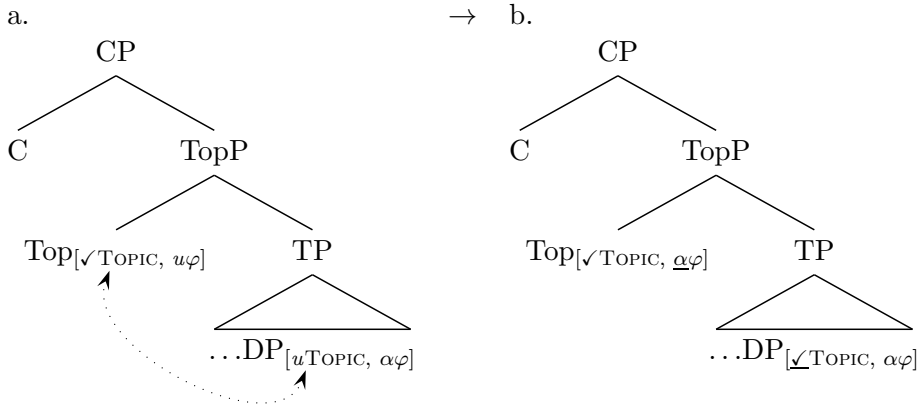
Once it has been established that (6) is the right approach to pronoun doubling, §5.3 develops a theory that allows a probe in a Chomskyian AGREE system to target

c. Rí$_i$ xá'antsya rà Juan kúú naranja$_i$.
it.AML cut.PRES he J. INTR(?) orange
Perhaps: 'The oranges are being Juan-cut.'

Setting aside the exact syntax of (iic), which is far from clear at this time, none of these valency changing processes are present in the data throughout, aside from a few isolated cases of object incorporation that are clearly labeled. Therefore, it does not seem to be possible to uphold the possibility that when pronoun doubling targets objects, any kind of covert passivization occurs.

topics. The core intuition behind this system is that topic agreement in this language acts similarly to a traditional abstract CASE system (see Chomsky and Lasnik 1993 and Chomsky 1995, as well as Miyagawa 2010, 2017), but this language is typologically head-marking rather than dependent marking (Nichols, 1986). The basic system is presented in (7).

(7)



In (7a), we see that in this language, the Topic head, similar to that proposed by Rizzi (1997), is a φ-probe. As such, it will search its c-command domain to find a DP that it can enter into an AGREE relation with. At the same time, I propose that the DP which will be interpreted as a topic enters the derivation with an unvalued [uTOPIC] feature. The idea here is that DPs in this language must stand in an AGREE relation with Top in order to be successfully interpreted as a topic at the semanticopragmatic interface. The topic-to-be DP also has a set of φ-features, naturally. Therefore, the unvalued features of both Top and the topic DP can be valued if they enter into an AGREE relation. As shown in (7b), this is precisely what I propose happens. It is this valued set of φ-features on Top that is ultimately Spelled-out as the doubled pronoun at the morphophonological interface.

Once this system is developed in §5.3, §5.4 presents some interesting consequences and empirical generalizations that emerge from analyzing pronoun doubling as

topic agreement, and concludes.

5.1 Pronoun doubling is φ-agreement

In §3 and §4, we saw a variety of negative evidence that topic agreement cannot be derived by a Move-and-Reduce analysis of clitic doubling. Let us review some positive evidence that pronoun doubling is a form of φ-agreement. To do this, recall the core diagnostics for φ-agreement from §2.1, repeated in (8).

(8)

	φ-agreement	Pronoun doubling
Obligatorily target highest nominal?	✓	**x**
Obeys the PIC?	✓	✓
Sensitivity to specificity or referentiality?	**x**	✓
Can affect variable binding?	**x**	**x**
One per clause?	✓	✓
Does coordination block it?	**x**	**x**

In (8), we see that the majority of diagnostics that clearly distinguish φ-agreement[2] indicate that pronoun doubling is a form of φ-agreement. In this section, we will discuss each in turn. §5.1.1 shows that pronoun doubling obeys the PIC. §5.1.2 demonstrates that pronoun doubling does not affect variable binding, as we have already seen in (2). §5.1.3 discusses the diagnostic of Baker (2012) that φ-agreement is compatible with only one marker per clause. §5.1.4 discusses coordination, and demonstrates that SMP Mixtec allows for Closest Conjunct Coordination, which appears to be a key feature of φ-agreement constructions.[3] §5.1.5 discusses the two features of pronoun doubling that do not fit in neatly with the existing typology of φ-agreement constructions: a

[2]I leave out the diagnostic of Preminger (2009) from (7) because, as discussed in §2.1.7, if no default occurs, than the process is compatible with either φ-agreement or clitic doubling. No default occurs in pronoun doubling constructions, and as such, we cannot reach any conclusions.

[3]That said, see the discussion of Mexican Spanish in §2.2.5.2, as well as Bošković (2018) for similar reported data from Peruvian Spanish. Note that the speakers of Peruvian Spanish I consulted did not confirm the data reported in Bošković (2018).

sensitivity to referentiality, and that both subjects and objects may be targeted, as shown in (5).

5.1.1 Pronoun doubling obeys the PIC

Recall from §2, particularly sections §2.1.2 and §2.2.3, that a key feature of φ-agreement is that it may never cross phase boundaries due to the Phase Impenetrability Condition (PIC) of Chomsky (2001). In this section, I will show that pronoun doubling also obeys the PIC, supporting its identification with φ-agreement. Note that most of this evidence was already presented in §3, where the context was such that these data were problematic. In the present context, they fall neatly into place. I take this as a welcome result.

First we must identify the phasal categories in this language. I do this in §5.1.1.1 with an in-depth discussion of pied-piping with inversion. Once we have determined that the phasal categories in this language are C, D, and P[4], §5.1.1.2 reviews the data from §3 that pronoun doubling may never cross these boundaries.

5.1.1.1 The phasal categories in SMP Mixtec: The view from pied-piping with inversion

In order to determine what phasal categories of SMP Mixtec, let us consider the distribution of a phenomenon referred to as pied-piping with inversion, discussed briefly in §3. First reported in Smith-Stark (1988) and investigated further in Black (1994), Eberhardt (1999), and Broadwell et al. (2006) in Oto-Manguean, and notably in Aissen (1996) and Coon (2009) for Mayan languages, pied-piping with inversion is defined in (9).

(9) **Pied-piping with inversion**: The phenomenon where [+WH] constituents obligatorily appear leftmost in their phrase when corresponding [-WH] constituents

[4]I remain agnostic about v because I do not have the tools at my disposal to investigate it.

do not.

In pied-piping with inversion constructions, a wh-phrase that is embedded within a larger phrase obligatorily appears at the left edge of that containing phrase. Consider the pattern demonstrated by possessors in Tzotzil, an unrelated Mayan language.

(10) DP pied-piping with inversion in Tzotzil (Aissen, 1996)

a. I- k- il -be [s- tot li Xun] -e.
COMP- I- see -him his- father the X. -ENC
'I saw Xun's father.'

b. * I- k- il -be [li Xun s- tot].

c. [**Buch'u** s- tot]$_i$ av- il -be t_i.
who his- father you- see -him
'Whose father did you see?'

d. S- tot **buch'u** $_i$ av- il -be t_i?

In (10a-b), we see that a non-wh-possessor like *Xun* must occur after the possessum *s-tot* 'his father.' In contrast, the wh-possessor *buch'u* 'whose' in (10c-d) must precede the possessum. In this sense, the wh-phrase is obligatorily inverted with respect to the corresponding non-wh-phrase.

In SMP Mixtec, three categories demonstrate pied-piping with inversion. The first, like Tzotzil, is DP. This is shown in (11).

(11) SMP Mixtec pied-piping with inversion in DP

a. Rà$_i$ xa xìn ì [táta ñá Julia]$_i$.
he PERF see.PAST I father she J.
'I already saw Julia's father.'

b. * Rà$_i$ xa xìn ì [ñá Julia táta].

c. [Yó táta]$_i$ xa xìñ ú t_i?
who father PERF see.PAST you

'Whose father have you already seen?'

d. Táta yó $_i$ xa xìñ ú t_i?

In (11) we see that non-wh-possessors like *ñá Julia* 'Julia' must follow the possessum while wh-possessors like *yó* 'who' must precede the possessum. This pattern is identical to the Tzotzil in (10).

An identical pattern is seen in PP, shown in (12).

(12) PP pied-piping with inversion in SMP Mixtec

a. Rà$_i$ íntu'u rà Juan$_i$ [xihǐn míí be'e].
he sit.PRES.SG he J. beside the house
'Juan is sit beside the house.'

b. * Rà$_i$ íntu'u rà Juan$_i$ [míí be'e xihǐn].

c. [Nă xihǐn]$_i$ íntu'u rà Juan t_i?
what beside sit.PRES.SG he J.
'What is Juan sitting beside?'

d. Xihǐn nă $_i$ íntu'u rà Juan t_i?

(12) demonstrates the same inversion pattern in PP that we saw in DP in (11). In (12a-b), we see that prepositions like *xihǐn* 'beside' obligatorily precede non-wh-complements like *míí be'e* 'the house.' In contrast, (12c-d) demonstrate that when the nominal complement of a preposition is a wh-phrase like *nă* 'what,' this order inverts.

Finally, the same pattern is seen in CP. This pattern is commonly referred to as clausal pied-piping, and is well-described outside of Mesoamerica. See Ortiz de Urbina (1989, 1993) and Arregi (2003) for clausal pied-piping in Basque, and Cole (1982) and Hermon (1984) for Imbabura Quechua. Just like in these languages, clausal pied-piping involves first movement of the wh-phrase to Spec,CP of the embedded clause, followed by movement of the entire embedded clause to Spec,CP of the matrix clause.[5]

[5]Clausal pied-piping exists in free variation with the extraction patterns demonstrated in §4. Most speakers do not notice any difference between the two strategies, although some speakers indicate a preference for clausal pied-piping.

(13) a. Káchi rà Juan [kasa'a ñá Maria tamale].
say.PAST he J. make.IRR she M. tamales
'Juan said that Maria would make tamales.'

b. [Nă$_i$ kasa'a ñá Maria t_i]$_j$ káchi rà Juan t_j?
what make.IRR she M. say.PAST he J.
'What did Juan say Maria would make?'

Finally, all of these kinds of pied-piping may be combined. For instance, if a wh-phrase is embedded within a PP, and that PP sits in an embedded clause, both PP pied-piping with inversion as well as clausal pied-piping can occur. This is shown in (14b).

(14) a. Ká'a rà Juan [nì ka'an ñá Maria [xí'in se'e ñá]].
think.PRES he J. PAST talk she M. with child her
'Juan thinks that Maria is talking with her child.'

b. [[Yó xí'in]$_i$ nì ka'an ñá Maria t_i]$_j$ ká'a rà Juan t_i?
who with PAST talk she M. think.PRES he J.
'Who does Juan think Maria talked with?'

This can produce rather striking cases of pied-piping, particularly when a wh-phrase is embedded within a PP that occurs in an intermediate clause. Consider (15).

(15) Pied-piping with inversion in PP and CP

a. Xìni so'o rà Julio [káchi ñá Carmen xí'in ñá Maria [kuu
see.PAST ear he J. say.PAST she C. with she M. COP.IRR
sahbì itsyààn]].
rain tomorrow
'Julio heard [that Carmen told Maria [that it will rain tomorrow]].'

b. [[Yó xí'in]$_i$ káchi ñá Carmen t_i [kuu sá'hbi itsyààn]]$_j$
who with say.PAST she C. COP.IRR rain tomorrow
xìni so'o rà Julio t_j?
see.PAST ear he J.
'Who did Julio hear Carmen tell that it will rain tomorrow?'

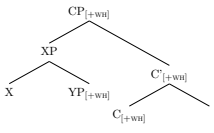
In (15a), the intermediate clause contains the PP *xí'in ñá Maria* 'to Maria.' In (15b), *ñá Maria* is replaced with the wh-phrase *yó* 'who.' When this happens, first pied-piping with inversion takes place within PP, yielding *yó xí'in* 'who with.' Then, this PP moves to Spec,CP of its containing clause, the intermediate clause. This produces *Yó xí'in káchi ñá Carmen...*. This intermediate clause then moves to Spec,CP of the root clause, pied-piping the most deeply embedded clause *kuu sahbì itsyààn* 'it will rain tomorrow.'

From this, we see that three categories license pied-piping with inversion in SMP Mixtec: DP, PP, and CP. The fact that it is these categories and only these categories that license this phenomenon is certainly not an accident. Particularly, these are the categories that are plausible phase heads in this language (Chomsky 2000, 2001; Abels 2003; Heck and Zimmermann 2004; Svenonius 2004).

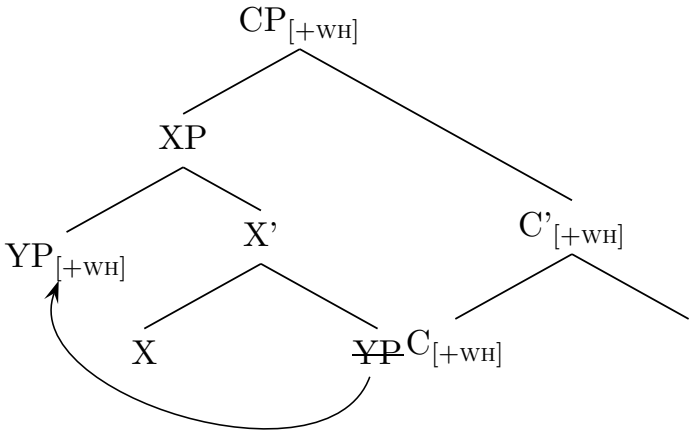
Indeed, the connection between phase theory and pied-piping with inversion is hinted at in Aissen (1996), although phase theory had not been developed at the time. Aissen proposes that wh-phrases must be in a highly local relationship with [+WH] C. This local relationship is blocked if a wh-phrase remains too deeply embedded within a containing phrase. This forces the wh-phrase to move to the specifier of its containing phrase in order to be accessible to [+WH] C. This is schematized in (16).

(16) Aissen's (1996) derivation of pied-piping with inversion

a. Ungrammatical configuration



b. Grammatical configuration

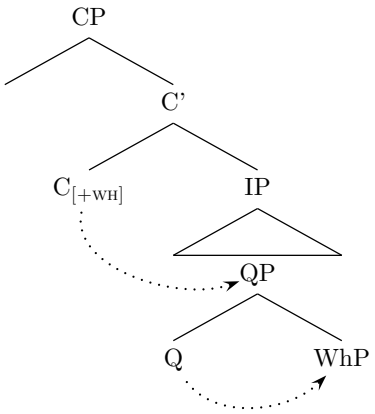


In (16a), the [+WH] YP is embedded within XP. Aissen proposes that in this position, it is too deeply embedded to enter into a necessary abstract agreement relation with $C_{[+WH]}$. Therefore, YP is forced to move to Spec,XP. From this position, it is accessible to abstract agreement with $C_{[+WH]}$.

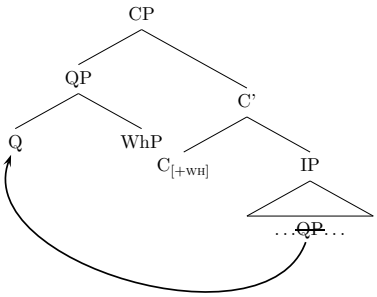
Aissen's core idea can be readily transposed both into modern phase theory and into modern theories of pied-piping, notably that developed by Cable (Cable 2007, 2010, 2012, 2013). In Cable's system, pied-piping with inversion is not a property of [+WH] constituents, but rather a category Q. Q and the wh-phrase must enter into an AGREE relation with each other, and $C_{[+WH]}$ and QP must AGREE as well. The agreement between QP and $C_{[+WH]}$ licenses movement of QP to Spec,CP, giving the appearance of wh-pied-piping. This system is demonstrated in (17).

(17)

a.



b.



Using the agreement and movement mechanisms demonstrated in (17), Cable (2010) establishes a typology of languages with wh-pied-piping along two vectors. The first is whether or not agreement between Q and the wh-phrase is necessary. The second has to do with where QP is allowed to merge. In some languages, it may merge anywhere in an extended projection in the sense of Grimshaw (2000, 2005), while in other languages, QP must merge with a complete extended projection. This typology

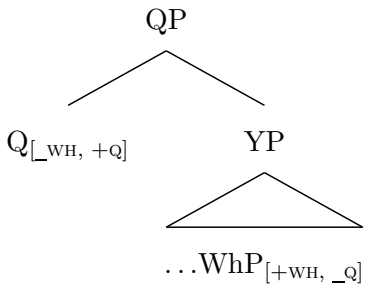
is shown in (18).

(18) Typology of QP

	AGREE: ✓	AGREE: **x**
Merge above E.P.	?	Ch'ol (Coon, 2009)
Merge Inside E.P.	English	Tlingit

Now, let us consider the typological gap in (18). This sort of language is sketched in (19).

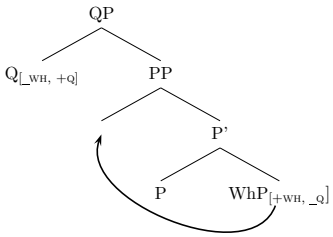
(19) Where YP is the highest category in an extended projection



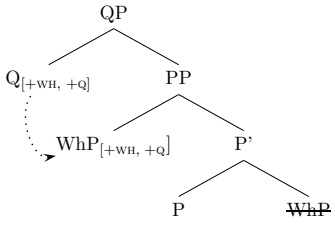
In (19), QP and the wh-phrase must enter into an AGREE relation. Importantly, the wh-phrase is embedded within the YP extended projection. Let us assume that YP is a phasal category. This means that the wh-phrase embedded within YP will not be able to agree with QP if it remains in situ, according to the PIC. I propose that this motivates movement of the wh-phrase to Spec,YP, escaping the phase. This movement allows the wh-phrase to enter into an AGREE relation with QP and derives the inversion in pied-piping with inversion. This is sketched for a PP in (20).

(20)

a.



b.



In (20), the wh-phrase enters the derivation as a complement of P, while QP is required to merge above PP. If P is a phase head, the wh-phrase will be unable to enter into an AGREE relation with QP, per the PIC. Therefore, the only way both QP and the wh-phrase can receive a value for their unvalued features is if the wh-phrase undergoes movement to Spec,PP. As the specifier is outside of the phase demarcated by P, the wh-phrase and QP may enter into an AGREE relation and subsequently the whole containing phrase may undergo A'-movement. This derives both the obligatory nature of pied-piping with inversion, as well as the inversion itself.

If this analysis is on the right track, we identify three categories that are phasal in SMP Mixtec: CP, DP, and PP. This is because these are the three categories that induce pied-piping with inversion, as shown in (11-13). With this established, let us reconsider the patterns of pronoun doubling that we saw in §3.

5.1.1.2 Pronoun doubling may not occur into a phase

In §3, we saw that pronoun doubling may not reach into any of the categories which we just identified as phases, namely CP, DP, and PP. Let us review these data.

First, pronoun doubling may never target objects of prepositions. See §3.3.3 for more data, but (21) suffices to demonstrate this conclusion.

(21) a. **Rà**$_i$ íyǐ'bi rà$_i$ [nuhŭ nána rà].
he afraid.PRES he of mother his
'He's afraid of his mother.'

b. * **Ñá**$_i$ íyǐ'bi rà [nuhŭ nána$_i$ rà].
she afraid.PRES he of mother his

c. ☒ Kusijǐ [in ì].
happy inside me
'I am happy.'

d. * Yù'u$_i$ kusijǐ [in ì$_i$].
I happy inside me

In (21b), we see that pronoun doubling may not target the complement of a prepositional object, in this case *nuhŭ nána rà* 'of his mother.' Likewise, in (21d) we see that complements of prepositional subjects, here *in ì* 'in me,' may not be targeted either. I interpret this observation as showing that pronoun doubling obeys the PIC with respect to the PP phase.

§3.3.3 also presents data which show that pronoun doubling also respects the DP phase. Here, we saw that possessors may never be targeted by pronoun doubling. This can be accounted for naturally if, as argued above, DP is a phase and pronoun doubling obeys the PIC. These data are reproduced in (22).

(22) a. **Rí**$_i$ xa nùhmĭ ísi'i rà Macario tsĭna$_i$ lo'o.
it.AML PERF hug.PAST wife he M. dog little
'Macario's wife already hugged the puppy.'

b. * **Rà**$_i$ xa nùhmĭ ísi'i rà Macario$_i$ tsĭna lo'o.
he PERF hug.PAST wife he M. dog little
Intended: 'Macario's wife already hugged the puppy.'

c. **Ñá**$_i$ xa xín ì ísi'i$_i$ rà Mario, so tá kon ì ísi'i rà Juán.
she PERF see.PRES I wife he M. but IRR.NEG see I wife he J.
'I already know Mario's wife, but I still haven't met Juan's wife.'

d. * **Rà**$_i$ xa xín ì ísi'i rà Mario$_i$ ...
he PERF see.PRES I wife he M.
Intended: 'I haven't met Mario's wife...'

In (22b) and (22d), we see that possessors may not be targeted by pronoun doubling. I take this to be further evidence that this process obeys the PIC.

Finally, in §3.3.1, we saw that pronoun doubling is clause bounded. In other words, pronoun doubling may not cross the CP phase. In that section, we saw this in two ways. The first comes from the inability of pronoun doubling to target nominals within finite embedded clauses. These data are shown in (23), particularly (23c).

(23) a. Tát ì ndatu'un rà Juán xí' ìn áto ba'á ñá Maria xí'a nì
wait.PRES I chat.IRR he J. with me if good she M. after PAST

tahan.
quake
'I hope that Juan tells us that Maria is well after the earthquake.'

b. **Rà**$_i$ xa ndàtu'un rà Juán$_i$ xí' ìn [íyo ba'á ñá Maria], so
he PERF said he J. with me be.PRES.SG good she M but
xi̱n ì xa'á ñá Julia.
know.NEG I about she J.
'Juan already told me that Maria is fine, but I don't know about Julia.'

c. * **Ñá**$_i$ xa ndàtu'un rà Juan xí' ìn [íyo ba'á ñá Maria$_i$]...

In (23), we create a context in which *ñá Maria* is highly topical. Despite this, pronoun doubling may not target it, as we see in (23c). I take this to indicate that pronoun doubling obeys the PIC with respect to finite CP.

Embedded irrealis clauses introduced by *ná* which do not involve Obligatory Control, argued to have a CP layer in §3.1, also block pronoun doubling. Consider (24).

(24) a. Ná kúndàà ini Roberto níhìn kwê ñá Maria, koni rà ná
SUBJ know.PRES inside R. skinny very she M. want.IRR he SUBJ
kaxi ñá chocolate.
eat.IRR she chocolate
'Should Roberto know how skinny Maria is, he'd want her to eat chocolate.'

b. **Rà**$_i$ xa kôni rà$_i$ [ná kaxi ñá Maria chocolate].
he PERF want.PRES he SUBJ eat.IRR she M. chocolate
'He already wants Maria to eat chocolate!'

c. * **Ñá**$_i$ xa kôni rà [ná kaxi ñá Maria$_i$ chocolate].
she PERF want.PRES he SUBJ eat.IRR she M. chocolate

d. * **Ñà**$_i$ xa kôni rà [ná kaxi ñá Maria chocolate$_i$].
it PERF want.PRES he SUBJ eat.IRR she M. chocolate

In (24c-d), we see that pronoun doubling may not occur into these embedded CPs either.

From this, we can conclude that pronoun doubling obeys the PIC with respect to each of the phases we have identified in this language. I consider this to be supporting evidence that pronoun doubling is a form of φ-agreement.

5.1.2 Pronoun doubling does not affect variable binding

den Dikken (1995), among many others (see Rezac 2010 for an overview), observes that φ-agreement does not affect anaphoric relations in the same way that phrasal movement does. Consider the paradigm in (25).

(25) (den Dikken 1995)

a. Some applicants$_i$ seem -$\boxtimes_i$ to each other$_i$ to be eligible for the job.

b. * There seem -$\boxtimes_i$ to each other$_i$ to be some applicants$_i$ eligible for the job.

In (25a), we see that A-movement of 'some applicants' creates a new A-position from which the reciprocal anaphor 'each other' in matrix experiencer position can be bound. In contrast, in (25b) we see that simply agreeing with 'some applicants' in the absence of A-movement is insufficient to license the reciprocal. This yields the Condition A violation in (25b).

Similarly, φ-agreement does not create an A-position from which variables may be bound. Consider the paradigm in (26), inspired by that from den Dikken (1995) in (25).

(26) a. Some women$_i$ seem -$\boxtimes_i$ to their$_i$ peers to be eligible for the job.

b. * There seem -$\boxtimes_i$ to their$_i$ peers to be some women$_i$ eligible for the job.

In (26a), we see that movement of 'some women' allows it to bind the variable 'their' in the matrix experiencer. Crucially, φ-agreement with 'some women' without movement is insufficient to bind the variable, as (26b) shows.

From this, we can conclude that φ-agreement does not affect binding in the same way that phrasal movement can. This matters for our purposes because we have already seen that pronoun doubling has no affect on binding relations. In those earlier discussions, notably in §3.3.2, this lack of an effect on binding was seen as problematic. But if pronoun doubling is φ-agreement, the absence of such effects is expected.

5.1.3 Only one doubled pronoun per clause

In §2.1.5, we discussed the claim in Baker (2012) that φ-agreement can be reliably diagnosed by the simple cardinality of exponence. Baker claims, based on the mechanics of Chomskyian AGREE, that φ-agreement yields only one marker per clause. This is because once a probe that needs φ-features receives them, there is no way to expone agreement again without a second probe.

In SMP Mixtec, pronoun doubling may occur only once per clause. (28) is based on (5) above.

(27) a. **Rà**$_i$ xá'antsya rà Juan$_i$ chìkí.
he cut.PRES he J. tuna
'Juan is cutting *tunas*.'

b. **Rí**$_i$ xá'antsya rà Juan chìkí$_i$.
it.AML cut.PRES he J. tuna
'Juan is cutting *tunas*.'

c. * **Rà**$_i$ **rí**$_j$ xá'antsya rà Juan$_i$ chìkí$_j$.

d. * **Rí**$_j$ **rà**$_i$ xá'antsya rà Juan$_i$ chìkí$_j$.

In (28c-d), we see that multiple nominals may not be pronoun doubled within a single clause, no matter what order they might occur in. From this, it is clear that only one nominal may be doubled per clause, and that there may be exactly one pronoun double per clause. This is consistent with φ-agreement.

As we will see shortly in the coming section, topicality has an important effect on pronoun doubling. With this in mind, consider (29).

(28) a. Ndása kù'u ná, ñá Juana ra ñá Maria, kóni ná ndása
two sister they.F she J. and she M. love.PRES they.F two
ñanì, rà Roberto ra rà Julio.
brother he R. and he J.
'Two sisters, Juana and Maria, love two brothers, Roberto and Julio.'

b. Ñá Juana, **ñá**$_i$ xa nùhmǐ **ñá**$_i$ rà Roberto, so ñá Maria, xa kò
she J. she PERF hug.PAST she he R. but she M. PERF NEG
ní nuhmǐ ñá rà Julio.
PAST hug she he J.
'As for Juana, she already hugged Roberto.'

c. Ñá Juana, **rà**$_i$ xa nùhmǐ ñá **rà** **Roberto**$_i$, so ...
she J. he PERF hug.PAST she he R. but
'As for Juana, she already hugged Roberto, but...'

d. * Ñá Juana, **ñá**$_j$ **rà**$_i$ xa nùhmǐ ñá$_j$ **rà Roberto**$_i$, so...

e. * Ñá Juana, **rà**$_i$ **ñá**$_j$ xa nùhmǐ ñá$_j$ **rà Roberto**$_i$, so...

(29a) sets up a context in which both *ñá Juana* and *rà Roberto* are highly topical. As such, in (29b-c), we see that either may be doubled. Crucially, in (29d-e), we see that they cannot both be doubled at the same time. Therefore, the effect in (28) holds even in cases where there are multiple topics in the same clause. This further supports the conclusion that only one doubled pronoun may occur per clause, which is consistent with φ-agreement.[6]

5.1.4 Pronoun doubling freely allows Closest Conjunct Agreement

In §2.1.7 and §2.2.5, we saw that φ-agreement and clitic doubling behave differently in coordinate structures cross-linguistically. Where φ-agreement never has any trouble targeting coordinate structures and frequently yields Closest Conjunct Agreement, clitic doubling out of a coordinate structure is much more marginal. Usually, languages simply prohibit clitic doubling out of coordinate or disjunct structures, as we saw with the Puerto Rican Spanish in §2.2.5.1.[7] This result falls out naturally if clitic doubling is derived by movement, as movement from a coordinate structure would

[6]There is a possible confound here. As we saw in chapter 1, there are certain prosodic conditions that block pronoun doubling from occurring. It may be the case that multiple doubled pronouns violate these prosodic requirements, while only one satisfies them. Disentangling this confound would lead us too far astray. Therefore, I must acknowledge the potential issue and continue.

[7]See Bošković (2018) for some other potential cases of clitic doubling from coordinate structures.

violate the Coordinate Structure Constraint (Ross 1967).

In SMP Mixtec, pronoun doubling freely targets coordinate structures. To begin, SMP Mixtec has two coordinators that may occur with DPs. The first is *ra* 'and,' which may be used to coordinate both DPs and other categories of constituents. The second is *xí'in*. This usually means 'with,' but it may also mean 'and' when coordinating DPs. The *xí'in* 'with/and' coordinator can only occur with DPs. Both coordinators are shown in (30).

(29) a. **Nà**$_i$ nàki -tá'àn [ñá Maria **ra** rà Juan]$_i$.
they meet.PAST -RECIP she M. and he J.
'Maria and Juan met each other.'

b. **Nà**$_i$ kàni -tá'àn [ñá Maria **xí'in** rà Juan]$_i$.
they fight.PAST -RECIP she M. with he J.
'Maria and Juan argued (lit. fought each other).'

In (30), we see that both coordinate structures, conjoined either by *ra* 'and' or *xí'in* 'with,' behave like true coordinate structures because both bind the reciprocal incorporated anaphor *-tá'àn* 'each other.' Importantly, both allow a plural pronoun, in this case *nà* 'they' to double them.

Interestingly, this language also allows a pattern identical to Closest Conjunct Agreement in which only the leftmost conjunct is doubled, but this pattern seems to be restricted to the *xí'in* 'with' coordinator. Compare the grammatical (31a) and (30) to the ungrammatical (32).

(30) a. **Ñá**$_i$ nàki -tá'àn [ñá Maria$_i$ **xí'in** rà Juan].
she meet.PAST -RECIP she M. with he J.
'Maria and Juan met each other.'

b. **Ñá**$_i$ kàni -tá'àn [ñá Maria$_i$ **xí'in** rà Juan].
she fight.PAST -RECIP she M. with he J.
'Maria and Juan argued.'

(31) a. * **Ñá**$_i$ nàki -tá'àn [ñá Maria$_i$ **ra** rà Juan].
she meet.PAST -RECIP she M. and he J.
Intended: 'Maria and Juan met each other.'

b. * **Ñá**$_i$ kàni -tá'àn [ñá Maria$_i$ **ra** rà Juan].
she fight.PAST -RECIP she M. and he J.
Intended: 'Maria and Juan argued.'

In comparing (31) to (32), we see that Closest Conjunct Agreement is compatible only with coordinate structures that use the coordinator *xí'in* 'with.' Crucially, it cannot be the case that in (31) the coordinate structures with *xí'in* 'with' is interpreted as a preposition instead of a coordinator. If it were a preposition here, we would be unable to bind the reciprocal *tá'àn* 'each other,' yielding an ungrammatical interpretation similar to '*Maria met each other with Juan' for (31a). As we know that (31a) is grammatical, it must be the case that these coordinate structures with *xí'in* 'with' are true coordinate structures.

Additionally, disjunctive structures do not block pronoun doubling. Indeed, disjunction requires Closest "Disjunct" Agreement.

(32) a. **Ñá**$_i$ ntá'yi [ñá Maria$_i$ á rà Juan].
she cry.PRES she M. or he J.
'Maria or Juan is crying.'

b. * **Nà**$_i$ ntá'yi [ñá Maria á rà Juan]$_i$.
they cry.PRES she M. or he J.
Intended: 'Maria or Juan is crying.'

c. * **Rà**$_i$ ntá'yi [ñá Maria á rà Juan$_i$].
she cry.PRES she M. or he J.
Intended: 'Maria or Juan is crying.'

If pronoun doubling were derived from movement, we would expect (30) and (32) to be ungrammatical due to the Coordinate Structure Constraint of Ross (1967). This stands in contrast with φ-agreement systems, which commonly allow Closest Conjunct Agreement. See McCloskey and Hale (1984), McCloskey (1986), Munn (1993,

1999), Aoun et al. (1994), Soltan (2006), Marušič et al. (2007), Benmamoun et al. (2009), Boškavić (2009), and Bhatt and Walkow (2013), among many others.

5.1.5 Features of pronoun doubling that are not easily fit with φ-agreement

So far, we have seen that pronoun doubling in SMP Mixtec displays many commonalities with φ-agreement. Furthermore, we have seen in chapters 3 and 4 that pronoun doubling displays characteristics that are hard to reconcile with movement-based theories of the relation between the pronoun and the nominal it doubles.

It seems, then, that we have good reason to adopt (33).

(33) Pronoun doubling is φ-agreement.

But we must also face some observations that, at first blush, seem difficult to reconcile with (33).

First, as we saw in (5), pronoun doubling may target both subjects and objects depending on the conversation in which the sentence is uttered. (5) is repeated in (34).

(34) a. **Rà**$_i$ xá'antsya rà Juan$_i$ chìkí.
he cut.PRES he J. tuna
'Juan is cutting *tunas*.'

b. **Rí**$_i$ xá'antsya rà Juan chìkí$_i$.
it.AML cut.PRES he J. tuna
'Juan is cutting *tunas*.'

At first glance, the variability in (34) is problematic for a φ-agreement account because AGREE is subject to strong locality restrictions. If the subject can satisfy the presumed φ-probe, as we see in (34a), it is unclear why this same probe would ever be able to interact with the object, as in (34b).

Second, pronoun doubling shows semantic restrictions of the sort that are commonly found in clitic doubling systems. See the discussion in §2.2.2 and the works cited

there. In SMP Mixtec, this restriction is extremely similar to that observed by Jaeggli (1982) for accusative clitic doubling in Rioplatense Spanish: only referential nominals, or D-linked nominals in the sense of Pesetsky (1987), may be doubled. The clearest way to see this is by considering quantificational expressions. Universal quantifiers may be doubled only if there is a specific, contextually salient set corresponding to the restriction of the quantifier. Contrast the grammatical (35) with the ungrammatical (36).

(35) Ñǔù koni xinti' ì be'e ñá *amigo* ña' ì, so ní kǔu
night yesterday sit.PAST.SG I house she friend POSS my but PAST can.NEG
ti míí ì kùsì. **Nà**$_i$ sîsò xitsin ntsi'i nà$_i$!
EMPH the I sleep.IRR they boil.PRES nose every they
'Last night, I spend the night at a friend's house, but I couldn't sleep. Everyone was snoring (lit. boiling nose)!'

(36) Kǔntsi'ì iñ ú, se'e. (***Nà**$_i$) sîsò xitsin ntsi'i nà$_i$.
worry.NEG.IRR inside you child they boil.PRES nose every they
'Don't worry, child (*no te preocupes, mijo*), everyone snores (lit. boils nose).'
Context: A child is embarrassed because they found out they snore, and you are trying to comfort them.

In (35), we see that when *ntsi'i nà* 'everyone' refers to a clear, context salient set, it may be agreed with. In contrast, in the context in (36) in which *ntsi'i nà* 'everyone' does not refer to a context salient set, it may not be targeted for agreement. This sort of sensitivity to the pragmatic context is unexpected from a run-of-the-mill φ-agreement system, as discussed in §2.1.3.

In the rest of this chapter, I try to show that these unexpected aspects of pronoun doubling can be integrated into a φ-agreement system in an interesting way once we make one key observation: Pronoun doubling specifically targets topics.

5.2 Pronoun doubling targets topics

In this section, I will demonstrate that pronoun doubling targets topics. In making this case, of course, we immediately face the difficulty that the term 'topic' has been applied to many things, dating back at least to Aristotle. Even within the relatively recent history of generative linguistics, a wide variety of phenomena have been labeled 'topics.' These range from the contrastive topics of English (Chafe 1976) and Japanese (Kuno 1976), the "free" topics of Mandarin, Lahu, Korean, and Japanese (Li and Thompson 1976; Huang et al. 2009), as well as the well-known "aboutness" topics of Strawson (1964), Reinhart (1981), Aissen (1992), Vallduví (1992), Vallduví and Engdahl (1996), and Krifka (2008).[8]

We can ignore free topics because SMP Mixtec lacks them. Compare the Lahu and Mandarin in (37) to the ungrammatical SMP Mixtec in (38).

(37) Free topics (Li and Thompson 1976)

a. [Hɛ chi tê pêʔ] ɔ̄ dàʔ jâʔ.
field this one CLASS rice very good
'This field (topic), the rice is very good.' (Lahu)

b. [Nèi -chang huǒ] xìngkui xīaofangduì laí de kuài.
that -CLASS fire fortunate fire.brigade come ADV quick
'That fire (topic), fortunately the fire brigade came quickly.' (Mandarin)

(38) No free topics in SMP Mixtec

a. * [Yúkŭ kân], ba'ă kwê'e chìchí.
forest that good very avocado
Intended: 'That forest, the avocados are very good.'

b. * [Ñû'ù kân], ba'ă kwê'e kamá kìxàà míí nà *bombero*.
fire that good very fast come.PAST the they fire.fighter
Intended: 'That fire, thankfully the fire fighters came quickly.'

[8] Frascarelli and Hinterhölzl (2007) introduce an additional category of topic which they term "familiar" topics. I will not discuss this sort of topic, because it is not entirely clear how to distinguish these from backgrounded aboutness topics.

In (37), we see that Lahu and Mandarin allow so-called free topics, which are not directly connected to the argument structure of the predicate. See Huang et al. (2009), chapter 6, for an overview from the perspective of Mandarin. In (38), we see that SMP Mixtec, like English, does not allow free topics.

So let us ask what kinds of topics SMP Mixtec actually does have. As Aissen (To appear) observes, the simplest way to identify and elicit topics cross-linguistically is to examine the answer to a wh-focus question. See Dik (1978) as well. Consider the example provided in (39).

(39) a. Where did the children go?

b. [The children]$_{Topic}$ went (somewhere$_x$) [x= to school]$_{Focus}$

In (39b), the prepositional phrase 'to school' is in focus, as this provides the information requested by the wh-element 'where' in (39a). What is left over in (39b), 'the children went,' must contain a topic because this is the information relative to which the focused locative 'to school' is meant to be interpreted. Within this phrase, there is only one entity, specifically 'the children.' Therefore, 'the children' is a topic.

Important for our purposes, pronoun doubling in SMP Mixtec faithfully tracks topics. To see this, recall that speakers judge the sentences in (5), repeated in (34), as truth-conditionally equivalent, but the difference between them *depende en la conversación* 'depends on the conversation.' Specifically, the choice between them depends on which is the topic. This is demonstrated in (40-41).

(40) a. Nă xá'antsya rà Juan?
what cut.PRES he J.
'What is Juan cutting.'

b. **Rà**$_i$ xá'antsya [rà Juan$_i$]$_{Top}$ [chìkí]$_{Foc}$.
he cut.PRES he J. tuna
'Juan is cutting *tunas*.' *Answer to (37a).*

c. #? **Rí**$_j$ xá'antsya [rà Juan$_i$]$_{Top}$ [chìkí$_j$]$_{Foc}$.
it.AML cut.PRES he J. tuna
Intended: 'Juan is cutting *tunas*.' *Intended as an answer to (37a).*

(41) a. Yó nà xá'antsya chìkí?
who they cut.PRES tuna
'Who is cutting the tunas?'

b. **Rí**$_j$ xá'antsya [rà Juan]$_{Foc}$ [rí chìkí$_j$]$_{Top}$.
it.AML cut.PRES he J. it.AML tuna
'Juan is cutting *tunas*.' *Answer to (38a).*

c. #? **Rà**$_i$ xá'antsya [rà Juan$_i$]$_{Foc}$ [rí chìkí]$_{Top}$.
he cut.PRES he J. it.AML tuna
Intended: 'Juan is cutting *tunas*.' *Intended as an answer to (38a).*

(40a) provides a wh-question in which the wh-phrase *nǎ* 'what' corresponds to the direct object. Therefore, in (40b), the answer to (40a), the direct object is in focus while the subject is the topic. Importantly, we see that pronoun doubling must track the topic DP. This is seen in comparing the grammatical (40b), in which pronoun doubling tracks the topic, to the ungrammatical (40c), where pronoun doubling attempts to track the focused element. (41) shows an identical pattern, except that the object is the topic in the target sentence in (41b), rather than the subject.

With this, we establish two properties about pronoun doubling. First, as stated in (34), pronoun doubling is φ-agreement. Second, from (40-41) we see that pronoun doubling faithfully tracks topics. Therefore, I propose that pronoun doubling is a form of topic agreement, defined in (42).[9]

[9]The term 'topic agreement' has been used to refer to other phenomena that are clearly not meant to be analyzed with φ-agreement of the sort assumed here. Notable among these is the use of this term by Bresnan and Mchombo (1987) to refer to incorporated pronouns in Chichewa that do not stand in a formal syntactic relation to sentence topics. As such, I do not engage with their work here or throughout, because they clearly have in mind something different from φ-agreement or clitic doubling of the sense understood here. Regardless, their work is certainly related to mine here in spirit.

Additionally, as Chung (p.c.) observes, there seems to be a connection between topic agreement and the so-called obviation constructions in Algonquian. See Aissen (1997) and the citations therein for a comprehensive overview of the phenomenon. While I agree that there is, on an abstract spiritual level,

(42) Topic agreement: A φ-agreement system in which topics are the targets for agreement.

While topic agreement may seem exotic, it turns out to be more common cross-linguistically than one might expect. A particularly clear example of topic agreement in another language comes from Dalrymple and Nikolaeva (2005), who show that object agreement in Northern Khanty[10], a Uralic language of Russia, is only possible when the object is construed as a topic. This is seen in comparing the ungrammatical (43b) to the grammatical (43c).

(43) Northern Khanty topic agreement (Dalrymple and Nikolaeva 2005)

a. Mati kalaŋ$_i$ we:l -əs / *we:l -s -əlli$_i$?
which reindeer kill -PAST.3SG.SUBJ kill -PAST -OBJ.3SGSUBJ
'Which reindeer did he kill?'

b. [Tam kalaŋ]$_{Foc}$ we:l -əs / *we:l -s -əlli$_i$.
this reindeer kill -PAST.3SG.SUBJ kill -PAST -OBJ.3SGSUBJ
'He killed this reindeer.' *Answer to (40a).*

c. [Tam kalaŋ$_i$]$_{Top}$ we:l -s -e:m$_i$ / *we:l -s -əm.
this reindeer kill -PAST -OBJ.1SGSUBJ kill -PAST -1SGSUBJ
'I killed this reindeer.' *Answer to 'What did you do to this reindeer?'*

In (43b), we see that when the object *tam kalaŋ* 'this reindeer' is in focus, it cannot be agreed with. In contrast, when *tam kalaŋ* 'this reindeer' is the topic, as in (43c), it must be agreed with. This is extremely reminiscent of the pattern in (40-41).

Northern Khanty is not the only language to demonstrate such a pattern. Famously, Oromo, a Cushitic language of Ethiopia, only allows verbs to agree with the subject if the subject is a topic. This is shown in (44).

certainly some connection the two notions (see especially Goddard 1984, 1990 and Dryer 1994), the term 'obviation' usually implies a very strict definition, namely, the syntactic device that distinguishes between animate third person nominals in a discourse. As the definition in (42) makes clear, I have something more broad in mind.

[10]Formally known as Northern Ostyak.

(44) Oromo topic agreement (Clamons et al. 1999)

a. [Ati ifi Salma -á -n]$_i$ nyataa godhu tur **-tan**$_i$.
you.SG and S. -SUBJ -TOP food make were -2PL.PAST
'You and Salma were cooking.'

Answer to Eessaa'n ture hoga gurbaan dhufe? *'Where was I when the boy came?'*

b. [Ati ifi Salma -á]$_i$ nyaata godu tur **-e**.
you.SG and S. -SUBJ food make were -PAST
'You and Salma were cooking.'

Answer to Eenyuu nyaata godhu ture? *'Who was cooking?'*

In Oromo, topics are clearly marked by the topic marker *-n*, as in (44a). In comparing (44a) to (44b), we see that subject agreement is possible only if the subject is marked as a topic.

In addition to the Northern Khanty in (43) and the Oromo in (44), similar topic agreement patterns have been widely attested. Languages reported to have topic agreement include Kinyarwanda (Bantu, Morimoto 2006), Albanian (Indo-European, Kallulli 2016), Maithili (Indo-Aryan, Stump and Yadav 1988 and Dalrymple and Nikolaeva 2005), Nonstandard French (Indo-European, Lambrecht 1981), Ripano (Indo-European, D'Alessandro 2017), Tsez (Nakh-Daghestanian, Polinsky and Potsdam 2001), Innu-aimûn (Algonquian, Branigan and Mackenzie 2002), Najdi Arabic (Afroasiatic, Alshamari 2017), and Netherlands Sign Language (French Sign, Crasborn et al. 2009). Miyagawa (2017) also discusses several important connections between φ-agreement and topics, and Givón (1976, 1983) proposes that all φ-agreement systems derive historically from topic-comment constructions (see Lehmann 1988, Corbett 1995, 2006, Hopper and Traugott 1993, Roberts and Roussou 2003, and Fuß 2005 as well). I add to that typological literature here and I attempt to show that a theoretically satisfying analysis is made available within the terms of contemporary versions of the minimalist framework for syntax.

Now that we have seen that the notion of topic is broadly compatible with the targets of pronoun doubling, let us refine our notion. Below, I will demonstrate that topics in SMP Mixtec have the following set of properties.

(45) A topic in SMP Mixtec is

a. an established referent in the Common Ground.

b. identifiable, in the sense of Chafe (1976).

c. the subject of categorical judgments (see Kuroda 1972; McNally 1998, and Jäger 2001).

The sense of topic intended in (45) is fairly specific. Instead of referring to a broad, abstract notion of aboutness at the level of discourse, topic for me is a highly local notion restricted to single clauses. In this way, topics of SMP Mixtec closely resemble the 'sentence topics' of Davison (1984), the 'clause level topics' of Downing (1991), the 'clause level themes' of much work in the Prague School (see Tomlin et al. 1997 for an overview of this connection), among many others. My goal here is not to adjudicate amongst these theories of topichood, or even to provide the most succinct semantic generalization for topics in this language. Rather, my goal is to observe that, at their core, they all describe, more or less, the same class of objects, and this local notion of a clausal topic is useful here. It is my hope that this work can be used as a stepping stone towards a more careful semantic analysis in the future.

By analyzing pronoun doubling as topic agreement, several other welcome results fall out naturally. Chief among these are the interpretive restrictions on nominals that may be pronoun doubled. Let us examine the definition of topic from (39) more closely. In this system of topichood, there must be a specific entity within the Common Ground, in the sense of Stalnaker (2002), for new information to be associated with. To use the metaphor of Heim (1982), building on work in Reinhart (1981), there must be an established 'file card' in the Common Ground that can host new information. As

Chafe (1976) puts it "identifiable" is a better term, meaning "I think you already know and can identify the particular referent I have in mind" (pg. 39).[11] As such, I will use the term "identifiable" in the rest of the discussion.

At the empirical level, analyzing pronoun doubling as topic agreements means that only identifiable nominals should be targeted. This turns out to be the case. To see this, let us consider the behavior of various quantificational and non-referential elements. First, consider. *iin ña'à* 'something.' Like its English translation, *iin ña'à* may be interpreted as specific and referential, i.e., identifiable, or non-specific and non-referential, or non-identifiable. When interpreted as identifiable, *iin ña'à* 'something' must be pronoun doubled, but when non-identifiable, *iin ña'à* 'something' may not be doubled. This is shown in (46).

(46) a. ☒ Nàkààba iin ña'à, so xǐn ì nă kíí á.
fall.PAST one thing but know.PRES.NEG I what COP.PRES it
'Something fell, but I don't know what it is.'

Context: We are sitting in a room and hear something fall outside. I do not recognize the object based on the sound it makes falling.

b. # **Ñà**$_i$ nàkààba iin ña'à$_i$, so xǐn ì nă kíí á.
it fall.PAST one thing but know.PRES.NEG I what COP.PRES it
'Something fell, but I don't know what it is.'

Context: Same as (46a). Judged as infelicitous. Speaker responded by picking up her handbag, making direct eye contact with me, and dropping the handbag. She then said "See? It doesn't make sense. That makes it sound like you know what you're talking about, but then saying you don't know."

c. **Ñà**$_i$ nàkààba iin ña'à$_i$, ra xín ì nă kíí á.
it fall.PAST one thing and know.PRES I what COP.PRES it
'Something fell, and I know what it is.'

[11] See Strawson (1964) as well.

Context: We are sitting in a room and hear something fall outside. By the sound it makes when it hits the floor, I recognize what it is.

d. ?? ☒ Nàkààba iin ña'à$_i$, ra xín ì nă kíí á.
fall.PAST one thing and know.PRES I what COP.PRES it
'Something fell, and I know what it is.'

Context: Same as (46c). Judged as a bit better than (46b) to (46a), although having the doubled pronoun, as in (46c), is preferred.

In (46a-b), we see that when *iin ña'à* 'something' is meant to be construed as non-identifiable, it may not be doubled. With this in mind, let us consider (46b). Here, *iin ña'à* 'something' is pronoun doubled, even though a non-identifiable interpretation is still forced by the context and by the adjunct clause *so xĭ̌n ì nă kíí á* 'but I don't know what it is.' Because pronoun doubling is topic agreement, it requires that the nominal it cross-references be interpreted as identifiable. This leads to a contradiction in (46b). Now, let us consider (46c-d). Here, the context supports a identifiable interpretation of *iin ña'à* 'something.' As such, pronoun doubling is strongly preferred.

Through this lens, let us also return to (35-36). These are repeated as (47-48).

(47) Ñŭù koni xinti' ì be'e ñá *amigo* ña' ì, so ní kŭu
night yesterday sit.PAST.SG I house she friend POSS my but PAST can.NEG
ti míí ì kùsì. **Nà**$_i$ sîsò xitsin ntsi'i nà$_i$!
EMPH the I sleep.IRR they boil.PRES nose every they
'Last night, I spend the night at a friend's house, but I couldn't sleep. Everyone was snoring!'

(48) Kŭntsi'ì iñ ú, se'e. (***Nà**$_i$) sîsò xitsin ntsi'i nà$_i$.
worry.NEG.IRR inside you child they boil.PRES nose every they
'Don't worry, child (*no te preocupes, mijo*), everyone snores (lit. boils nose).'

Context: A child is embarrassed because they found out they snore, and you are comforting them.

In (47), we create a context where the set which is in the restictor for *ntsi'i nà* 'everyone'

is salient in the Common Ground: these are the people who were inside the friend's house last night. Furthermore, we see that in this identifiable context, *ntsi'i nà* 'everyone' may be targeted for topic agreement. This contrasts with the use of *ntsi'i nà* 'everyone' in (48). Here, the context requires that *ntsi'i nà* 'everyone' refer to individuals both internal to and external to the Common Ground. Therefore, by definition, the referents of *ntsi'i nà* 'everyone' in (48) are not identifiable in the Common Ground. This explains why *ntsi'i nà* 'everyone' may not be targeted for topic agreement in this context.

Another way in which identifiability determines patterns of topic agreement can be seen in what I call *nà*-impersonal constructions, discussed briefly at several earlier points. These are roughly equivalent in discourse function to the English passive, or the *se*-passive of Spanish (see Suñer 1976 and D'Alessandro 2004 for a more recent discussion). The *nà*-impersonal construction is shown in (49), repeated from (1) in §3.

(49) a. Ndànuûn nà tsǐn nuhŭn *kama*.
put.on.SG.PAST they mouse on bed
'They put the mouse on the bed.'

Provided for: 'The mouse was put on the bed.'

b. Nì ixăni rà xà'ani nà rà.
PAST dream he kill.PAST they him
'He dreamed that he was killed.'

In (49), we see several sentences that were translated by speakers using a passive. In SMP Mixtec, these involve an overt pronominal subject *nà* 'they,' which is not anaphorically related to any discourse referent. This yields an impersonal interpretation.

Because impersonal-*nà*, by definition, does not pick out any identifiable referent in the Common Ground, we predict that it should not be able to be targeted for topic agreement. This turns out to be the case.

(50) a. **Tún**$_i$ chí'i nà ndòho$_i$ Ñûù Kă'anu.
it.WOOD sow.PRES they sugarcane town big

'Sugarcane is grown in San Martín Peras (lit. the big town).'

Provided for '*Se cultivan la caña de azúcar en San Martín Peras.*'

b. # **Nà**$_i$ chí'i nà$_i$ ndòho Ñûù Kă'anu.
they sow.PRES they sugarcane town big
Intended: 'Sugarcane is grown in San Martín Peras.'

Judgment: "*Está bien si hablamos de unas personas y quiero decir 'Ellos cultivan la caña de azúcar en San Martín Peras'.*" "It's good if we're talking about some people and I want to say 'They grow sugarcane in San Martín Peras'."

c. **Rí**$_i$ xáxi nà chìchí$_i$ nù Ñûù Ndŏ'ba.
it.AML eat.PRES they iguana by town Oaxaca
'Iguana is eaten by Oaxaca City.'

Provided for *Se comen iguanas cerca de la Ciudad de Oaxaca.*

d. # **Nà**$_i$ xáxi nà$_i$ chìchí nù Ñûù Ndŏ'ba.
they eat.PRES they iguana by town Oaxaca
Intended: 'Iguana is eaten by Oaxaca City.'

Judgment similar to (49b).

In (50a) and (50c), we see that it is impossible to double impersonal-*nà*. If pronoun doubling is attempted with the subject, as in (50b,d) the impersonal reading of *nà* 'they' is obligatorily lost. This makes sense if topic agreement requires an identifiable discourse referent to serve as topic. This explains the forced referential interpretation observed in (50b, d).[12]

In a similar vein, *ní iin nà* 'no one' may never be pronoun doubled, even if a context salient set is present. Consider (51).

[12]These data also have an explanation in the thetic/categorical distinction. Impersonal constructions are most likely judged as thetic. Thetic judgments, as shown in (57), are incompatible with topic agreement. The speaker judgment that there must be a specific set of individuals for (50b,d) to be licit can be seen as a requirement that these sentences be judged as categorical. See the discussion around (55-56).

(51) a. Hospitál kíchiñ ì. Xîni ntsi' ì enfermera.
hospital work.PRES I know.PRES all I nurse
'I work in a hospital. I know all the nurses.'

b. (*$\mathbf{Nà}_i$) kwĕ'e ní iin $nà_i$.
they polite.NEG not.even one them
'None of them is rude.'

(51a) sets up a context in which there is an identifiable set of nurses, while (51b) expresses that no member of that set is rude. Therefore, *ní iin nà* 'no one' in (51b) cannot pick out any identifiable referent, even though there is a contextually salient set. This derives why *ní iin nà* 'no one' cannot be targeted for topic agreement: topic agreement requires an identifiable referent, while *ní iin nà* 'no one,' by definition, may not pick one out.[13]

Furthermore, by analyzing pronoun doubling as topic agreement, we derive why local persons must always be doubled. Consider (52).

(52) a. *($\mathbf{Yù'u}_i$) nì ixutsya -k $ì_i$.
I PAST swim -INTR I
'I was swimming.'

b. *($\mathbf{Yô'o}_i$) ndààba -k $ú_i$.
you jump.PAST -INTR you
'You were jumping.'

c. (**Rà**) xákù $rà_i$.
he laugh.PRES he
'He is laughing.'

In (52a-b), we see that the local persons *yù'u* 'I' and *yô'o* 'you' must always be doubled. This has a natural explanation within our theory of topic agreement. As local persons are always identifiable within the Common Ground, they may always serve as topics. Non-local persons, such as *rà* 'he' in (52c), require additional contextual information

[13] Additionally, as Rizzi (1997) observes, 'no one' cross-linguistically makes a bad topic. This also explains why *ní iin nà* 'no one' cannot be targeted for topic agreement.

of the sort now under examination in order to be identifiable. This lets us understand why non-local persons need not always be doubled, in contrast with local persons.

The identifiability requirement also derives another effect related to partitives. To see this, first consider contrasts like those in (53).

(53) a. *? **Rí**$_i$ ntá'yi iin tsǐna$_i$, so xǐn ì ntsyâ rí kúu
it.AML cry.PRES one dog but know.NEG.PRES I which it.AML COP
rí.
it.AML
Intended: 'A dog is barking, but I don't know which one it is.'

b. # Ntá'yi iin tsǐna, so xǐn ì ntsyâ rí k'uu rí.

c. **Rí**$_i$ ntá'yi iin tsǐna$_i$, ra xín ì ntsyâ rí kúu rí.
it.AML cry.PRES one dog and know..PRES I which it.AML COP it.AML
'A dog is barking, and I know which one it is.'

d. *? ⊠ Ntá'jy iin tsǐna, ra xín ì ntsyâ rí kúu rí.

(53) presents a paradigm similar to (46). Here, we see that when a specific dog cannot be identified or introduced into the Common Ground, as in (53a-b), pronoun doubling is judged as severely degraded. In contrast, when the speaker can identify a specific dog to associate the information that it is barking to, as in (53c-d), pronoun doubling is required.

Interestingly, this requirement of identifiability disappears with partitive expressions. Consider (54).

(54) **Rí**$_i$ xínu iin míí tsǐna$_i$, so xǐn ì ntsyâ rí kúu
it.AML run.PRES one the dog but know.PRES.NEG I which it.AML COP.PRES
rí.
it.AML
'One of the dogs is running, but I don't know which dog it is.'

Context: We both know that I have dogs, and they have come up in our conversation. One day we're in my house and we hear something happening outside. I recognize that the noise is one of the dogs running, but I can't see them, so I

don't know which dog is running. You ask me "What's happening?" and I reply with (54).

In (54), we see that the context provides a salient set of dogs. So long as the speaker can identify the dog that is barking as a member of that set, even without any further specific knowledge, pronoun doubling is perfectly licit. This recalls the Partitive Constraint of Jackendoff (1977), which requires that partitive expressions be semantically definite, or in our language, identifiable. This inherent identifiability is clearly enough to license topic agreement.

Finally, the account in terms of topic agreement makes important predictions related to the thetic/categorical distinction, as first introduced to generative linguistics by Kuroda (1972) (see as well Ladusaw 1994 and Jäger 2001, among many others). Let us begin with categorical judgments.

Categorical judgments express a value relative to some identifiable referent. As Ladusaw (1994) phrases this, "The subject of a categorical judgment cannot be a nonspecific indefinite, its reference is "'presupposed"." (pg. 223) Therefore, we predict that utterances which clearly express categorical judgments should always host topic agreement. This turns out to be the case.

First, consider generics, which are always judged as categorical (Kuroda, 1972).

(55) a. * (**Rí**$_i$) xínu tsǐna$_i$.
it.AML run.PRES dog
'Dogs run.'

b. * (**Rí**$_i$) ntá'yi ntsintsiki$_i$.
it.AML cry.PRES cow
'Cows moo (lit. cry).'

In (55), we see that generic expressions, which are strongly categorical, require topic agreement. This makes sense both within the system of topics developed here, and within the theory of categorical expressions put forward by Kuroda (1972).

Additionally, Ladusaw (1994) observes that individual level predicates in the sense originally defined by Carlson (1977) and developed further by Diesing (1992) and Kratzer (1995)[14] should be incompatible with thetic judgments. Instead, individual level predicates determine categorical judgments. As expected based on the discussion above, we predict that this should mean that individual level predicates should require pronoun doubling. This turns out to be the case.

(56) a. * (**Rà**$_i$) sókǒn rà Juan$_i$.
he tall he J.
'Juan is tall.'

b. * (**Ñá**$_i$) doktóra ñá Maria.
she doctor she M.
'Maria is a doctor.'

In (56), we see that individual level predicates like *sókǒn* '(be) tall' and *doktóra* '(be a) doctor' require topic agreement, just like the generics in (55).

In both (55) and (56), it is clear that categorical judgments require pronoun doubling. Given the semantics of categorical judgments as presented in Kuroda (1972) and Ladusaw (1994), this is as predicted. The semantics of categorical judgments require that the predicate take an argument that is identifiable, "presupposed" to use the term from Ladusaw (1994), in the Common Ground. This requirement of identifiability is the same one that is imposed on topics in the system developed here. Therefore, we have a straightforward derivation of why categorical judgments require topic agreement.

Now let us consider thetic judgments. Thetic judgments refer to general states of affairs, without predicating the property of being in such a state of any specific or identifiable entity. In other words, thetic judgments are incompatible with topics in the sense appealed to here (see McNally 1998; Jäger 2001). Therefore, we predict that thetic judgments will be incompatible with topic agreement.

[14]Kratzer (1995) is often cited in unpublished manuscript form as Kratzer (1988) or Kratzer (1989).

This prediction proves to be true: thetic judgments are incompatible with topic agreement. Consider the paradigms in (57) and (58). Note that *tsihmyà* 'devil' is animal gender.

(57) a. ☒ Íyo iin tsyŏmí yúkŭn.
be.PRES.SG one owl forest
'There's an owl in the forest.'

Provided for Spanish *Hay un tecolote en el bosque.*

b. # **Rí**$_i$ íyo iin tsyŏmí$_i$ yúkŭn.
it.AML be.PRES.SG one owl forest
Intended: 'There's an owl in the forest.'

Judged good as meaning 'An owl is in the forest,' specifically as an answer to *Ntsyâchi íyo tsyŏmí?* 'Where is there an owl?'

(58) a. ☒ Kò tà'àn tsihmyà ndê.
NEG exist.PRES devil be.PRES.PL
'Devils don't exist (lit. it is not the case that there are devils).'

b. * **Rí**$_i$ kò tà'àn tsihmyà ndê.
it.AML NEG exist.PRES devil be.PRES.PL
Intended: 'Devils don't exist.'

Judgment: 'The *rí* [topic agreement] doesn't go here. It would sound better in something like [57c].'

c. **Rí**$_i$ ndê tsihmyà$_i$ yó'o.
it.AML be.PRES.PL devil here
Translated by speaker as *los diablos están aquí*, 'The devils are here.'

In (57b) and (58b), we see that topic agreement is incompatible with a thetic interpretation. When topic agreement is forced into a sentence that is meant to be judged as thetic, a categorical judgment is forced. This is evidenced by the speaker's judgment in (57b), and by the speaker providing the categorical (58c) when presented with (58b).

The incompatibility between topic agreement and the thetic judgment-type is exactly what we expect. This is because topic agreement requires an identifiable

referent, but the semantics of a thetic judgment is incompatible with the presence of such an identifiable referent. This conflict renders topic agreement unacceptable in the context of a thetic judgment, as we see in (57a) and (58a).

From this investigation, we see that appealing to topic agreement to derive pronoun doubling accounts for several important facts. First, topic agreement is a form of agreement, and as such, is expected to behave like AGREE in the sense of Chomsky (2000). This is indeed the case, as we saw in §5.1. Second, topic agreement, naturally, targets topics. As we have seen in this section, this identification yields an understanding of which nominals are doubled and which resist doubling. Therefore, I conclude that pronoun doubling is topic agreement.

5.3 Towards a formal treatment of topic agreement

In this chapter, we have established two key properties of pronoun doubling which should guide any attempt to construct a formal theory of the phenomenon. They are stated in (59).

(59) a. Pronoun doubling is derived through φ-agreement, which we understand as an interaction between a probe and a goal in the sense of Chomsky (2000, 2001).

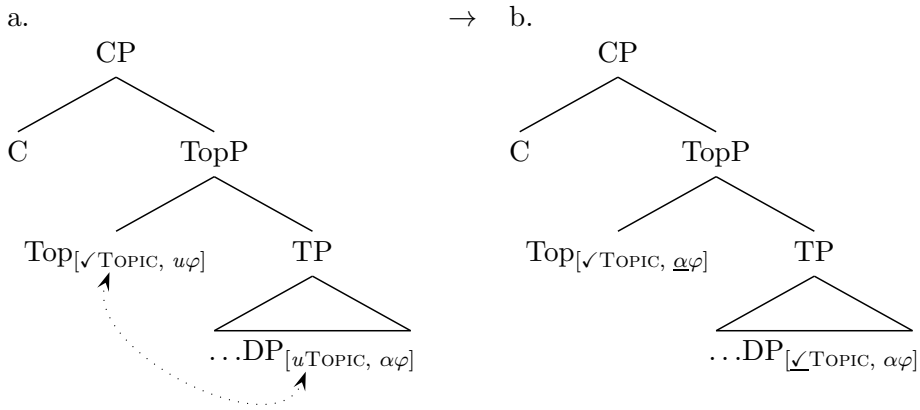
b. This particular φ-agreement process targets topics.

In essence, the insight so far is that pronoun doubling, despite all initial appearances, does not involve a true pronoun. While this relabeling represents a step forward, a true, formal theory of topic agreement remains to be stated.

The core proposal will be that topic agreement works in ways that are essentially like the traditional minimalist CASE system (Chomsky and Lasnik 1993; Chomsky 1995, 2001). The primary difference is that abstract CASE features are not implicated,

but rather information structure features like [Topic].[15] The basics of this theory are presented in (60), which is repeated from (7).

(60)



As in any theory of φ-agreement, two sorts of decisions needed to be made in (60). The first decision is the placement and the character of the probe. Likewise, we needed to decide the nature of the goal. Let us consider each of these in turn, beginning with the probe.

It is difficult to ascertain exactly where our probe will sit in this topic agreement system because of the prosodic conditions on the appearance of topic agreement, discussed in §1. Recall that these prosodic conditions essentially all derive from the fact that topic agreement markers are phonologically clitics, as is to be expected of functional elements like free-standing agreement markers (Selkirk, 1995). Specifically, if the leftmost prosodic phrase ϕ is binary, either because of a branching predicate or an initial adjunct that is a prosodic word or larger, topic agreement is prohibited. For instance, consider (61) and (62), which demonstrate these two options for blocking pronoun doubling.[16]

[15]Naturally, this approach requires that information structure features be present in the syntax. This has a strong precedent in the syntactic literature. For particularly influential minimalist literature, see Chomsky (1993, 1995), Rizzi (1997), López (2009), Aboh (2010), Mikkelsen (2015), Samko (2016). But this position is not without controversy. In particular, see Chomsky (2008), Horvath (2010), Fanselow and Lenertová (2011).

[16]Note that (61b) does have a grammatical parse. This is because *sóko* 'placenta' has a homophone

(61) a. *(**Ñà**$_i$) [sóko] yá$_i$.
it placenta it
'It's a placenta.'

b. (***Ñà**$_i$) [sóko ñá] yá$_i$.
it placenta her it
'It's her placenta.'

(62) a. *(**Yù'u**$_i$) [kárakono -k] ì.
I run.PRES -INTR I
'I am running.'

b. (***Yù'u**$_i$) [kamá kárakono -k] ì.
I quickly run.PRES -INTR I
'I am running quickly.'

In (61a), we have a nominal predicate *sóko* 'placenta.' Because the nominal predicate is not branching, the clitic topic agreement marker is permitted. Furthermore, the semantics of this sentence as a categorical judgment require topic agreement. In contrast, in (61b) we see that topic agreement is prohibited. The reasons for this prohibition cannot be semantic, since (61b) is a categorical judgment of equivalent status to (61a). Instead, the apparent inability of topic agreement to occur here is a reflection of prosodic factors. Here, the predicate *sóko ñá* 'her placenta' is binary. This binary predicate prevents topic agreement.

(62) demonstrates a similar pattern. In (62a), we see that topic agreement is obligatory. We already understand why, as local persons are always identifiable referents in the Common Ground and a categorical judgment is forced, making topic agreement necessary. But now from this perspective, it is something of a mystery why topic

sóko 'to carry something on one's shoulder.' This grammatical meaning is given as (i).

(1) Ñà$_i$ [sóko] ñá yá$_i$.
it carry.on.shoulder.PRES she it
'She is carrying it on her shoulder.'

Crucially, in (i) the predicate is not binary. This explains why topic agreement is permitted here.

agreement is prohibited in (62b). All that has changed is the addition of the adjunct *kamá* 'quickly.' I am aware of no reason, theoretical or empirical, why an adjunct should block AGREE. But from the perspective of prosody, we can understand this pattern: the prosodic unit *kamá kárakono -k* 'runs quickly,' forms a branching structure much like what we saw in (61b).[17]

Once we control for these prosodic concerns, we can see that our probe must occur between CP and TP. Consider (63).

(63) Á **rí**$_i$ nì xi'i tsǐna$_i$?
Q it.AML PAST die dog
'Did the dog die?'

In (63), we see that the topic agreement marker *rí*, which cross-references the animal *tsǐna* 'dog,' occurs immediately between the question marker *á* and the past tense marker *nì*.[18] Assuming that the question marker occupies C, we can propose that the

[17]These prosodic considerations are also responsible for the apparent root clause effect. For instance, topic agreement is generally marked in embedded clauses.

(1) ?? Ká'a rà Juan [**yù'u**$_i$ keba'a -k ì$_i$].
think.PRES he J. I win.IRR -INTR I
'Juan thinks that I will win.'

But if the embedded clause is fronted, or if a pause occurs before an in situ embedded clause, topic agreement greatly improves.

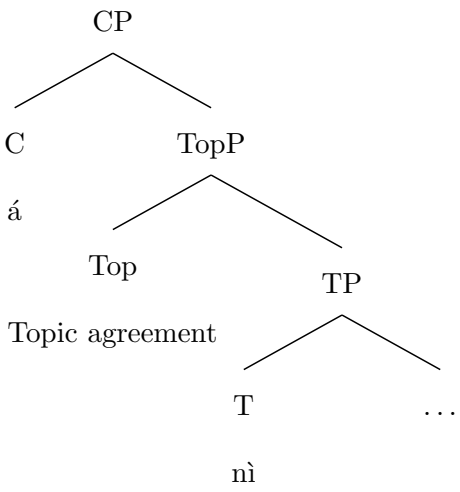
(2) a. [**Yù'u**$_i$ keba'a -k ì$_i$] ká'a rà Juan.
I win.IRR -INTR I think.PRES he J.
'Juan thinks that I will win.'
b. Ká'a rà Juan, # **yù'u**$_i$ keba'a -k ì$_i$.

[18]As mentioned in §1, it is not totally clear if these morphemes properly refer to tense, or rather to aspect, as in other Mixtec languages. Historically, the marker *nì* seems to be a cognate with completive aspect markers in other Mixtec languages. For instance, compare the verbal conjugation of the closely related Silacayoapan Mixtec in (ia) to the San Martín Peras forms in (ib).

(1) a. ka'an, ká'an, **na** ka'an
speak.POT speak.CON COM speak
'speak (potential aspect), speak (continuous aspect), speak (completive aspect)'
b. ka'an, ká'an, **nì** ka'an
speak.IRR speak.PRES PAST speak
'speak (irrealis), speak (present), speak (past)'

topic agreement probe, which I label 'Top,' will be contained with a head immediately below C in the extended projection of the clause and above T, as shown in (64).

(64)



The position of the probe in (64) is largely standard for theories that propose left peripheral topic projections, such as Rizzi (1997), Kiss (2002), and Aboh (2010). At the same time, we know that the topic projection in SMP Mixtec will be different than in other languages because Top in SMP Mixtec must be a φ-probe. Topic agreement in this language copies the φ-features of the topic into the topic probe. Consider the paradigm in (65) for gender agreement in the third person, in which topic agreement is bolded.

(65) a. **Rà**$_i$ kànakààba rà Juan$_i$.
he fall.PAST he J.
'Juan fell.'

b. **Ñá**$_i$ kànakààba ñá Maria$_i$.
she fall.PAST she M.
'Maria fell.'

Regardless of whether these morphemes turn out to be tense or aspect, the core features of this analysis hold, namely, that the probe will be selected for by C.

c. **Rí**$_i$ kànakàààba míí ntsibá'yi$_i$.
it.AML fall.PAST the coyote
'The coyote fell.'

d. **Tún**$_i$ kànakàààba míí tubi'intsyá$_i$.
it.WOOD fall.PAST the cactus
'The cactus fell.'

e. **Rá**$_i$ kànakàààba sahbì$_i$.
it.LIQ fall.PAST rain
'The rain fell.'

f. **Ñà**$_i$ kànakàààba míí yùhŭ$_i$.
it fall.PAST the rock
'The rock fell.'

SMP Mixtec has six grammatical genders for nominals in the third person: masculine humans, feminine humans, animal, wooden, liquid, and a catchall 'neutral' gender for everything that does not fit into one of the other categories. In (65), we see that topic agreement agrees in gender with the topic, no matter which grammatical gender it falls into.

In addition to gender, topic agreement agrees with person and number as well. We have seen this multiple times throughout, but (66) exemplifies this point again.

(66) a. **Yù'u**$_i$ kànakàààba -k ì$_i$.
I fall.PAST -INTR I
'I fell.'

b. **Yé**$_i$ kànakàààba -k yé$_i$.
we.INCL fall.PAST -INTR we.INCL
'We (INCL) fell.'

c. **Ndú**$_i$ kànakàààba ndú$_i$.
we.EXCL fall.PAST we.EXCL
'We (EXCL) fell.'

d. **Yô'o**$_i$ kànakàààba -k ú$_i$.
you fall.PAST -INTR you

'You fell.'

e. **Ndó**$_i$ kànakààba ndó$_i$.
you.PL fall.PAST you.PL
'You (PL) fell.'

f. **Ná**$_i$ kànakààba ná sí'i$_i$.
they.FEM fall.PAST they.FEM woman
'The women fell.'

(66a-e) demonstrates the paradigms for local persons in which we see that topic agreement implicates both number and person.[19] From this, we can conclude straightfor-

[19]For most speakers, there is no gender agreement in local persons, but some speakers accept and readily produce forms like (i) as possible alternatives for (66a) and (66d).

(1) a. Rà$_i$ kànakààba -k ì$_i$.
he fall.PAST -INTR I
'I (a man) fell.'
b. Ñá$_i$ kànakààba -k ì$_i$.
she fall.PAST -INTR I
'I (a woman) fell.'
c. Rà$_i$ kànakààba -k ú$_i$.
he fall.PAST -INTR you
'You (a man) fell.'
d. Ñá$_i$ kànakààba -k ú$_i$.
she fall.PAST -INTR you
'You (a woman) fell.'

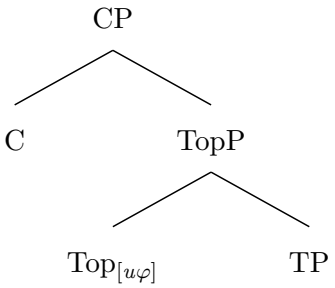
In (i), as opposed to (66a) and (66d), we see that topic agreement agrees for gender, but not person. For speakers that accept the pattern in (i), it is frequently produced, and often preferred.

This local gender agreement is only available for the singular persons. Contrast the grammatical sentences in (i) with the ungrammatical sentences in (ii).

(2) a. * Nà$_i$ kànakààba -k yé$_i$.
they fall.PAST -INTR we.INCL
Intended: 'We (INCL, mixed group) fell.'
b. * Ná$_i$ kànakààba -k yé$_i$.
they.FEM fall.PAST -INTR we.INCL
Intended: 'We (INCL, a group of women) fell.'
c. * Nà$_i$ kànakààba ndú$_i$.
they fall.PAST we.EXCL
Intended: 'We (EXCL, mixed group) fell.'
d. * Ná$_i$ kànakààba ndú$_i$.
they.FEM fall.PAST we.EXCL
Intended: 'We (EXCL, group of women) fell.'

wardly that the Top head in topic agreement constructions includes a φ-probe. This is shown in (67), which minimally updates (64).

(67)



By adding this φ-probe, though, we run into a problem when we return to sentences in which topic agreement is prohibited. For instance, let us consider again the contrast between thetic and categorical judgments from (57), reproduced in (68).

(68) a. ☒ Íyo iin tsyǒmí yúkǔn. (Thetic)
be.PRES.SG one owl forest
'There's an owl in the forest.'

b. **Rí**$_i$ íyo iin tsyǒmí$_i$ yúkǔn. (Categorical)
it.AML be.PRES.SG one owl forest
'An owl is in the forest.'

The two sentences in (68) are identical except for the absence and presence respectively of topic agreement. In (68a), we see that when topic agreement does not occur, the sentence is judged thetic, while in (68b), we see that with topic agreement, the sentence is judged categorical.

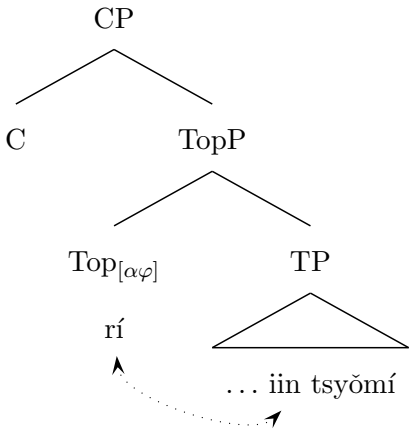
So far, our account can easily handle the categorical sentence in (68b). The analysis is as in (69).

---

e. * Nà$_i$ kànakààba ndó$_i$.
they fall.PAST you.PL
Intended: 'You (PL, mixed group) fell.'

f. * Ná$_i$ kànakààba ndó$_i$.
they.FEM fall.PAST you.PL
Intended: 'You (PL, group of women) fell.'

(69) Categorical judgment

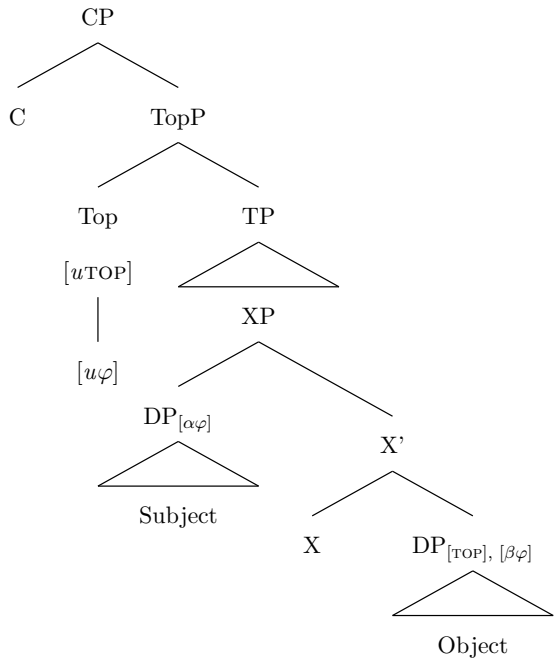


In (69), topic agreement functions like any other φ-agreement system. So far so good.

Something further must be said to account for (68a), though. Clearly, we cannot apply the analysis sketched (69) to (68a), because we need topic agreement, and by extension φ-agreement, not to occur in thetic judgments. In other words, we need some way of turning off φ-agreement when there is no accessible topic within TP.

There are, in principle, two paths we could take to prevent φ-agreement in thetic sentences. First, let us imagine that TopP is always present in every structure, whether a sentence topic may occurs or not. If we assume this, then we could say that our topic probe is structured in such a way that agreement with topics is forced, and φ-agreement is only possible if agreement with a topic has taken place. For instance, let us consider a structure such as (70).

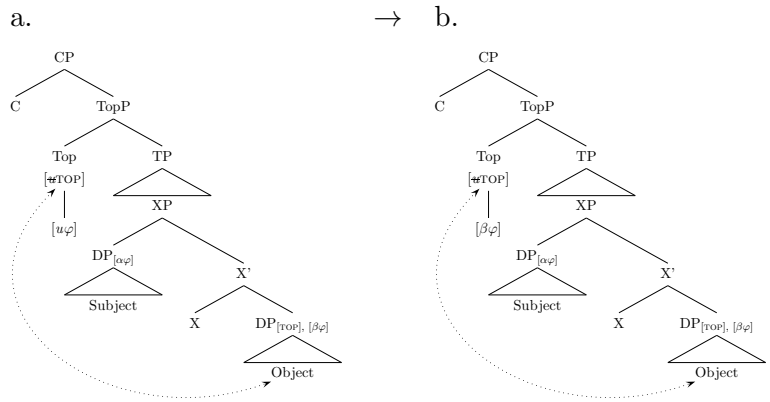
(70)



Several things are happening in (70). First, a DP, in this case the object, is syntactically marked as a topic in bearing a feature [Topic]. The assignment of this feature would, presumably, seem random from a predictive, computational perspective, but would have a logic from the conversational, pragmatic perspective. Second, our topic agreement probe is structured in a way similar to that found in work by Manetta (2006) and Martinović (2015). Here, the φ-probe is embedded under a probe that specifically targets topics, while I label [uTop]. This will require the Top-probe to probe first, making the φ-probe parasitic on the results of this earlier search.

Consider (71), which demonstrates the Agree relation from (70).

(71)



In (71), the object is marked [Topic]. Therefore, the [uTop] probe will enter into an Agree relation with the object. In this system, because the node Top enters into an Agree relation with a DP bearing φ-features, the φ-probe has automatic access to those φ-features. Therefore, if the [uTop] probe cannot find anything to enter into an Agree relation with, the [$u\varphi$] probe will be unable to receive a value.

Now, let us imagine what would happen if there were no nominals with the feature [Topic]. This would mean that our probe Top would finish the syntax without a value for its feature [uTop]. Presumably, the system would need to tolerate this, much as in the system developed by Preminger (2014). This would allow the unvalued [$u\varphi$] and [uTop] to, then, simply be deleted, as is standardly thought to be necessary (Chomsky 2000, 2001, and Pesetsky and Torrego 2007).

That said, I find this analysis unappealing for two reasons. First, it requires several moving pieces to work, from the feature [Topic] that occurs on DPs, to the structured probes and the deletion of unvalued features. While each of these has independent application in the literature, a more appealing analysis would rely less on technical elaboration. Second, there is something conceptually unappealing about this line of analysis. In (71), topicality does not originate on the Topic head. Rather, in this system, Top is nothing more than an Agr head, providing no apparent content for the interfaces. Therefore, in a real way, (71) is profoundly incompatible with minimalist

ideals. See Chomsky (2000, 2001). While this, *a priori*, can hardly count as an argument against the proposal, it does make one wonder if it would be possible to do better, in appealing to a more spartan descriptive technology.

With this desideratum in mind, let us explore an alternative solution to the problem of preventing φ-agreement in a clause where no topic may occur. Let us say that the TopP projection simply does not occur when there is no topic. In other words, C would have the two subcategorization frames in (72).

(72) a. C, [__ TopP]

Yields a categorical judgment

b. C, [__ TP]

Yields a thetic judgment

If our topic agreement probe simply were not present in thetic judgments, or in other instances when no topics are possible, then we would have no problem. Such an approach has a certain appeal.

At the same time, though, we run the risk of facing a severe Look Ahead problem, because we need to ensure that TopP will be selected only in structures which include sentence topics. We can avoid this potential issue if we incorporate another empirical insight: sentence topics must be targeted for agreement. To see this, let us return to paradigms like (73) and (74).

(73) a. Nă xá'antsya rà Juan?
what cut.PRES he J.
'What is Juan cutting.'

b. *(**Rà**$_i$) xá'antsya [rà Juan$_i$]$_{Top}$ [chìkí]$_{Foc}$.
he cut.PRES he J. tuna
'Juan is cutting *tunas*.'

(74) a. Yó nà xá'antsya chìkí?
who they cut.PRES tuna

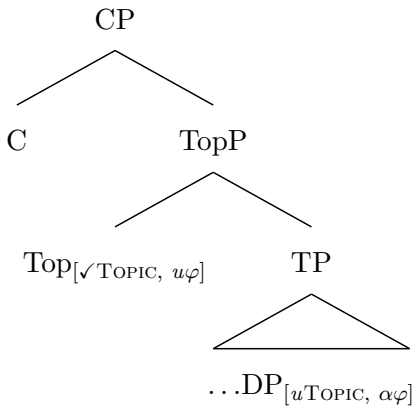
'Who is cutting the tunas?'

b. *(**Rí**$_i$) xá'antsya [rà Juan]$_{Foc}$ [rí chìkí$_i$]$_{Top}$.
it.AML cut.PRES he J. it.AML tuna
'Juan is cutting *tunas*.'

(73) and (74) present a pattern that is familiar by now: φ-agreement in this language targets topics. At the same time, the (b) sentences highlight another important facet of this pattern: topics must be targeted for agreement, as failure to agree with the topic results in ungrammaticality. With this, it seems that topic agreement and topics exist in a symbiotic relation: sentence topics need to be agreed with, and topic agreement can only target sentence topics. Therefore, an adequate theory should capture the symbiotic character of this relationship.

To accomplish this goal, as well as to avoid the problems raised against the other potential analyses, let us give more content both to our Top probe as well as to the topics themselves. Consider the configuration in (75).

(75)



In (75), we have a set up in which both the probe and the goal contribute to the Agree relation they will ultimately share. Starting with the Top head, we see that it bears a feature which I call "[✓Topic]." This feature functions as a signal to the interpretive interface that the sentence within its scope contains a topic. We might take it to be a

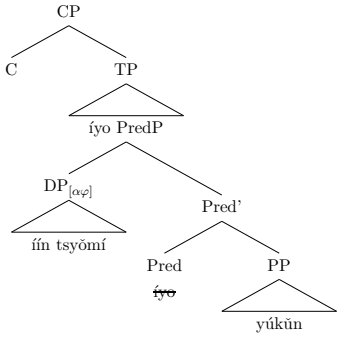
correlate of the fact that topichood is as much a property of structures as it is nominals, implying that it bifurcates the clausal material into two fields: topic and focus, theme and rheme, and so on. There is now no Look Ahead problem, in the sense that the interpretive properties of the topic head are now compatible only with sentences that do, indeed, contain a topic.

We should also consider the featural composition of the goal. As a DP, it will bear a set of valued φ-features. Additionally, I propose that it enters the derivation with an unvalued [*u*Topic] feature. This will require that it stand in an Agree relation with a Top head in order to value its topic feature. This is, broadly, similar to Rizzi's (1997) Criterial approach, although cached out formally in an Agree relation, rather than in a uniform specifier-head relation. The presence of the two complementary feature sets will ensure that the Top head and the sentence topic will complete each other, intuitively capturing the pattern in (73-74).

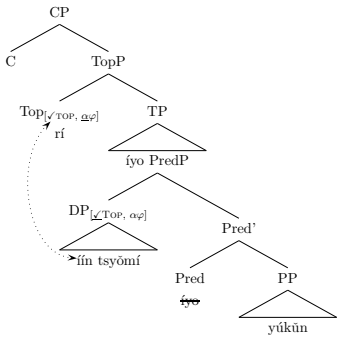
To see how this will work, we will review several derivations. First, let us return to our contrast between thetic and categorical sentences from (68). Based on the discussion above, particularly from (72), we can model the difference between these two types of sentences as in (76). (76a) provides the proposed structure for the thetic judgment in (68a), while (76b) shows the structure for the categorical judgment in (68b). Note that verb movement of *íyo* 'be.pres.sg' is not shown in either structure.

(76)

a. Thetic judgment



b. Categorical judgment

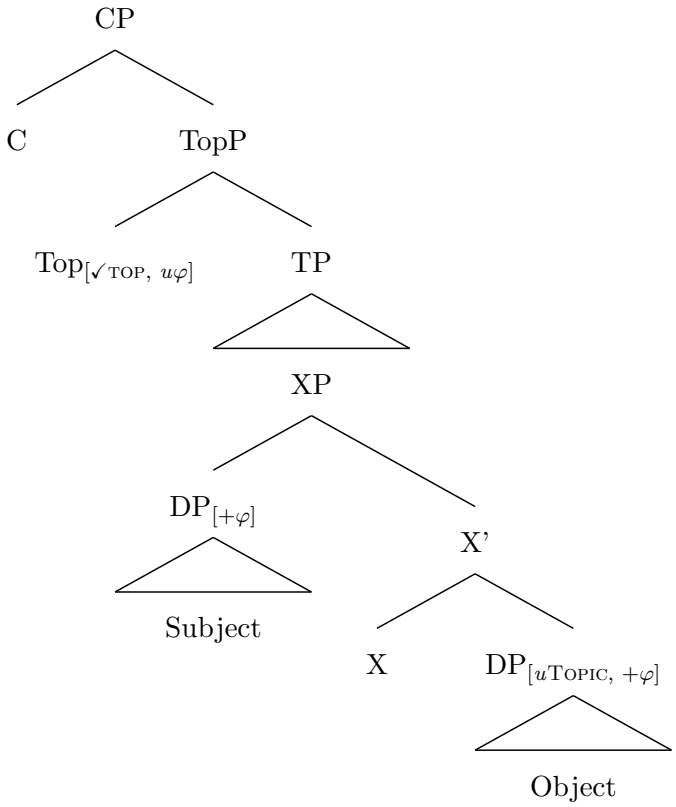


For the thetic judgment in (76a), we see that no Top projection is present. The absence of the topic agreement probe derives why no φ-agreement takes place. Likewise, the lack of a Topic projection in a thetic judgment makes sense, as thetic judgments are incompatible with sentence topics (Jäger 2001).

Categorical judgments, on the other hand, require predication over an identifiable referent in the discourse context. In other words, categorical judgments require a topic. Therefore, the Top projection must appear in the structure in (76b) in order to license the nominal which will serve as topic. But the topic probe brings with it a φ-probe, and we therefore understand why topic agreement must occur in categorical judgments.

The second core aspect of topic agreement that we must be sure our analysis can handle is the mandatory object agreement that occurs when the object is a topic. This is the sort of pattern we saw in (74b). The trick here is to have our φ-probe ignore the φ-features of the subject in favor of the φ-features of the topical object. Setting aside the exact structure of the lower portion of the clause, the type of configuration we must account for is given in (77).

(77)



In (77), the topical object is commanded by the subject, but both DPs have a set of φ-features that could value the topic probe's unvalued $[u\varphi]$. The challenge, of course, is to explain why the topic probe can bypass the apparently more accessible φ-features of the subject in order to agree with the object topic.

There are various approaches one might take to this problem, but the driving intuition must sure be that the object is favored in such circumstances because it, and it alone, bears the topic feature which matches that of the higher probe. I will therefore tentatively assume (78).

(78) A probe π shares feature-values with the most prominent syntactic object whose label maximally matches the features of π.

In the case at hand in (77), it is the object whose featural composition maximally matches the topic probe, since along with φ-features it has an unvalued topic

feature, which the subject crucially lacks. Since we assume, with Chomsky (2000, 2001), that the presence of unvalued features results in ungrammaticality, an AGREE relation is forced and topic agreement is obligatory.[20]

In this way, we see that agreement with the object is possible so long as the object is to be interpreted as a topic. This is clearly the right result, based on the empirical terrain we have examined so far.

In this section we have seen that topic agreement can be analyzed using the standard tools of Chomskyian AGREE, so long as sentence topics enter the syntactic derivation bearing a feature which can receive a value only from the topic agreement probe. In this way, topic agreement functions formally in a way that is very close to the traditional minimalist understanding of Case licensing (Chomsky and Lasnik 1993 and Chomsky 1995), with the crucial difference that Case in an inherently uninterpretable feature while topic is, self-evidently, interpretable. The fact that agreement is ultimately realized in the form of the head which hosts the probe can be seen as one reflection of the general head-marking character of the language (in the sense of Nichols 1986 and much subsequent typological work).

5.4 Concluding and reflecting on topic agreement

In this chapter, we have considered the possibility that pronoun doubling, despite all initial appearances to the contrary, does not involve a pronoun, but is best analyzed as a form of φ-agreement. We saw that pronoun doubling has several properties that suggest a φ-agreement analysis. In particular, §5.1.1 demonstrated that pronoun doubling is phase bounded, but lacks the signature properties of both A-movement and

[20] Alternatively, we could say that the probe will Interact with both the subject and the object, in the sense of Hiraiwa (2002, 2005), Béjar (2003), Rezac (2003, 2004), Chomsky (2008), Béjar and Rezac (2009), Nevins (2011), Preminger (2014), and Deal (2015b), but the unvalued feature on the object will force ungrammaticality unless it ultimately serves as the goal. I am not sure how to empirically disambiguate between these two hypotheses, and as such I offer the more minimal solution in (78).

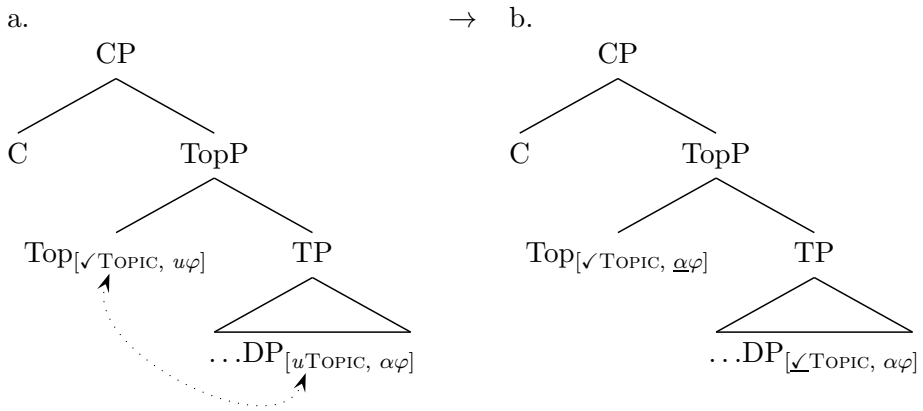
A'-movement. These are the properties we expect of a probe-goal interaction without concomitant raising (exactly the analysis developed in this chapter).

At the same time, to analyze pronoun doubling as a form of φ-agreement presents its own problems. For instance, in §5.1.5, we considered two potentially problematic features of pronoun doubling that present difficult in analyzing this process as φ-agreement, such as this processes restriction to specific, contextually salient referents, and its ability to violate usual locality restrictions on AGREE. In §5.2, we saw that both of these apparent difficulties can be overcome by recognizing that φ-agreement in this language targets topics.

From this perspective, the semantic restrictions on the possible targets of φ-agreement in this language reflect semantic and pragmatic restrictions on topics cross-linguistically. Likewise, because this process targets topics, the supposed violations of locality can be understood.

In §5.3, we developed a theory of topic agreement. This is shown in (79), which is repeated from (7).

(79)



In the system developed in §5.3 and shown in (79), topic agreement functions formally like a traditional minimalist Case system in the sense of Chomsky and Lasnik (1993) and Chomsky (1995): both the probe and the goal need a feature value from the other, and

so the two pieces complete each other by establishing an AGREE relationship. The only formal difference between (80) and a traditional minimalist Case system is the features involved. While both systems involve φ-agreement, traditional minimalist Case systems use abstract Case features where (80) uses an information structural feature [TOPIC].

An important property of the general approach is that it aligns very closely with the work of Miyagawa (2010, 2017), who observes that features such as Case may have analogs in other languages in information structural notions such as topic and focus. Our discussion of SMP Mixtec fills in an expected typological gap in which an information structural feature like topic serves the same purpose as Case in more familiar languages.

Miyagawa calls this analog to φ-agreement "discourse-configurational agreement," and identifies a new set of features parallel with φ-features, which he calls δ-features, or "discourse" features. δ-features include information structural notions like topic and focus. Clearly, the system developed here is in this spirit and fits naturally into the typological landscape that such a theory would lead us to expect to find. In this general framework of understanding, the bundling of φ-features and a topic feature into a single probe is expected and natural, rather than anomalous.

Finally, SMP Mixtec topic agreement may well offer useful insights into the theory of topichood itself since, if the analysis offered here is on the right track, it provides an independent and reliable diagnostic for what is or is not a topic. To take one example where such a diagnostic may be useful, there has been some debate in the literature as to whether non-referential, universally quantified generics like "All triangles have a sum of 180 degrees" behave as thetic or cagetorical judgments. Kuroda (1972), for example, argues that such sentences are thetic in German but categorical in Japanese. In SMP Mixtec, as diagnosed by the unavailability of topic agreement, these sentences seem to be thetic. (80) is repeated from (48).

(80) Kŭntsi'ì iñ ú, se'e. (*$\mathbf{Nà}_i$) sîsò xitsin ntsi'i $nà_i$.
worry.NEG.IRR inside you child they boil.PRES nose every they
'Don't worry, child (*no te preocupes, mijo*), everyone snores (lit. boils nose).'
Context: A child is embarrassed because they found out they snore, and you are comforting them.

In (80), we have a non-referential, universally quantified generic that patterns identically to the thetic judgments discussed above. Therefore, it seems that languages can differ in the semantics/pragmatics they assign to these sorts of sentences, with some languages utilizing a semantics that requires a categorical judgment, like Japanese, while other languages employ semantic representations that yield thetic judgments, like German and SMP Mixtec. Presumably, these sentences mean comparable things cross-linguistically, which indicates that multiple possible semantic representations must be available to yield the meaning of a non-referential, universally quantified generic. Determining what exactly these converging semantic representations are must be left to future research.

Second, SMP Mixtec topic agreement might suggest that more information is stored in the discourse context than one might have imagined. For instance, many weather and environmental predicates in SMP Mixtec require an overt pronominal subject *yá* 'it,' much like English. This "ambient *yá*," to adapt the terminology of Bolinger (1973), is obligatorily targeted by topic agreement.

(81) a. (*$\mathbf{\tilde{N}à}_i$) nì tahan $yá_i$.
it PAST quake it
'There was an earthquake.'

b. (*$\mathbf{\tilde{N}à}_i$) xa kùnaha $yá_i$.
it PERF become.night it
'It became night.'

c. (*$\mathbf{\tilde{N}à}_i$) xa kùtswíín $yá_i$.
it PERF darken it
'It got dark.'

Above, we saw that nominals must refer to an identifiable referent in the discourse context in order to be targeted by topic agreement. Therefore, the obligatory topic agreement with ambient-*yá* in (82) strongly suggests that the ambient environment of a discourse can be construed as an indentifiable discourse reference, on par with other referents.

Finally, we should consider issues of morphology and exponence. In SMP Mixtec, topic agreement morphemes are morphophonologically identical to a regular pronoun. This is, of course, what lead me to name topic agreement "pronoun doubling" in the first place. Consider the non-local series.

(82)

	Topic agreement			Pronoun	
a.	rà [ɾà]	3.MASC	a'.	rà [ɾà]	'he' ∼ 'they (all male group)'
b.	ñá [ɲá]	3SG.FEM	b'.	ñá [ɲá]	'she'
c.	ná [ná]	3PL.FEM	c'.	ná [ná]	'they (all female)'
d.	rí [ɾí]	3.ANIMAL	d'.	rí [ɾí]	'it (an animal)' ∼ 'they (animals)'
e.	túún [tũ]	3.WOOD	e'.	túún [tũ]	'it (wooden things)' ∼ 'they (wooden things)'
f.	rá [ɾá]	3.LIQUID	f'.	rá [ɾá]	'it (a liquid)' ∼ 'they (liquids)'
g.	ñà [ñà]	3SG	g'.	ñà [ñà]	'it'
h.	nà [nà]	3PL	h'.	nà [nà]	'they'

Going back at least to Uriagereka (1995), who credits early unpublished work by Esther Torrego, the sort of morphophonological identity in (82) is considered a diagnostic for clitic doubling. From this perspective, it is clear why so much ink in this dissertation was devoted to establishing that pronoun doubling should not be unified with superficially similar and well-studied cases of clitic doubling. Simply put, pronoun doubling looks like clitic doubling.

Such considerations might give rise to skepticism about the general approach developed here. However, much recent work has challenged any easy inference from morphophonological form to syntactic analysis. As discussed in §2.2.4, one particularly

clear example comes from Yuan (2017a,b), who presents a careful, cross-dialectal study of Inuktitut dialects. In this work, she demonstrates that a morphophonologically stable set of affixes in some dialects are best analyzed as φ-agreement, and the same affixes in other dialects are best analyzed as clitic doubling. Compare the Inuktitut in (83), which Yuan demonstrates is derived through clitic doubling, to the same morphemes in Kalaallisut, shown to be derived through φ-agreement. Note that the alternation between $j \sim v$ is a regular sound correspondence in these two dialects.

(83) Inuktitut (clitic doubling)

a. **-jara** '1S.S/3S.O/DECL'

b. **-jait** '2S.S/3S.O/DECL'

c. **-jarma** '2S.S/1S.O/DECL'

(84) Kalaallisut (φ-agreement)

a. **-vara** '1S.S/3S.O/DECL'

b. **-vait** '2S.S/3S.O/DECL'

c. **-varma** '2S.S/1S.O/DECL'

In comparing the morphemes in (83-84), we see that, apart from regular sound correspondences, the two sets are identical, even though they are demonstrably derived by different means (Yuan 2017a,b). There is now every reason to be skeptical, then, that morphophonological diagnostics can be taken to be reliable when attempting to distinguish between φ-agreement and clitic doubling.

If the analysis developed here is roughly correct, then SMP Mixtec is a similar case. We have tried to establish that being guided purely by the morphophonology leads us astray. Rather, we needed a careful examination of the syntactic and interpretive properties of this process to see that pronoun doubling (a term which, at this point seems inappropriate) to establish that, despite appearances, it does not actually seem to involve a pronoun at all. Rather, we are dealing with morphophonological syncretism between the Top head and pronouns. We can model this easily with Vocabulary Entries such as those in (85).

(85) $[\alpha\varphi] \Leftrightarrow$ rà, ñá, rí, ná...

(85) captures identity of form between pronouns and topic agreement morphemes seen in (83) by providing Vocabulary Entries in which category is irrelevant. Instead, (85) demonstrates a class of Vocabulary Entries that simply link phonological exponents with sets of φ-features, regardless of whether those features appear on Top or on D. If the general line of analysis is correct, we must agree with Yuan that the morphophonology of exponence cannot be taken as a reliable diagnostic for syntactic derivation.

Naturally, further puzzles remain. To conclude this dissertation, I will demonstrate two particularly interesting open phenomena within pronoun doubling constructions that do not receive an obvious treatment in the analysis presented here.

Given that topics in SMP Mixtec have been demonstrated to be obligatorily specific, identifiable, and referential, we might now expect something of a "definiteness" hierarchy, with more definite DPs preferentially being targeted over less definite DPs. Consider the hierarchy from Farkas and Brasoveanu (2013), developed from Aissen (2003).

(86) Personal pronoun > Proper name > Definite DP > Specific indefinite > Non-specific indefinite

Interestingly, we do indeed see the effects of this hierarchy, but the expected accessibility is reversed. The agreement accessibility hierarchy in SMP Mixtec is shown in (87).

(87) Specific indefinite > Definite DP > Proper name > Personal pronoun

(87) is entirely the same hierarchy as (86), although the predicted accesibility is reversed. In a structure which includes any two of the elements in (87), topic agreement obligatorily targets the higher of the two on the hierarchy. This is shown in (88-90), in which, in each sentence, the object outranks the subject.[21]

[21]Identical data were found if the subject outranks the object, although in those cases subject agree-

(88) a. **Rí$_i$** xa xìn ì iin ntsibá'yi$_i$. Kă'nu kwê'e rí.
it.AML PERF see.PAST I a coyote big very it
'I saw a coyote. It was very big.' (Specific indefinite > personal pronoun)

b. * **Yù'u**$_i$ xa xìn ì$_i$ iin ntsibá'yi. Kă'nu kwê'e rí.

c. **Rí**$_i$ xa xìn ì míí ntsibá'yi$_i$.
it.AML PERF see.PAST I the coyote
'I saw the coyote.' (Definite DP > personal pronoun)

d. * **Yù'u**$_i$ xa xìn ì$_i$ míí ntsibá'yi.

e. **Rà**$_i$ xa xìn ì rà Juan$_i$.
he PERF see.PAST I he J.
'I saw Juan.' (Proper name > personal pronoun)

f. * **Yù'u**$_i$ xa xìn ì$_i$ rà Juan.

(89) a. **Rì$_i$** xa xìni rà Juan iin ntsibá'yi$_i$. Kă'nu kwê'e rí.
it.AML PERF see.PAST he J. a coyote. big very it.AML
'Juan saw a coyote. It was very big.' (Specific indefinite > proper name)

b. * **Rà$_i$** xa xìni rà Juan$_i$ iin ntsibá'yi. Kă'nu kwê'e rí.

c. **Rí**$_i$ xa xìni míí ñá sí'i iin chkwíín$_i$. Kă'nu kwê'e rí.
it.AML PERF see.PAST the she woman a fox big very it.AML
'The woman saw a fox. It was very big.' (Specific indefinite > definite DP)

d. * **Ñá**$_i$ xa xìni míí ñá sí'i$_i$ iin chkwíín. Kă'nu kwê'e rí.

e. **Rí**$_i$ xa xìni rà Juan iin chkwíín$_i$. Kă'nu kwê'e rí.
it.AML PERF see.PAST he J. a fox big very it.AML

---

ment becomes obligatory. The full paradigm is not shown for reasons of space, but the critical examples are shown in (i).

(1) a. **Rí**$_i$ xa xìni iin tsĭna$_i$ yù'u. Kă'nu kwê'e rí.
it.AML PERF see.PAST a dog me big very it.AML
'A dog saw me. It was very big.'

b. * **Yù'u**$_i$ xa xìni iin tsĭna yù'u$_i$. Kă'nu kwê'e rí.

c. **Rí**$_i$ xa xìni míí tsĭna$_i$ rà Juan.
it.AML PERF see.PAST the dog he J.
'The dog saw Juan.'

d. * **Rà**$_i$ xa xìni míí tsĭna rà Juan$_i$.

'Juan saw a coyote. It was very big.' (Specific indefinite > proper name)

f. * **Rà**$_i$ xa xìni rà Juan$_i$ iin chkwíín. Kă'nu kwê'e rí.

(90) a. **Rí**$_i$ xa xìni rà Juan míí tsĭna$_i$.
it.AML PERF see.PAST he J. the dog
'Juan saw the dog.' (Definite DP > proper name)

b. * **Rà**$_i$ xa xìni rà Juan$_i$ míí tsĭna.

This apparent reversed hierarchy is surprising from the perspective of topic agreement, as we would predict personal pronouns to make better topics than specific indefinites because personal pronouns, particularly the local person pronoun *ì* 'I' is the most identifiable referent in the Common Ground. Despite this, personal pronouns are outranked by everything else on the hierarchy. Likewise, specific indefinites outrank definite DPs and proper names as well, and definite DPs also outrank proper names. This is surprising from a semantic perspective, but the judgments in (88-90) are robust.

Finally, recall that pronoun doubling is obligatory for second person subjects in intransitive clauses.

(91) * (Yô'o$_i$) tsí'bĭ -k ú$_i$.
you sweep.PRES -INTR you
'You are sweeping.'

Despite this, second person transitive subjects, unlike the first person transitive subjects in (88), block the object to be targeted for topic agreement, and likewise cannot be targeted themselves. This is true no matter where the object sits on the hierarchy in (87). This is demonstrated in (92) with a specific indefinite object.

(92) a. ☒ Xa xìñ ú iin ntsibá'yi. Kă'nu kwê'e rí.
PERF see.PAST you a coyote big very it
'You saw a coyote. It was very big.'

b. * **Rí**$_i$ xa xìñ ú iin ntsibá'yi$_i$. Kă'nu kwê'e rí.

c. * **Yô'o**$_i$ xa xìñ ú$_i$ iin ntsibá'yi. Kă'nu kwê'e rí.

Interestingly, this second person blocking effect only holds if the second person pronoun is in subject position. If the second person pronoun is in object position, then the subject may be targeted, but the object is still unavailable. This may be a result of the low position of personal pronouns on the hierarchy in (87). Regardless, these final data are shown in (93).

(93) **Rí**$_i$ xa xìni iin ntsibá'yi$_i$ yô'o. Kă'nu kwê'e rí.
it.AML PERF see.PAST a coyote you big very it
'A coyote saw you. It was very big.'

Clearly, there are still many more mysteries remaining to be solved within SMP Mixtec topic agreement, but I hope this dissertation has opened the path for future work on this important language.

Appendix A

Appendix: The syntax of the pseudo-relative

In this appendix, I will sketch the syntax of pseudo-relatives. These were introduced in §4.3 as SMP Mixtec's repair for external argument wh-extraction. (1), repeated from §4.3, demonstrates that external arguments may not undergo wh-movement, while (2) demonstrates the pseudo-relative construction.

(1) a. * Yó$_i$ kàni __ rà Julio?
who hit.PAST he J.
Intended: 'Who hit Julio?'

b. * Yó$_i$ káchûn __$_i$?
who work.PRES
Intended: 'Who is working?'

c. * Ntsyâ míí nà tsyahá yó'o ba'á kíchiñu __?
what the they man this well work.PRES
Intended: 'Which of these men is working well?'

(2) a. [Yó (kúú) nà] [kàni __ rà Julio]?
who COP.PRES they hit.PAST he J.
'Who are they that hit Julio?'

b. [Yó (kúú) nà] [káchûn __]?
who COP.PRES they work.PRES
'Who are they that are working?'

c. [Ntsyâ nà (kúú) míí nà tsyahá yó'o] [ba'á kíchiñu __]?
which they COP.PRES the they man this well work.PRES
'Which of them are these men that are working well?'

(2) demonstrates the pseudo-relative construction under investigation here. I will refer to the bracketed portion of such sentences, like *yó kúú nà* 'who are they' in (2a) as the cleft, with the remainder of the sentence, such as *kàni* __ *rà Julio*, roughly, 'that hit Julio,' as the pseudo-relative. The primary goal of this appendix is to provide the evidence that I have gathered that indicates that the gap inside the pseudo-relative is not derived by A'-movement.

Before investigating the syntax of these constructions further, it is worth noting that the pseudo-relative construction obligatory for external argument wh-dependencies, demonstrated in (2) above, is generally available for internal argument wh-constructions as well. In particular, as we will see here, the pseudo-relative construction provides many of the repairs for the ungrammatical internal argument wh-constructions. That said, there are restrictions on when a pseudo-relative can be used for internal argument wh-constructions.

First, pseudo-relatives are freely available for unaccusative subject wh-dependencies.

(3) a. Yó$_i$ ínchichi __ xihĭn be'e?
who stand.PRES.SG beside house
'Who is standing beside the house?'

b. Yó kúú nà [ínchichi __ xihĭn be'e]?
who COP.PRES they stand.PRES.SG beside house
'Who are they that are standing beside the house?'

c. Nă xá'anu __ yé'è?
what grow.PRES garden
'What is growing in the garden?'

d. Nă kúú yá [xá'anu __ yé'è]?
what COP.PRES it grow.PRES garden
'What is it that is growing in the garden?'

The sentences in (3) seem to be in free variation with the wh-movement strategy shown in §4.3. But it is not generally the case that all internal argument wh-constructions can use a pseudo-relative. For instance, consider the transitive pairs in (4).

(4) a. Nă chì'i Julia __?
what sow.PAST J.
'What did Julia sow?'

b. Nă (kúú) yá [chì'i Julia __]?
what COP.PRES it sow.PAST J.
'What is it that Julia sowed?'

c. Yó kàni rà Julio __?
who hit.PAST he J.
'Who did Julio hit?'

d. * Yó (kúú) nà [kàni rà Julio __]?
who COP.PRES they hit.PAST he J.
Intended: 'Who did Julio hit?,' Good as: 'Who hit Julio?'

In (4a-b), we see that the internal argument wh-dependency can involve either wh-movement, as in (84a), or the pseudo-relative construction, as in (84b). Again, in this case, the two options seem to be in free variation. In contrast, in (84c-d), we see that only wh-movement of the internal argument is available. While the string in (3d) has a grammatical parse in which the wh-dependency targets the external argument within the pseudo-relative, it may not have a reading in which the pseudo-relative gap is in internal argument position.

It seems to be the case that the difference is related to recoverability of the gap. In (3), there is only one argument of the predicate, and as such, there is only one position to which the gap can be interpreted. In (4b), the only plausible site to interpret the gap is in internal argument position. This is because the alternative reading, 'What

sowed Julia?' in which the gap is in external argument position, is entirely implausible, as human beings obviously cannot be sown. In contrast, in (4d), it is plausible that the gap could be in either external or internal argument position, as it is possible that either Julio hit someone else, or someone else hit Julio. When this type of potential ambiguity arises, using a pseudo-relative in internal argument wh-constructions becomes unavailable, presumably to avoid this ambiguity.

First in §A.1 I provide a brief overview of the syntax of the cleft. In §A.2, I provide a slew of evidence that suggests that the gap inside the pseudo-relative is not derived by A'-movement. To preview this discussion, the bulk of this evidence comes from pseudo-relatives being this language's repair for ill-formed A'-dependencies. These include Crossover effects in §A.2.1 and intervention effects in §A.2.2. With that said, §A.3 presents the evidence that the pseudo-relative gap is derived by movement. Interestingly, though, this movement seems to be a mix of A- and A'-movement. In §A.3.1, I show that pseudo-relative gap dependencies are unbounded and island-sensitive, like the A'-dependencies in §4 but unlike the A-dependency in §3. But in §A.3.2, I show that the pseudo-relative gap does not behave like an A'-gap in that is correspond to a possessor or the object of a preposition. This is like the A-gap in §3, but unlike A'-gaps.

Before getting into the weeds of the syntax of the pseudo-relative, some comments about the constituency in (2) are in order. Specifically, how do we know that things are as in (2), and not, say, as in (5)?

(5) [Yó kúú] [nà$_i$ xà'antsya $\_\__i$ ítǔn]?
who COP.PRES they cut.PAST tree
'Who are they who cut the tree?'

In (5), what appears to be the subject of the cleft, in this case *nà* 'they,' forms a constituent with what I have referred to as the pseudo-relative. If the nominal forms a constituent with the pseudo-relative, then we, perhaps, have a straightforward account of the gap, as we would have a nominal that is interpreted in gap within the local

constituent.

There are two reasons to think that the constituency in (5) is incorrect. The first comes from SMP Mixtec's equivalent of sluicing, shown in (6).

(6) a. Iin nà xà'antsya míí ítŭn, so xĭn ì [yó kúú nà
one them cut.PAST the tree but know.NEG.PRES I who COP.PRES they
].

'Someone cut down the tree, but I don't know who it was.'

b. * Íín nà xa'antsya míí ítŭn, so xĭn ì [yó (kúú)].

c. Iin ña'à nàkààba, so xĭn ì [nă kíí yá].
one thing fall.PAST but know.NEG.PRES I what COP.PRES it
'Something fell, but I don't know what it is.'

d. * Íín ña'à nàkààba, so xĭn ì [nă (kíí)].

In (6), we see that SMP Mixtec uses these same wh-clefts in scenarios where other languages would use a sluice. Inside of these wh-clefts, we see that the subject is obligatory. I take this as evidence that this unit is a constituent in this language. Therefore, when we return to comparing (2) to (5), the null hypothesis should be that this same constituent is present.

Second, there is good evidence that the subject of the wh-cleft does not originate within the pseudo-relative, as (5) predicts. This evidence comes from the lack of morphological connectivity effects between the wh-cleft subject and the gap inside of the pseudo-relative. To see this, consider the predicates in (7).

(7) a. Sata be'e **ínchichi** ítŭn.
beside house stand.PRES.SG tree
'The tree is beside the house.'

b. Sata be'e **ita** ítŭn.
beside house stand.PRES.PL tree
'The trees are beside the house.'

c. Iin rà tsyahá nì **kìta**.
one he man PAST enter.SG
'One man entered.'

d. Kwa'ă kwê' á yìbi nì **kèhè**.
many very it people PAST enter.PL
'A lot of people entered.'

e. **Chìnuû** rà Juan míí líbro nŭ mésa.
put.PAST.SG he J. the book on table
'Juan put the book on the table.'

f. **Kòsò** rà Juan míí líbro nŭ mésa.
put.PAST.PL he J. the book on table
'Juan put the books on the table.'

(7) demonstrates a number of predicates that exhibit a pattern commonly referred to as participant number suppletion. Participant number suppletion is a common form of suppletion, particularly in Mesoamerica, in which the subject of an intransitive or the object of a transitive trigger verbal suppletion if either is plural (Veselinova 2006, Toosarvandani 2015, Bobaljik and Harley 2017, inter alia). In SMP Mixtec, this class of predicates contains 11 members, and is largely restricted semantically to predicates of motion, location, and for transitives, verbs of putting.[1] (8) provides the complete list.

(8)

[1]SMP Mixtec, like many other Mixtec languages, also has three adjectives that show number suppletion.

(1) Adjective number suppletion

	Singular	Plural	
a.	kă'anu	nă'anu	'big'
b.	lo'o	bálí	'little'
c.	ka'nì	na'nì	'long'

In the relevant respect, adjectives pattern identically to verbs. I will only discuss the verbs here.

	Singular	Plural	
a.	ínchichi	ita	'be standing'
b.	íntu'u	ndê	'be sitting'
c.	nakuntu'u	nakundê	'sit down (eventive)'
d.	íyo	ndê	'be located (generic)'
e.	ndanuû	íntoso	'be on top of'
f.	naka	ñu'u	'be inside of'
g.	xíka	kákû	'walk'
h.	kìta	kèhè	'enter'
i.	kanahta	kanakwê	'exit'
j.	chikàà	tàhààn	'put inside'
k.	chinuû	kòsò	'put on top of'

Importantly for our purposes, participant number suppletion is normally obligatory. This is exemplified in (9).

(9) a. Nà nì kèhè nà.
they PAST enter.PL they
'They entered.'

b. * Nà nì kìta nà.
they PAST enter.SG they
Intended: 'They entered.'

c. Ñá nì kìta ñá.
she PAST enter.SG she
'She entered.'

d. * Ñá nì kèhè ñá.
she PAST enter.PL she
Intended: 'She entered.'

In (9a-b) we see that plural suppletion is obligatory with a plural subject, while (9c-d) demonstrate that the plural suppletive allomorph is unavailable if the subject is singular.

With this background on participant number suppletion, let us return our attention to distinguishing the structures in (2) and (5). If (5) were correct and what appears to be the subject of the wh-cleft originated within the pseudo-relative, then it should trigger participant number suppletion. This turns out to not be the case.

To begin, wh-phrases may never trigger plural suppletion.

(10) a. Yó$_i$ **ínchichi** __$_i$?
who stand.PRES.SG
'Who is standing?'

b. * Yó$_i$ **ita** __$_i$?
who stand.PRES.PL
Intended: 'Who is standing?'

c. Nă$_i$ **chinuû** rà Juan __$_i$ nuhŭ mesa?
what put.on.PAST.SG he J. on table
'What did Juan put on the table?'

d. * Nă$_i$ **kòsò** rà Juan __$_i$ nuhŭ mesa?
what put.on.PAST.PL he J. on table
Intended: 'What did Juan put on the table?'

In (10), we see that wh-expressions may never trigger plural suppletion. In a context where the speaker wants an answer with multiple individuals, a pseudo-relative must be used. Interestingly, the verb inside the pseudo-relative never shows plural suppletion in spontaneous speech, although speakers generally judge plural forms as fine if asked directly. This is shown in (11).

(11) a. Yó kúú nà **ínchichi** __?
who COP.PRES they stand.PRES.SG
'Who are they that is standing?'

b. ? Yó kúú nà **ita** __?
who COP.PRES they stand.PRES.PL
'Who are they that are standing?'

c. Nă kúú ntskû míí ña'à **chinuû** rà Juan __ nuhŭ mesa?
what COP.PRES all the thing put.PAST.SG he J. on table
'What are all the things that Juan put on the table?'

d. ? Nă kúú ntskû míí ña'à **kòsò** rà Juan __ nuhŭ mesa?
what COP.PRES all the thing put.PAST.PL he J. on table
'What are all the things that Juan put on the table?'

In (11), we see that singular forms of verbs within the pseudo-relative are totally unmarked. Plural forms, in contrast, are generally accepted as grammatical, although seem to be slightly degraded relative to singular forms. This is important because it indicates that the element that forms the gap is not plural. In particular, compare (11c-d) to (12).

(12) a. * Ntskû míí ña'à$_i$ **chinuû** rà Juan __$_i$ nuhŭ mesa.
all the thing put.PAST.SG he J. on table
Intended: 'Juan put all the things on the table.'

b. Ntskû míí ña'à$_i$ **kòsò** rà Juan __$_i$ nuhŭ mesa.
all the thing put.PAST.PL he J. on table
'Juan put all the things on the table.'

In (12), we see the by-now familiar pattern: plural objects trigger plural suppletion. Now, let us consider the two possible constituent structures for the question in (11c). These were first provided in (2) and (5), and are replicated for this question in (13).

(13) a. [Nă kúú ntskû míí ña'à] [**chinuû** rà Juan __ nuhŭ mesa]?
what COP.PRES all the thing put.PAST.SG he J. on table
'What are all the things that Juan put on the table?'

b. [Nă kúú] [ntskû míí ña'à$_i$ **chinuû** rà Juan __$_i$ nuhŭ mesa]?
what COP.PRES all the thing put.PAST.SG he J. on table
'What are all the things that Juan put on the table?'

Given the patterns of plural suppletion we have just examined, we can be sure that (13b) is not correct. If it were, we would expect plural suppletion within the second bracketed constituent to be obligatory, just as it is in (12). The failure of plural suppletion to be triggered here indicates that the subject of the wh-cleft, in this case *ntskû míí ña'à* 'all the things,' cannot be the source of the gap. Instead, it must be the case that something else derives the gap.

With this, I adopt the basic constituency in (13a). As discussed throughout, this structure has two basic pieces: the wh-cleft and the pseudo-relative. Let us first examine the syntax of the cleft in a cursory way before considering the true focus of this subsection: the gap in the pseudo-relative.

A.1 The syntax of the pseudo-relative cleft

Recall that in an external argument wh-question like (14), the first bracketed constituent is the cleft.

(14) [Yó kúú nà] [nì kìta __ be'e]?
who COP.PRES they PAST enter.SG house
'Who are they that entered the house?'

As (14) demonstrates, wh-clefts like *yó kúú nà* 'who are they' consists for three parts. The first is the wh-phrase, in this case *yó* 'who,' followed by the copula *kúú*, and finally a subject. The subject is most often *nà* 'they,' although it can be any nominal, depending on the context. (15) demonstrates a variety of contexts that license other subjects within the cleft.

(15) a. [Yó kíí **yá**] xíta __?
who COP.PRES it sing.PRES
'Who is it that is singing?' (Context: You hear one voice singing and want to know who it is.)

b. [Yó kúú **rà**] ndánuû __ sata kwáyi.
who COP.PRES he be.on.PRES.SG back horse
'Who is he who is on the horse?' (Context: You see someone on a horse, and you can tell from the shadow that they are a man.)

c. [Ntsyâ rí kúú míí kitsì bálí] chintsyě __ yé?
which it.AML COP.PRES the animal little.PL help.IRR us.INCL
'Which of them are the little animals that will help us?'

In (15), we see that the subject of the pseudo-relative cleft can be any nominal expression, depending on the what the contexts provides as the expected answer to the question. Additionally, note that the copula in the cleft is invariably marked present in natural speech, although it is generally judged as acceptable with past or irrealis marking.

(16) [Yó { kúú, ?xikuu, ?kuù } nà] . . .
who COP.PRES COP.PAST COP.IRR they
'Who { are, ?were, will be } they who. . .'

In addition to the TAM marking on the copula inside of the wh-cleft, its form is also revealing. SMP Mixtec has two elements that can be translated as the copula in English[2], and the choice between them depends on the category of the non-verbal predicate. In addition to the verbal predicates focused on throughout this dissertation, SMP Mixtec also allows nominal and adjectival predicates. These, as well as the form of the copula in each of the three TAM categories, are shown in (17).

(17) a. Rà mástro { kúú, xikuu, kuù } tát ì.
he teacher COP.PRES COP.PAST COP.IRR father my
'My father {is, was, will be} a teacher.' (Nominal predicate)

b. Ñà kwǎ'àn { kúú, nì xihyo, kuù } be' ì.
it red COP.PRES PAST COP.PAST COP.IRR house my
'My house {is, was, will be} red.' (Adjectival predicate)

In (17), we see instances of nominal and adjectival predication. Crucially, we see in (17a) that the form of the copula in nominal predicate constructions is identical to the form of the copula in the clefts in (16). Given this, it seems that the default assumption should be that the cleft should have the same syntax as a nominal predicate.

[2]In the traditional Mixtec literature, several other verbs fall into the category of "equative verbs" in addition to these two. These are *nduu* 'to become' and *náni* 'to be called.' These other two are not relevant for our purposes here.

Nominal predication in SMP Mixtec is complex, and to investigate its syntax in too much detail here would lead us too far astray. Regardless, some brief comments are in order. Consider the further examples of nominal predication in (18).

(18) a. Maestro siki (kúú) tát ì.
teacher funny COP.PRES father my
'My father is a funny teacher.'

b. Sutu kuù se' ì.
priest COP.IRR child my
'My son will be a priest.'

c. Juan (kúú) táta ñá.
J. COP.PRES father her
'Her father is Juan.'

d. Sntáro ba'ă (kúú) míí Juan.
solider good COP.PRES the J.
'Juan is a solider.'

e. Doktór xikuu táta lá'n ì.
doctor COP.PAST father old my
'My grandfather was a doctor.'

In (18), we see that nominal predication uses the same pieces as clefts in pseudo-relative constructions, with the only obvious difference being that a DP occurs in initial position instead of a wh-phrase. Note also that the present tense copula *kúú* is always optionally silent in nominal predication constructions, just as it is always optionally silent in the pseudo-relative cleft. I take this as further evidence for the parallelism between these two constructions.

There is good evidence that nominal predication is derived by movement of the nominal predicate to pre-copular position. This can be seen clearly in that this movement is blocked in a variety of environments, particular in subjunctive clauses and under negation.[3]

[3]There is a negated nominal predicate construction in which the predicate moves, although this is

(19) a. Kôn ì [kuù se' ì **sutu**].
want.PRES I COP.IRR child my priest
'I want my child to be a priest.'

b. * Kôn ì [kuù **sutu** se' ì].

c. * Kôn ì [**sutu** kuù se' ì].

d. Kò kúú rà Juan **sntáro ba'ă**.
NEG COP.PRES he J. solider
'Juan is not a good solider.'

e. * Kò kúú **sntáro ba'ă** rà Juan.

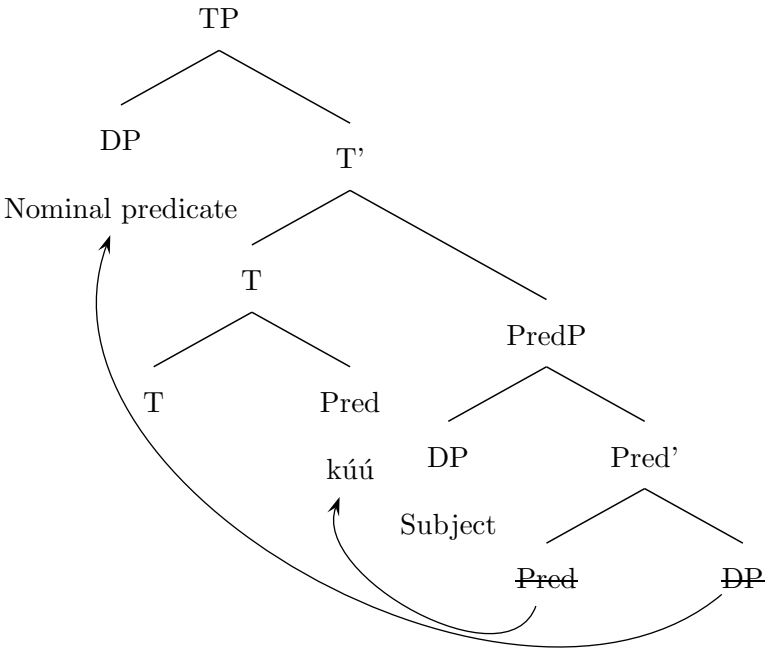
f. * **Sntáro ba'ă** kò kúú rà Juan.

In (19), we see that raising of the nominal predicate is blocked in subjunctive clauses, as in (19a-c), and by negation, as in (19d-f). As predicate fronting is blocked here, I assume the surface position of the nominal predicate in (19a, d) reflects its base position. This is shown in (20).

restricted to simplex nominal predicates. This involves the element *sĭ*, which occurs exclusively with negated nominal predicates in the present tense.

(1) Sĭ rà brujo rà.
NEG.NOM he sorcerer he
'He is not a sorcerer!'

(20)



(20) provides the proposed syntax for nominal predication and, by analogy, for clefts of the sort in (14-16). The only difference is, of course, that in pseudo-relative clefts, the nominal predicate is the wh-phrase.[4]

Let us now turn our attention in a much less trivial way to the syntax of the pseudo-relative.

A.2 Pseudo-relative gaps do not show A'-characteristics

To begin, let us review what exactly a pseudo-relative is. In a sentence like (21), the pseudo-relative is the second bracketed constituent. See the introduction to this appendix for evidence for this constituency.

(21) [Yó kúú nà] [nì kìta __ be'e]?
who COP.PRES they PAST enter.SG house

[4]It may be the case that wh-nominal predicates undergoes further movement to Spec,CP after usual predicate fronting to Spec,TP. Determining this is not productive towards our ultimate goal of deriving the gap in pseudo-relatives.

'Who are they that entered the house?'

In (22), the pseudo-relative is the constituent *nì kìta* __ *be'e*, "__ entered the house." Additionally, recall the reason why I refer to this constituent as a pseudo-relative, rather than a true relative clause. As we saw in §4.1, true relative clauses would require a resumptive pronoun in external argument position, rather than the gap in the pseudo-relative. As we saw in §4.2 above and as repeated in (22), pseudo-relative gaps are incompatible with resumption.

(22) a. * [Yó (kúú) nà] [kàni **nà** rà Julio]?
who COP.PRES they hit.PAST they he J.
Intended 'Who are they that (they) hit Julio?'

b. * [Yó (kúú) nà] [káchûn **nà**]?
who COP.PRES they work.PRES they
Intended: 'Who are they that (they) are working?'

c. * [Ntsyâ nà (kúú) míí nà tsyahá yó'o] [ba'á kíchiñu **nà**]?
which they COP.PRES the they man this well work.PRES they
Intended: 'Which of them are these men that (they) are working well?'

Therefore, pseudo-relatives cannot be reduced regular relative clauses.

This pseudo-relative gap is of great importance for our purposes here because it seems to violation this language's A'-extraction restriction, which prohibits external arguments from undergoing A'-movement. Therefore, if we want to uphold this extraction restriction, then we need to show that the gap inside the pseudo-relative is not derived by A'-movement. In this section I present evidence that this is, indeed, the case.

A.2.1 The gap in a pseudo-relative does not induce Crossover

Throughout §4, we saw that A'-movement is reliably diagnosed in this language through Crossover phenomena. Let us begin with Strong Crossover in wh-movement, repeated from §4.

(23) * Yó$_i$ ká'an ná$_i$ [kú'u __]?
who think.PRES they sick.PRES
Intended: 'Who$_i$ do they$_i$ think is sick?'

In (23), we see that wh-movement induces Strong Crossover. As discussed in §4.3.1, this is good evidence that wh-movement is derived through A'-movement.

Interestingly, pseudo-relative gaps do not induce Strong Crossover. In fact, using this sort of construction is the repair for (23), as shown in (24b).

(24) a. Yó (kúú) nà$_i$ [ká'an nà$_i$ kíchiñu __$_i$ ba'ă]?
who COP.PRES they think.PRES they work.PRES well
'Who are they$_i$ that they$_i$ think __$_i$ work well?'

b. Yó (kúú) nà$_i$ [ká'an nà$_i$ kú'u __$_i$]?
who COP.PRES they think.PRES they sick.PRES
'Who are they$_i$ that they$_i$ think __$_i$ are sick?'

In (24), we see that no Strong Crossover occurs in pseudo-relatives. This would be very unexpected if the pseudo-relative gap were derived through A'-movement, and as such, provides initial evidence that the pseudo-relative gap is not derived by A'-movement.

In addition, pseudo-relatives also fail to induce Weak Crossover. First, recall that wh-movement in this language routinely induces Weak Crossover. These data are repeated in (25) from §4.3.1.

(25) a. * Yó$_i$ kôni nána nà$_i$ __$_i$?
who love.PRES mother their
Intended: 'Who$_i$ does his$_i$ mother love __$_i$?'

b. * Yó$_i$ ká'an nána nà$_i$ [kàni rà Julio __$_i$]?
who think.PRES mother their hit.PAST he J.
Intended: 'Who$_i$ does their$_i$ mother think Julio hit __$_i$?'

As with Strong Crossover, using the pseudo-relative construction is the language's repair for avoiding Weak Crossover in these cases. In particular, compare (25b) to (26c).

(26) a. [Yó$_i$ kúú nà] ká'an nána nà$_i$ kàni __$_i$ rà Julio?
who COP.PRES they think.PRES mother their hit.PAST he J.
'Who$_i$ does their$_i$ mother think hit Julio?'

b. [Yó$_i$ kúú nà] ká'an nána nà$_i$ káchiñu ba'á __$_i$?
who COP.PRES they think.PRES mother their work.PRES hard
'Who$_i$ does their$_i$ mother think works hard?'

c. [Yó kúú nà] ká'an táta nà$_i$ chì'i __$_i$ chìkí?
who COP.PRES they think.PRES father their sow.PAST *tuna*
'Who is it that their$_i$ father thinks __$_i$ sowed *tunas*?'

From this investigation, we see that pseudo-relatives do not display the same characteristic restrictions of A'-movement that wh-movement does. Furthermore, this contrast is so robust that using the pseudo-relative is the language's primary method to avoid Crossover violations wholesale. I consider the fact as good initial evidence that these constructions are not derived through A'-movement.

That said, it has been known since at least Lasnik and Stowell (1991) that different kinds of A'-movement demonstrate different strengths of Crossover violations. Therefore, before completely ruling out the possibility of deriving pseudo-relatives with A'-movement, it behooves us to examine other characteristic areas of A'-movement for additional contrast.

A.2.1.1 Generalized intervention in SMP Mixtec and how the gap in a pseudo-relative circumvents it

Going back at least to Cinque (1990), it has been observed that various quantificational expressions, particularly negation and those that cannot be construed with a pair-list interpretation, may block wh-dependencies. The classic example of this comes from Beck (1996), who presents intervention effects in colloquial German. Consider the minimal pair of sentences in (27).

(27) German intervention effects with negation (Beck 1996)

a. ?? Was glaubt niemand wen$_i$ Karl __$_i$ gesehen hat?
what believes nobody whom K. seen has
Intended: 'Who does nobody believe that Karl saw?'

b. Wen$_i$ glaubt niemand dass Karl __$_i$ gesehen hat?
whom believes nobody that Karl seen has
'Who does nobody believe that Karl saw?'

Let us first consider (27a). Here, we see a wh-scope construction (see McDaniel 1989, Dayal 1994, 2000, Fanselow and Mahajan 2000, Bruening 2004, 2006, and Legate 2011 for an overview). Wh-scope marking involves a wh-phrase that does not undergo overt movement to Spec,CP of the matrix clause. Instead, the wh-phrase occupies Spec,CP of its originating clause, while a scope marker, *was* 'what' occurs in Spec,CP of the matrix clause where the content wh-word *wen* 'whom' is meant to be interpreted. In (27a), we see that this scope marking relation cannot take place around negation in the form of the negative subject *niemand* 'nobody.' Importantly, (27b) makes clear that the problem is not with having negation and a wh-dependency in the same clause.

Based on this distribution, Beck (1996) proposes that quantificational elements like negation are only interveners for LF movement. Therefore, if overt movement takes place before LF, as in (27b), there is no issue. At the same time, data from wh-in-situ languages, notably Korean, allows for an even sharper generalization of the data. Consider (28).

(28) Korean intervention effects (Beck and Kim 1997, through Beck 2006)

a. * Amuto **muôs** -ûl ilk -chi anh -ass -ni?
anyone what -ACC read -CHI not do.PAST -Q
Intended: 'What did no one read?'

b. **Muôs**$_i$ -ûl amuto __$_i$ ilk -chi anh -ass -ni?
what -ACC anyone read -CHI not -do.PAST -Q
'What did no one read?'

In (28a), we see an effect very similar to the German in (27). In Korean, a wh-phrase

may not be in the scope of another quantification expression, in this case *amuto* '(not) anyone.' In contrast, we see that in (28b), when the wh-element undergoes scrambling out of the scope of *amuto* '(not) anyone,' the question becomes fully grammatical.

Based on this distribution, Beck (2006) provides the characterization in (29) for the intervention effects in German and Korean.

(29) A wh-phrase may not have a [focus/negation] operator as its closest c-commanding potential binder. (Beck 2006, pg. 17)

In (29), we see that intervention effects are analyzed on the basis of closest c-command, specifically closest c-command of the highest copy at Spell-Out. This derives the pattern in (27b) and (28b) because overt movement, either wh-movement in the former or scrambling in the latter, alters the c-command relations, allowing the restriction in (29) to be circumvented.

At the same time, it is not clear what in the theory requires the highest copy of a wh-expression to be universally privileged in this way. For instance, we know that there is significant cross-linguistic variation in intervention effects. One clear case of this is in which operators act as interveners. Two clear cases of this variation come from negation and 'often.' First, consider negation.

(30) a. * Was glaubt Hans nicht, wer da war?
what believes H. not who there was
Intended: 'Who does Hans not believe was there?' (German, Beck 1996)

b. Nít mây síi aray?
N. not buy what
'What didn't Nít buy?' (Thai, Ruangjaroon 2002)

In (30a), we see that negation is a strong intervener in German. In contrast, (30b) shows that negation is not an intervener in Thai.

Additionally, 'often' is a strong intervener in German, but not in Korean.

(31) a. * Luise zählt auf, welche Uni oft welche Linguisten eingeladen
L. enumerates which university often which linguistics invited
hat.
has
Intended: 'Luise enumerates which university often invited which linguists.' (German, Beck 2006)

b. Minsu -nûn chachu nuku -lûl p'ati -e teliko kass -ni?
M. -TOP often who -ACC party -DIR take PAST -Q
'Who did Minsu often take to the party?' (Korean, Beck and Kim 1997)

In addition to this sort of variation in the set of quantificational elements that trigger intervention effects, there are also languages which lack intervention effects entirely. Notable among these languages are Spanish (Uribe-Etxebarria 2002, Reglero 2007, Reglero and Ticio 2015), Amharic (Eilam 2008), and Shona (Zentz, 2016).

With this, we see that variation in intervention effects is the norm cross-linguistically. Therefore, there is a clear space in the theory of intervention effects for a language that does not privilege the highest copy of a wh-expression at Spell-Out, as German and Korean do. In this sort of language, intervention effects would manifest as in (32).

(32) Hypothetical language: No copy of a wh-phrase may be c-commanded by a focus/negation operator.

A language like (32) would be similar to German or Korean, but overt movement of the wh-expression would not be sufficient to obviate intervention.

Interestingly, SMP Mixtec seems to be just such a language. In SMP Mixtec, it is the case that no copy of a wh-expression may be c-commanded by a focus or negation operator. Let us first consider negation.

(33) a. Yó$_i$ nàkààba __$_i$?
who fall.PAST
'Who fell?'

b. Yó$_i$ ká'an rà Juan [nàkààba __$_i$]?
who think.PRES he J. fall.PAST
'Who does Juan think fell?'

c. *? Yó$_i$ kò ní nakààba __?
who NEG PAST fall
Intended: 'Who didn't fall?'

d. *? Yó$_i$ kò ká'an rà Juan [nàkààba __]?
who NEG think.PRES he J. fall.PAST
Intended: 'Who does Juan not think fell?'

In (33a-b), we see the basic cases of wh-movement. In (33a), we see a simple case of monoclausal wh-movement, while in (33b), we see a standard sort of movement from an embedded clause to the matrix clause. Importantly, in (33c-d), we see that negation does not c-command the surface copy of the wh-expression *yó* 'who.' Nonetheless, the result is ungrammatical, presumably because negation does c-command lower copies of the wh-phrase.[5] It seems clear that this is a kind of intervention effect, with negation acting as an intervener much as it does in German. That said, this is a much stronger intervention effect, as the highest copy of the wh-expression is not c-commanded by negation.

Also as in German and Korean, 'no one' also acts as an intervener in SMP Mixtec. But, as expected, overt movement over 'no one' does not ameliorate the intervention effect.[6]

[5]Note that SMP Mixtec is a strongly NEG-raising language. By this, I mean that negation can never occur within clauses embedded under verbs like *ka'an* 'think.'

(1) a. Kò ká'an rà Juan [nàkààba míí ítŭn yó'o].
NEG think.PRES he J. fall.PAST the tree this
'Juan doesn't think that this tree fell ∼ Juan thinks that this tree did not fall.'

b. * Ká'an rà Juan [kò ní nakààba míí ítŭn yó'o].
think.PRES he J. NEG PAST fall the tree this
Intended: 'Juan thinks that this tree did not fall.'

Therefore, the position of negation in (33c-d) is the only possible place it may occur.

[6]The fronting of *ní iin nà* 'no one' to the preverbal position in (34a 34c), as well as its post-verbal occurrence in (34b,3 4e), is not a relevant part of the pattern. The fronting is an instance of the general operation of quantifier fronting, discussed in §3.

(34) a. Ní iin nà ní xixi nduchì.
not.even one them PAST eat beans
'No one ate beans.'

b. * Nă$_i$ ní xixi ní iin nà__$_i$?
what PAST eat not.even one them
Intended: 'What did no one eat?'

c. Ní iin nà ká'an [kú'u ñá Margarita].
not.even one them think.PRES be.sick.PRES she M.
'No one thinks that Margarita is sick.'

d. Yó$_i$ ká'an ñá doktóra [kú'u __]?
who think.PRES she doctor be.sick.PRES
'Who does the doctor think is sick?'

e. * Yó$_i$ ká'an ní iin nà [kú'u __]?
who think.PRES not.even one them be.sick.PRES
Intended: 'Who does no one think is sick?'

In (34), we see that wh-movement may not take place over *ní iin nà* 'no one.' Again, this is clearly a manifestation of the same intervention effect reported for German and other languages, and demonstrated in (27-28). The primary difference between German and SMP Mixtec, then, is that overt movement over the offending quantifier does not ameliorate the intervention.

Rather, the repair for the ungrammatical sentences of (33-34) is to use a pseudo-relative. This is shown in (35).

(35) a. Yó$_i$ nàkààba __$_i$?
who fall.PAST
'Who fell?'

b. * Yó$_i$ kò ní nakààba __$_i$?
who NEG PAST fall
Intended: 'Who didn't fall?'

c. [Yó kúú nà] kò ní nakààba __?
who COP.PRES they NEG PAST fall

'Who is it that didn't fall?'

d. * Yó$_i$ ká'an ní íín nà [kú'u __$_i$]?
who think.PRES not.even one them sick.PRES
Intended: 'Who does no one think is sick?'

e. [Yó kúú nà] ká'an ní íín nà [kú'u __]?
who COP.PRES they think.PRES not.even one them sick.PRES
'Who is it that no one thinks is sick?'

(35c, e) demonstrate that the pseudo-relative construction repairs the intervention effects in (35b, d). Before considering why this should be the case, let us investigate further intervention effects in this language.

In addition to negation, 'only' also acts as a common intervener cross-linguistically. Consider the Korean in (36).

(36) 'Only' as an intervener in Korean (Beck and Kim 1997, through Beck 2006)

a. Minsu -num **nuku** -lûl po -ass -ni?
M. -TOP who -ACC see -PAST -Q
'Who did Minsu see?'

b. * Minsu -man **nuku** -lûl po -ss -ni?
M. -only who -ACC see -PAST -Q
Intended: 'Who did only Minsu see?'

c. **Nuku** -lûl minsu -man po -ass -ni?
who -ACC M. -only see -PAST -Q
'Who did only Minsu see?'

In (36), we see a now familiar pattern: the wh-expression may not be c-commanded by the focus operator *-man* 'only.' To get around this, the wh-phrase must be scrambled outside of the c-command domain of *-man* 'only.'

Testing 'only' in SMP Mixtec has proven tricky, as there seem to be several strategies for expressing focus-sensitive uniqueness, and not all speakers have the same set of strategies. The three possibilities, as meaning roughly the same thing, are given in (37).

(37) a. Nùhmĭ **míí** Maria tsĭna lo'o.
hug.PAST the M. dog little
'(Only) Maria hugged the puppy.'

b. Maria **úún ba** nùhmĭ tsĭna lo'o.
M. only EMPH hug.PAST dog little
'Only Maria hugged the puppy.'

c. **Íín** Maria (kúú) ñá nùhmĭ tsĭna lo'o.
one M. COP.PRES she hug.PAST dog little
'It was just Maria who hugged the puppy.'

(37) presents the three general strategies that speakers use to translate sentences with 'only.' In (37a), the phrase which is meant to be interpreted with focus-sensitive uniqueness is marked with *míí* 'the.' *Míí* is not an exact translation of 'only,' as it does not denote a set of focus-alternatives (Rooth 1992). In (37b), we see the word *úún*, which does seem to be a good translation of 'only' in that it does denote a set of focus-alternatives, followed by the emphatic clitic *ba*. As is typical for quantified phrases, the phrase with *úún ba* undergoes quantifier movement. See §3. In the last translation of 'only,' (37c), a cleft is used in which the word *iin* 'one' denotes uniqueness.

To test whether 'only' in this language acts as an intervener, only the strategies in (37a-b) were tested. The strategy in (37c) was not tested because it involves a cleft, and is therefore not syntactically compatible with wh-movement.

Interestingly, only the strategy in (37b) induces an intervention effect. This has a straightforward explanation based on the analysis of intervention effects in Beck (2006) because *míí*-marked phrases do not denote a set of focus-alternatives, and are therefore not predicted independently to demonstrate intervention.

(38) a. Nă$_i$ nùhmĭ míí Maria __$_i$?
what hug.PAST the M.
'What did Maria hug?'

b. * Nă$_i$ nùhmĭ Maria úún ba __$_i$?
what hug.PAST M. only EMPH

Intended: 'What did only Maria hug?'

In (38a), we see that *míí*, 'the,' does not induce intervention effects. Following Beck (2006), I take this to be a straightforward consequence of *míí* not inducing a set of focus alternatives. In contrast, in (38b), we see that *úún ba* 'only,' which does induce focus alternatives, acts as an intervener. This is further confirmation that this effect is indeed best considered a kind of intervention, on par with German, Korean, and others.

Importantly, as in (36), the pseudo-relative ameliorates the intervention effect of *úún ba* 'only.' This is shown in (39).

(39) a. * Yó$_i$ nùhmı̌ Maria úún ba __?
who hug.PAST M. only EMPH
'Who did only Maria hug?'

b. [Yó kúú nà] nùhmı̌ Maria úún ba __?
who COP.PRES thhey hug.PAST M. only EMPH
'Who is it that only Maria hugged?'

The final intervention effect I will consider is the relation between wh-expressions and universal quantifiers. Beck (1996) observes that for many German speakers, (40a) is grammatical, but is only compatible with a pair-list answer.

(40) a. Was glaubt jeder wen Karl gesehen hat?
what believes everyone whom K. seen has
'Who does everyone believe Karl saw?'

b. Luise believes Karl saw Bernard, Hans believes Karl saw Detmar...

c. ?? Bill.

(40a), in contrast to (27a), is fully grammatical, but it requires a particular kind of answer. (40b) demonstrates that the right kind of answer to (40a) involves listing each individual in the set denoted by *jeder* 'everyone' with each individual's respective answer

to the question. I refer to this response as a pair-list answer.[7] In contrast, (40c), which lacks the possibility of a pair-list reading, is judged as severely degraded.

A similar effect emerges in SMP Mixtec. Consider the paradigm in (41), which mirrors (40).

(41) a. Ntsi'i nà xìxi xità.
all them eat.PAST tortilla
'Everyone ate tortillas.'

b. Nă$_i$ xìxi ntsi'i nà __$_i$?
what eat.PAST all them
'What did everyone eat?'

c. Xìxi rà Juan kôñù, xìxi ñá Maria xità...
eat.PAST he J. meat eat.PAST she M. tortilla
'Juan ate meat, Maria ate tortillas...'

d. ?? Xità.
'Tortillas.'

In (41), we see that universal quantifiers do not act as interveners for wh-movement, but they do enforce a pair-list reading. Following Beck (1996), I interpret this as a form of intervention effect as well.

Here, again, the repair for the intervention effect is to use a pseudo-relative. This is shown in (42).

(42) a. Nă$_i$ xìxi ntsi'i nà __$_i$?
what eat.PAST all them
'What did everyone eat?' (✓pair-list, **x** singleton)

b. [Nă kúú yá] xìxi ntsi'i nà __?
what COP.PRES it eat.PAST all them
'What did everyone eat?' (✓pair-list, ✓singleton)

[7]Beck (1996) observes that another answer to (40a) is something 'Everyone believes Karl saw Bill.' Importantly, she argues that this is a form of pair-list answer as well, just every member of the set denoted be 'everyone' maps to a single individual.

Throughout this discussion, we have seen that SMP Mixtec demonstrates a typologically novel form of intervention effect. Unlike other languages where the restriction is merely that the highest copy of a wh-expression may not be c-commanded by a focus or negative operator, no copies of a wh-expression may be within the domain of the quantifier in SMP Mixtec.

But how can we derive the amelioration effect of the pseudo-relative? First, recall the proposed intervention generalization for this language, repeated in (43).

(43) **SMP Mixtec wh-intervention generalization**: No copy of a wh-phrase may be c-commanded by a focus/negation operator.

As it turns out, we can strengthen the generalization in (43) to that in (44).

(44) **Generalized intervention in SMP Mixtec**: No copy of an A'-chain may be c-commanded by a focus/negation operator.

The much stronger generalization in (44), though certainly cross-linguistically novel, is appropriate for this language. To see this, let us consider the second A'-movement process of §4: relativization in restricted clauses.First, negative operators are degraded in relative clauses, just as in wh-movement constructions.

(45) a. ?? Rà$_i$ ntá'yi rà$_i$ lo'o [kò ní nakààba __$_i$].
he cry.PRES he little NEG PAST fall
Intended: 'The boy that didn't fall is crying.'

b. *? Bèhè kwê'e míí *piano*$_i$ [kǔǔ naki'i ní íín nà __$_i$].
heavy very the piano can.NEG lift.IRR not.even one them
Intended: 'The piano that no one can lift is very heavy.'

In (45), we see that negation inside of a relative clause is severely degraded. This is consistent with the strong strong intervention generalization in (44).[8]

Úún ba 'only' is also banned from relative clauses. This is shown in (46).

[8]The repair for sentences like (133) does not involve resumption. This is expected within the overall context of this language, which usually prohibits resumption of internal arguments under all circum-

(46) * Rí ntá'yi míí tsǐna lo'o$_i$ [nùmhǐ ñá Maria úún ba __$_i$].
it.AML cry.PRES the dog little hug.PAST she M. only EMPH
Intended: 'The puppy that only Maria hugged is barking.'

Therefore, the much stronger intervention generalization in (44) is accurate for this language.

Deriving the difference between SMP Mixtec and German or Korean is beyond the scope of this dissertation. But, for our purposes, (44) is strong enough to serve as a diagnostic for A'-movement in SMP Mixtec. Through this lens, it becomes clear that any movement in pseudo-relatives cannot be A'-movement.

With this, we have two strong avenues of evidence that argue that pseudo-relative gaps are not derived by A'-movement. With this much in place, let us consider what sort of movement, if any, can be diagnosed in pseudo-relatives.

A.3 Pseudo-relative gaps are derived by mixed movement

To preview the findings of this section, I will argue for two statements. First, the gap in pseudo-relatives is derived by movement. Second, along with the results of §A.2, this movement is certainly not A'-movement. That said, this movement is also not obviously A-movement. Rather, it seems to demonstrate a mixed set of properties.

stances. Instead, the only repair available seems to be periphrasis. These data are shown in (i).

(1) a. ?? Rà$_i$ ntá'yi rà$_i$ lo'o [kò ní nakààba **rà**$_i$].
he cry.PRES he little NEG PAST fall he
Intended: 'The boy that (he) didn't fall is crying.'
b. Rà ntá'yi rà lo'o kân, so kò ní nakààba rà.
he cry.PRES he little that but NEG PAST fall he
'That boy is crying, but he didn't fall.'
c. *? Bèhè kwê'e míí *piano*$_i$ [kǔǔ naki'i ní íín nà **ñà**$_i$].
heavy very the piano can.NEG lift.IRR not.even one them it
Intended: 'The piano that no one can lift (it) is very heavy.'
d. Bèhè kwê'e míí *piano*. Ní íín nà kǔǔ naki'i ñà.
heavy very the piano not.even one them can.NEG lift it
'The piano is very heavy. No one can lift it.'

Regardless of the inherent interest of this mixed movement system, the critical conclusion for §4 is that this movement is not A'-movement. In this way, pseudo-relative movement, no matter its ultimate derivation, does not violate SMP Mixtec's otherwise inviolable restriction against A'-extraction of external arguments.

A.3.1 The gap in a pseudo-relative is unbounded and island sensitive

The gap inside of the pseudo-relative cannot occur within an island. This is the case for the three islands we saw in §4, each shown in (47).

(47) a. * Yó$_i$ kúú nà [kíchiñú rà Juan [ná kuxi { __/ nà }$_i$
who COP.PRES they work.PRES he J. SUBJ eat.IRR they
kǒñù ntsikihi]]?
meat everyday
Intended: 'Who$_i$ does Juan work hard so that (they$_i$) can eat meat everyday?' (so-that clause island)

b. * Yó$_i$ kúú nà [kúú sihĭ ini ñá Isabela [chi
who COP.PRES they COP.PRES happy inside she I. because
sàsiki {__/ nà }$_i$]]?
play.PAST they
Intended: 'Who$_i$ was Isabela happy because (they$_i$) were playing?' (because clause island)

c. * Yó$_i$ kúú nà [xĭñ ú [náchûn kàni {__/ nà}$_i$
who COP.PAST they know.PRES.NEG you why hit.PAST they
rà Julio]]?
he J.
'Who is it that you do not know why (they) hit Julio?' (wh-island)

In each of the examples in (47), we see that the gap inside of the pseudo-relative cannot be within an island. This is an important contrast with true external argument relative clauses, which, as we saw in §4.2.2, freely allow resumptive pronouns in islands. These data are repeated in (48).

(48) a. Kwátsyá ntsya'a ñá sí'i$_i$ [kíchûn rà Juan [ná kuxi **ñá**$_i$ kǒñù
happy very she woman work.PRES he J. SUBJ eat.IRR she meat
ntsikihi]].
everyday
'The woman [who Juan works [so that she can eat meat everyday]] is very happy.'

b. Míí rà tsyahá$_i$ [kwátsyá kwê'e ñá Maria [chi kíchûn **rà**$_i$]],
the he man happy very she M. because work.PRES he
Juan náni rà.
J. called.PRES he
'The man who Maria is happy because he is working is named Juan.'

c. Rà$_i$ xáku míí rà lo'o$_i$ [xǐn ì [yó kàni **rà**$_i$]].
he laugh.PRES the he little know.PRES.NEG I who hit.PAST he
'The boy who I don't know who he hit is laughing.'

I consider the contrast between pseudo-relatives and external argument resumption to be strong evidence that movement is involved in pseudo-relatives. That said, simply positing movement is not especially elucidating within the present context, because the character of the movement as either A'-movement or something else is paramount.

A.3.2 Pseudo-relative movement is not A-movement either

Throughout §4, we investigated the A'-extraction restriction in SMP Mixtec, repeated in (50).

(49) **A'-extraction restriction in SMP Mixtec**: External arguments may not undergo A'-extraction.

As we have seen, the restriction in (50) is certainly true: external arguments may not undergo A'-extraction either in relative clauses or in wh-constructions. With that said, (50) is a sub-restriction within a broader generalization about A'-extraction, introduced in §3.2.3. This broader generalization is made in (51).

(50) **Broad A'-extraction restriction in SMP Mixtec**: External arguments, objects of prepositions, and possessors may not undergo A'-extraction.

In (51), we see that a more comprehensive generalization can be made about which sorts of phrases can undergo A'-extraction. Rather than just external arguments, objects of prepositions and possessors are both also categorically banned from undergoing A'-extraction. Consider (52) and (53).[9]

(51) No A'-extraction of objects of prepositions

a. * Kúú sihǐ ini míí sa'ba [íyakuu míí tsǐna sata __].
COP.PRES happy inside the frog lick.PRES the dog on
Intended: 'The frog that the dog is licking is happy.'

b. * Yó$_i$ nì ka' ú xí'in __$_i$?
who PAST talk you with
Intended: 'Who did you talk to?'

(52) No A'-extraction of possessors

a. * Ntsí'i kwê'e ini míí rà lo'o [kú'u nána __].
sad very inside the he little be.sick.PRES mother
Intended: 'The boy whose mother is sick is very sad.'

b. * Yó nì xi'i tsǐna sâna __?
who PAST die dog POSS.AML
Intended: 'Whose dog died?'

In the (a) examples of (52) and (53), we see that neither objects of prepositions nor possessors may undergo relativization with a gap, which we saw in §4.2, is an A'-movement process. Likewise, in the (b) examples, we see that objects of prepositions nor possessors may undergo wh-movement either.

[9]Note that, in addition to being banned from A'-extraction, external arguments are also islands. See §4.1.3. As such, the discussion throughout with regards to both prepositions and possessors is restricted to internal arguments.

As expected given the strong generalization in (51), the repair for the relativization cases in (52a) and (53a) is a resumptive pronoun, just as it is for external arguments, as shown in §4.2.2. This is shown in (54).

(53) a. Kúú sihǐ ini míí sa'ba [íyakuu míí tsǐna sata **rí**].
COP.PRES happy inside the frog lick.PRES the dog on it.AML
'The frog that the dog is licking (it) is happy.'

b. Ntsí'i kwê'e ini míí rà lo'o [kú'u nána **rà**].
sad very inside the he little be.sick.PRES mother his
Intended: 'The boy who (his) mother is sick is very sad.'

The repair for wh-movement is a bit more intricate. This is because there are two options, each of theoretical interest. The first is pied-piping with inversion, a common process throughout Mesoamerica (see Smith-Stark 1988, Black 1994, Aissen 1996, Eberhardt 1999, Broadwell et al. 2006, Coon 2009, and Cable 2012, 2013 for a discussion of pied-piping with inversion both within Oto-Manguean and Mesoamerica more broadly. See §5.1 as well). The pied-piping with inversion cases are presented in (55).

(54) a. [Yó xí'in]$_i$ nì ka' ú $\_\_{}_i$?
who with PAST talk you with
'Who did you talk to?'

b. [Yó tsǐna sâna]$_i$ nì xi'i $\_\_{}_i$?
who dog POSS.AML PAST die
'Whose dog died?'

In (55), we see that pied-piping with inversion avoids violating the restriction in (51) because the object of the preposition nor the possessor themselves under A'-extraction. Rather, their entire containing constituent moves, essentially ameliorating the violation in (51) by A'-extracting an internal argument, which we have seen throughout this chapter is perfectly grammatical in this language.

The second way that (52b, 53b) can be ameliorated, and of particular interest for us here, is by using a pseudo-relative.

(55) a. [Yó (kúú) nà] nì ka' ú xí'in __?
who COP.PRES they PAST talk you with
'Who is it that you talked with?'

b. [Yó (kúú) nà] nì xi'i tsǐna sâna __?
who COP.PRES they PAST die dog POSS.AML
'Who is it that (their) dog died?'

In naturally occurring sentences of SMP Mixtec, the pseudo-relative repair in (56) is by far the most common, and an extremely robust way that the language uses to avoid violating (51).

At this point, we have a way of understanding why (56) is an available repair to avoid violating (51). With pied-piping with inversion, the language avoids violating (51) by extracting the entire internal argument, rather than illicit subextraction of either a possessor or an object of a preposition. In (56), the repair avoids (51) by using movement, but a movement process that is demonstrably not A'-movement, as discussed in §A.2.

For purposes of comparison, recall quantifier fronting, which I argued in §3 was A-movement.

(56) a. Ntsi'bǐ$_i$ kí'i ñá __$_i$.
egg pick.up.PRES she
'She's gathering eggs.'

b. Ba'á [áto ntskû nà$_i$ kamá karakono __$_i$].
good if all them fast run.IRR
'It would be good if everyone ran fast.'

c. Ní íín nà$_i$ ǐ sókŏn __$_i$.
not.even one them NEG tall
'No one is tall.'

In (57), we see quantifier movement having applied to a variety of quantified expressions, from the bare indefinite in (57a) to *ní íín nà* 'no one' in (57c).

As discussed in §3.2.3 quantifier movement has another interesting property that distinguishes it from A'-movement: quantifier movement, like pseudo-relative movement, can extract possessors and objects of prepositions. This is shown in (58) and (59).

(57) Quantifer movement may extract possessors

a. U'un ná sí'i$_i$ kútó ì *empresa* ña'a $\_\_{}_i$.
five they.FEM woman like.PRES I business POSS
'I like the five women's businesses.'

b. Kòhmǐ rà tsyahá nì xi'i tsǐna sâna $\_\_{}_i$.
four he man PAST die dog POSS.AML
'The four men's dog died.'

(58) Quantifier movement may extract objects of prepositions

a. Ní íín nà$_i$ kò ní ka'a ñá Maria xí'in $\_\_{}_i$.
not.even one them NEG PAST speak she M. with
'Maria did not speak with anyone.'

b. Ntskû míí rí kitsì$_i$ kàyu'u rà lo'o nuhŭ $\_\_{}_i$.
all the it.AML animal shout.PAST he little at
'The boy shouted at all the animals.'

(58) and (59) are perfectly grammatical and natural, and indicate that quantifier movement may target both possessors and objects of prepositions.[10] Again, this makes sense if quantifier movement is a form of A-movement, as argued in §3. As A-movement, quantifier movement can extract objects of prepositions and possessors without violating (51). With this parallelism between quantifier movement and pseudo-relative movement established, we might wonder if pseudo-relative movement is A-movement too.

[10]Naturally, as discussed in §3, (58) and (59) require that the raised quantifier take scope. For instance, in (58a), there must be a business for each of the five women, yielding five businesses in total, while, likewise, in (58b) there must be four dead dogs, one that belongs to each of the four men.

That said, it seems clear that pseudo-relative movement cannot be reduced to quantifier movement because pseudo-relative movement is not clause bounded, while quantifier movement is. See §3.2.1 for a discussion of the boundedness of quantifier movement.

(59) a. [Yó (kúú) nà] ká'an míí nà yìbi keba'a __?
who COP.PRES they think.PRES the they people win.IRR
'Who is it that the people think will win?'

b. [Yó (kúú) nà] ká'an míí nà mástro nì xă'nu __ mésa?
who COP.PRES they think.PRES the they teacher PAST break table
'Who is it that the teachers think broke the table?'

c. [Yó (kúú) nà] ká'an míí mástro kani __ míí Sebastian
who COP.PRES they think.PRES the teacher hit.IRR the S.
bítsí?
today
'Who is it that the teacher thinks will hit Sebastian today?'

Therefore, we must conclude that pseudo-relative movement shows a mixed set of properties. These are summarized in (61).

(60) Movement properties in SMP Mixtec

		A-movement	A'-movement	Pseudo-relative movement
a.	Clause bounded?	✓	**x**	**x**
b.	Extract possessor & obejct of P?	✓	**x**	✓
c.	Induce Crossover?	**x**	✓	**x**
d.	Induce intervention effects?	**x**	✓	**x**

Through this lens, there are two potential analyses of pseudo-relative movement. First, we can simply accept that pseudo-relative movement shows a mix between A- and A'-properties. This would not be unheard of. See van Urk (2015) for an overview. Second, it may be that pseudo-relative is an unbounded A-movement. This would fit the properties in (61), and we saw in §2.2.1.1 that not all A-movement processes cross-linguistically are clause bounded. Regardless, it is unavoidable that pseudo-relative

movement is not A'-movement, which is the only necessary conclusion for the argumentation in §4.

Bibliography

Abels, Klaus. 2003. "Successive Cyclicity, Anti-Locality, and Adposition Stranding". Doctoral Dissertation, University of Connecticut.

Aboh, Enoch. 2010. "Information Structuring Begins with Numeration". Iberia: An International Journal of Theoretical Linguistics 2:12–42.

Aissen, Judith. 1992. "Topic and focus in Mayan". Language 68:43–80.

Aissen, Judith. 1996. "Pied-Piping, Abstract Agreement, And Functional Projections in Tzotzil". Natural Language and Linguistic Theory 14:447–491.

Aissen, Judith. 1997. "On the Syntax of Obviation". Language 73:705–750.

Aissen, Judith. 2003. "Differential Object Marking: Iconicity vs. Economy". Natural Language and Linguistic Theory 21:435–483.

Aissen, Judith. To appear. "Documenting topic and focus". Language Documentation and Conservation. PDF available at https://people.ucsc.edu/∼aissen/PUBLICATIONS/doctopfoc.pdf.

Alexander, Ruth Mary. 1988. "A Syntactic Sketch of Ocotepec Mixtec", 151–304. Number 1 in Studies in the Syntax of Mixtecan Languages. Summer Institute of Linguistics of the University of Texas at Arlington.

Alexiadou, Artemia, and Elena Anagnostopoulou. 1998. "Raising without infinitives and the nature of agreement". Ms. Available at http://www.philology.uoc.gr/staff/anagnostopoulou/files/a2.pdf.

Alexiadou, Artemis, and Elena Anagnostopoulou. 1997. "Towards a uniform account of scrambling and clitic-doubling", volume *German: syntactic problems - problematic syntax*, 142–161. Tübingen: Max Niemeyer Verlag.

Alshamari, Murdhy. 2017. Topic Particles in the North Hail Dialect of Najdi Arabic. Doctoral Dissertation, Newcastle University.

Anagnostopoulou, Elena. 2003. The syntax of ditransitives: Evidence from clitics. Berlin: Mouton de Gruyter.

Anagnostopoulou, Elena. 2007. "Clitic doubling", 519–581. The Blackwell Companion to Syntax. Madlen: Blackwell Publishing.

Anderson, Stephen. 1976. "On the Notion of Subject in Ergative Languages", volume Subject and Topic, 1–24. New York: Academic Press.

Anderson, Stephen. 2005. Aspects of the Theory of Clitics. Number 11 in Oxford Studies in Theoretical Linguistics. Oxford, UK: Oxford University Press.

Aoun, Joseph, Elabbas Benmamoun, and Dominique Sportiche. 1994. "Agreement, word order and conjunction in some varieties of Arabic". Linguistic Inquiry 25:195–220.

Aoun, Joseph, Lina Choueiri, and Norbert Hornstein. 2001. "Resumption, Movement, and Derivational Economy". Linguistic Inquiry 32:371–403.

Arregi, K. 2003. "Clausal Pied-Piping". Natural Language Semantics 11:115–143.

Baier, Nico. 2014. "Spell-out, Chains, and Long Distance *Wh*-movement in Seereer". Unpublished Qualifying Exam Paper. Available at http://linguistics.berkeley.edu/∼nbbaier/documents/baier_qp_final.pdf.

Baker, Mark. 2012. "On the relationship of object agreement and accusative case: Evidence from Amharic". Linguistic Inquiry 43:255–274.

Baker, Mark. 2015. *Case: Its Principles and Parameters*. Cambridge University Press.

Baker, Mark, and Ruth Kramer. 2016. "Doubling Clitics are Pronouns: Reduce and Interpret". Available at http://www.rci.rutgers.edu/*sim*mabaker/clitics-are-pronouns-resubmitted.pdf.

Baker, Mark, and Ruth Kramer. 2018. "Doubled Clitics are Pronouns: Amharic Objects (and Beyond)". Natural Language and Linguist Theory .

Baker, Mark, and Willie Udo Willie. 2010. "Agreement in Ibibio: From Every Head, To Every Head". Syntax 13:99–132.

Barker, C. 2012. "Quantification binding does not require c-command". Linguistic Inquiry 43.

Beatham, Bruce, and Candice Beatham. 2014. Pequeño Diccionario Ilustrado Tu̲'un na̲ ñuu Ki̲yaa xí'in na̲ xiyo oeste ñuu Skuíya. Instituto Lingüístico de Verano, A.C. Tercera edicion.

Beck, Sigrid. 1996. "Quantified Structures as Barriers for LF Movement". Natural Language Semantics 4:1–56.

Beck, Sigrid. 2006. "Intervention Effects Follow from Focus Interpretation". Natural Language Semantics 14:1–56.

Beck, Sigrid, and Shin-Sook Kim. 1997. "On Wh- and Operator Scope in Korean". Journal of East Asian Linguistics 6:339–384.

Béjar, S., and M. Rezac. 2009. "Cyclic Agree". Linguistic Inquiry 40:35–73.

Béjar, Susana. 2003. Phi-Syntax: A Theory of Agreement. Doctoral Dissertation, University of Toronto.

Benmamoun, Elabbas, Archna Bhatia, and Maria Polinsky. 2009. "Closest Conjunct Agreement in Head-Final Languages". Linguistic Variation Yearbook 9:67–88.

Bennett, Ryan. 2012. Foot-conditioned phonotactics and prosodic constituency. Doctoral Dissertation, UC Santa Cruz.

Bhatt, Rajesh. 2005. "Long distance agreement in Hindi-Urdu". Natural Language and Linguistic Theory 23:757–807.

Bhatt, Rajesh, and Martin Walkow. 2013. "Locating agreement in grammar: an argument from agreement in conjunctions". Natural Language and Linguistic Theory 31:951–1013.

Bittner, Maria. 1994. Case, Scope and Binding. Dordrecht, the Netherlands: Kluwer.

Black, Cheryl. 1994. Quiegolani Zapotec Syntax. Doctoral Dissertation, University of California, Santa Cruz.

Bleam, Tonia. 2000. Leísta Spanish and the Syntax of Clitic Doubling. Doctoral Dissertation, Univeristy of Pennsylvania.

Bobaljik, Jonathan. 2006. "Where's ϕ?: Agreement as a post-syntactic operation".

Bobaljik, Jonathan, and Heidi Harley. 2017. "Suppletion is Local: Evidence from Hiaki", volume The structure of words at the interfaces. Oxford University Press.

Bobaljik, Jonathan, and Höskuldur Thráinsson. 1998. "Two Heads Aren't Always Better Than One". Syntax 1:37–71.

Boeckx, Cedric. 2004. "Long-distance agreement in Hindi: theoretical implications". Studia Linguistica 58:23–36.

Boeckx, Cedric. 2010. "On Long-Distance Agree". Iberia: An International Journal of Theoretical Linguistics 1:1–32.

Bolinger, Dwight. 1973. "Ambient *it* is meaningful too". Journal of Linguistics 9:261–270.

Borer, Hagit. 1984. Parametric syntax: Case studies in Semitic and Romance languages. Dordrecht: Foris.

Boškavić, Željko. 2009. "Unifying first and last conjunct agreement". Natural Language and Linguistic Theory 27:455–496.

Bošković, Željko. 2007. "On the Locality and Motivation of Move and Agree: An Even More Minimal Theory". Linguistic Inquiry 38:589–644.

Bošković, Željko. 2018. "On The Coordinate Structure Constraint, Across-the-Boad-Movement, Phases, and Labeling". Ms. Available at https://ling.auf.net/lingbuzz/003894.

Bošković, Željko, and Jairo Nuñes. 2007. "the copy theory of movement: A view from pf", volume The Copy Theory of Movement, 13–74. Amsterdam/Philadelphia: John Benjamins.

Branigan, Phil, and Marguerite Mackenzie. 2002. "Altruism, A'-movement, and object agreement in Innu-aimûn". Linguistic Inquiry 33:385–407.

Bresnan, Joan, and Sam A. Mchombo. 1987. "On Topic, Pronoun, and Agreement in Chichewa". Language 63:741–782.

Broadwell, G., M. Butt, and H. King. 2006. "Alignment, Precedence, and the Typology of Pied-Piping with Inversion". In Proceedings of the LFG06 Conference, ed. M. Butt and H. King. Stanford, CA: CSLI Publications.

Bruening, B. 2008. "Readme for Scope Elicitation". *Materials from the Scope Fieldwork Project*: http://udel.edu/∼bruening/scopeproject/materials.html.

Bruening, Benjamin. 2004. "Two types of *wh*-scope marking in Passamaquoddy". Natural Language and Linguistic Theory 22:229–305.

Bruening, Benjamin. 2006. "Differences between the *Wh*-Scope-Marking and *Wh*-Copy Constructions in Passamaquoddy". Linguistic Inquiry 37:24–49.

Bruening, Benjamin. 2014. "Precede-and-command revisited". Language 90:342–388.

Burzio, Luigi. 1986. Italian Syntax: A Government and Binding Approach. Studies in Natural Language and Linguistic Theory. Reidel.

Cable, Seth. 2007. The grammar of Q: Q-particles and the nature of *wh*-fronting, as revealed by the *wh*-questions of Tlingit. Doctoral Dissertation, MIT, Cambridge, MA.

Cable, Seth. 2010. "Against the Existence of Pied-Piping: Evidence from Tlingit". Linguistic Inquiry 41:563–594.

Cable, Seth. 2012. "Pied-Piping: Introducing Two Recent Approaches". Language and Linguistics Compass 6:816–832.

Cable, Seth. 2013. "Pied-Piping: Comparing Two Recent Approaches". Language and Linguistics Compass 7:123–140.

Campbell, Lyle, Terrence Kaufman, and Thomas Smith-Stark. 1986. "Meso-America as a Linguistic Area". Language 62:530–570.

Carlson, Gregory. 1977. Reference to Kinds in English. Doctoral Dissertation.

Carnie, Andrew. 1995. Non-Verbal Predication and Head Movement. Doctoral Dissertation, MIT.

Chafe, Wallace (ed.). 1980. The pear stories: Cognitive, cultural, and linguistic aspects of narrative production, volume III of Advances in Discourse Processes. Ablex.

Chafe, William. 1976. "Givenness, Contrastiveness, Definiteness, Subbjects, and Topics", volume Subject and Topic, chapter 2, 25–56. San Francisco: Academic Press, Inc.

Charnavel, Isabelle, and Dominique Sportiche. 2016. "Anaphor Binding: What French Inanimate Anaphors Show". Linguistic Inquiry 47:35–87.

Chomsky, Noam. 1980. "On binding". Linguistic Inquiry 11:1–46.

Chomsky, Noam. 1981. Lectures on government and binding. Dordrecht: Foris.

Chomsky, Noam. 1986. Barriers. Cambridge, MA: MIT Press.

Chomsky, Noam. 1993. "A Minimalist Program for Linguistic Theory". In The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger, ed. K. Hale and S. J. Keyser, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. "Bare phrase structure", volume Government and Binding Theory and the Minimalist Program of Generative Syntax, 383–439. Blackwell.

Chomsky, Noam. 2000. "Minimalist Inquiries: The Framework", volume Step by step: Essays on minimalist syntax in honor of Howard Lasnik, 89–156. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. "Derivation by phase", volume Ken Hale: A life in language, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2008. "On Phases", volume Foundational Issues in Linugistic Theory: Essays in Honor of Jean-Roger Vergnaud. Cambridge, MA: MIT Press.

Chomsky, Noam, and Howard Lasnik. 1993. The Minimalist Program, chapter "The Theory of Principles and Parameters". MIT Press.

Chung, Sandra. 1994. "Wh-Agreement and "Referentiality" in Chamorro". Linguistic Inquiry 25:1–44.

Cinque, Guglielmo. 1990. Types of Ā-Dependencies. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 1999. Adverbs and functional heads: A cross-linguistic perspective. Oxford Studies in Comparative Syntax. Oxford University Press.

Clamons, Robbin, Ann E. Mulkern, Gerald Sanders, and Nancy Stenson. 1999. "The Limits of Formal Analysis: Pragmatic motivation in Oromo grammar", volume Functionalism and Formalism in Linguistics: Volume II: Case studies, 59–76. John Benjamins Publishing Company.

Cole, Peter. 1982. Imbabura Quechua. London: Dover.

Collins, Chris, and Höskuldur Thráinsson. 1996. "VP-Internal Structure and Object Shift in Icelandic". Lingustic Inquiry 27:391–444.

Coon, Jessica. 2009. "Interrogative Possessors and the Problem with Pied-Piping in Chol". Linguistic Inquiry 40:165–175.

Coon, Jessica, and Alan Bale. 2014. "The interaction of person and number of Mi'gmaq". Nordlyd 40:85–101.

Corbett, Greville. 1995. Syntax: An International Handbook of Contemporary Research, volume 2, 1235–1244. Berlin: De Gruyter Mouton.

Corbett, Greville. 2006. Agreement. Cambridge, England: Cambridge University Press.

van Craenenbroeck, Jeroen, and Marjo van Koppen. 2008. "Pronominal doubling in Dutch dialects: big DPs and coordinations", volume Microvariation in syntactic doubling of Syntax and Semantics, 207–249. Bingley: Emerald.

Crasborn, Onno, Els van der Kooij, Johan Ros, and Helen de Hoop. 2009. "Topic agreement in NGT (Sign Language of the Netherlands)". The Linguistic Review 26:355–370.

Dahl, Östen. 1985. Tense and Aspect Systems. Basil Blackwell Ltd.

D'Alessandro, Roberta. 2004. Impersonal *si* constructions: Agreement and interpretation. Doctoral Dissertation, Universität Stuttgart.

D'Alessandro, Roberta. 2017. "Agreement stacking and topic oriented agreement: how to agree with adverbs, gerunds, and prepositions". Manuscript available at http://ling.auf.net/lingbuzz/003563.

Dalrymple, Mary, and Irina Nikolaeva. 2005. "nonsubject agreement and discourse roles", volume Oxford University Working Papers in Linguistics, Philology & Phonetics, 71–91. Oxford: Oxford University Press.

Daly, John, and Larry Hyman. 2007. "On the Representation of Tone in Peñoles Mixtec". International Journal of American Linguistics 73.

Davies, William. 1986. Choctaw Verb Agreement and Universal Grammar, volume 2 of Studies in Natural Language and Linguistic Theory. Dordrecht.

Davison, Alice. 1984. "Syntactic Markedness and the Definition of Sentence Topic". Language 60:797–846.

Dayal, Veneeta. 1994. "Scope marking as indirect wh-dependency". Natural Language Semantics 2:137–170.

Dayal, Veneeta. 2000. "scope marking: Cross-linguistic variation in indirect dependency", 157–193. John Benjamins Publishing Company.

Deal, Amy Rose. 2013. "Possessor Raising". Linguistic Inquiry 44:391–432.

Deal, Amy Rose. 2015a. "Interaction and satisfaction in φ-agreement". In NELS 45, ed. T. Bui and D. Ozyildiz, 1–14.

Deal, Amy Rose. 2015b. "Reasoning about equivalence in semantic fieldwork", volume Methodologies in Semantic Fieldwork, 157–174. Oxford University Press.

Deal, Amy Rose. 2016. "Syntactic Ergativity: Analysis and Identification". Annual Review of Linguistics 2:165–185.

DeLancey, Scott. 1981. "An Interpretation of Split Ergativity and Related Patterns". Language 57:626–657.

DeLancey, Scott. 1985. "On active typology and the nature of agentivity", volume Relational typology, 47–60. Berlin: Mouton.

DiCanio, Christian. 2008. The Phonetics and Phonology of San Martín Itunyoso Trique. Doctoral Dissertation, University of California, Berkeley.

DiCanio, Christian, and Ryan Bennett. To appear. "Prosody in Mesoamerican Languages". Oxford University Press.

Diesing, Molly. 1992. Indefinites. Number 20 in Linguistic Inquiry Monograph. MIT Press.

Diesing, Molly. 1997. "Yiddish VP Order and the Typology of Object Movement in Germanic". Natural Language and Linguistic Theory 15:369–427.

Dik, S. C. 1978. Functional grammar. Amsterdam: Foris.

den Dikken, Marcel. 1995. "Binding, Expletives, and Levels". Linguistic Inquiry 26:347–354.

Dixon, R. M. W. 1979. "Ergativity". Language 55:59–138.

Dobrovie-Sorin, Carmen. 1990. "Clitic Doubling, Wh-Movement, and Quantification in Romanian". Linguistic Inquiry 21:351–397.

Dobrovie-Sorin, Carmen. 1994. The syntax of Romanian. Comparative Studies in Romance. Berlin: Mouton de Gruyter.

Donohue, Mark. 2001. "Split Intransitivity and Saweru". Oceanic Linguistics 40:321–336.

Doron, Edit. 1982. "The Syntax and Semantics of Resumptive Pronouns". Texas Linguistics Forum 19:1–48.

Downing, Angela. 1991. "An alternative approach to theme: A systematic-functional perspective". Word 42:119–143.

Dryer, Matthew. 1994. "The discourse function of the Kutenai inverse", volume Voice and Inversion of Typological Studies in Language, 65–99. Amsterdam: John Benjamins.

Dunn, Michael John. 1999. A Grammar of Chukchi. Doctoral Dissertation, Australian National University.

Durie, Mark. 1985. A Grammar of Acehnese on the basis of a dialect of North Aceh. Number 112 in Koninklijk Instituut voor Taal-, Land- en Volkenkunde. Foris.

Eberhardt, R. 1999. "Questions and Inversion in Ocotepec Mixtec". Work Papers of the Summer Institute of Linguistics, University of North Dakota Session 43.

Egland, Steven, and Doris Bartholomew. 1983. "La Inteligibilidad Interdialectal en México: Resultados de Algunos Sondeos" .

Eilam, Aviad. 2008. "Intervention Effects: Why Amharic Patterns Differently". In Proceedings of the 27th West Coast Conference on Formal Linguistics, ed. N. Abner and J. Bishop, 141–149. Somverville, MA: Cascadilla Press.

Elfner, Emily. 2012. Syntax-prosody interactions in Irish. Doctoral Dissertation, University of Massachusetts, Amherst.

Elordieta, Gorka. 1997. Issues in the phonology and morphology of the major iberian languages, chapter Accent, tone and intonation in Lekeitio Basque, 3–78. Washington, D.C.: Georgetown University Press.

Elordieta, Gorka. 2007. "Minimum size constraints on Intermediate Phrases". Presentation at the XVI International Congress on Phonetic Sciences, Universität Saarland, Germany, August 2007.

Engdahl, Elisabet. 1983. "Parasitic Gaps". Linguistics and Philosophy 6:5–34.

Fanselow, Gisbert, and Denisa Lenertová. 2011. "Left peripheral focus: Mismatches between syntax and information structure". Natural Language and Linguistic Theory 29:169–209.

Fanselow, Gisbert, and Anoop Mahajan. 2000. "towards a minimalist theory of wh-expletives, wh-copying, and successive cyclicity", volume Wh-scope marking of Linguistik Aktuell/Linguistics Today. Amsterdam/Philadelphia: Benjamins.

Farkas, Donka, and Adrian Brasoveanu. 2013. "A Typology of Specificity". Revue Roumaine de Linguistique .

Farris, E. 1992. "A Syntactic Sketch of Yosondúa Mixtec", volume 4 of Studies in the Syntax of Mixtecan Languages, 1–172. Summer Institute of Linguistics of the University of Texas at Arlington.

Foris, David. 2000. A Grammar of Sochiapan Chinantec. Number 6 in Studies in Chinantec Languages. SIL International.

Frank, Paul Stephen. 1990. A Grammar of Ika, volume 94 of University of Texas at Arlington Publications in Linguistics. SIL International.

Franks, Steven, and Catherine Rudin. 2005. "Bulgarian Clitics as K^{o} Heads". In Formal Approaches to Slavic Linguistics: The South Carolina Meeting, ed. S. Franks, F. Gladney, and M. Tasseva-Kurktchieva, 106-118. Ann Arbor, MI: Michigan Slavic Publications.

Frascarelli, Mara, and Roland Hinterhölzl. 2007. "Types of Topics in German and Italian", volume *On Information Structure, Meaning and Form*, 87–116. Amsterdam: John Benjamins.

Freidin, Robert, and Howard Lasnik. 1981. "Disjoint reference and wh-trace". Linguistic Inquiry 12:39–53.

Fuß, Eric. 2005. The Rise of Agreement: A formal approach to the syntax and grammaticization of verbal inflection. Number 81 in Linguistik Aktuell. John Benjamins Publishing Company.

Gerfen, Henry. 1996. Topics in the Phonology and Phonetics of Coatzospan Mixtec. Doctoral Dissertation, The University of Arizona.

Givón, Talmy. 1971. "Historical syntax and synchronic morphology: An archaeologist's field trip". In Papers from the Seventh Regional Meeting Chicago Linguistics Society, 394–415. Chicago, IL: Chicago Linguistics Society.

Givón, Talmy. 1976. "Topic, pronoun and grammatical agreement", volume Subject and Topic, chapter 6, 149–188. San Francisco: Academic Press, Inc.

Givón, Talmy. 1983. Topic continuity in discourse. Amsterdam: John Benjamins.

Goddard, Ives. 1984. "The obviative in Fox narrative discourse". In Papers of the Fifteenth Algonquian Conference, ed. W. Cowan, 273–286. Ottawa: Carelton University.

Goddard, Ives. 1990. "Aspects of the Topic Structure of Fox Narratives: Proximate Shifts and the Use of Overt and Inflectional NPs". International Journal of American Linguistics 56:317–340.

Grimshaw, Jane. 2000. "Locality and Extended Projection", volume Lexical Specification and Insertion of IV - Current Issues in Linguistic Theory. John Benjamins Publishing Company.

Grimshaw, Jane. 2005. Words and Structure. Number 151 in CSLI Lecture Notes. Stanford, CA: CSLI Publications.

Grohmann, Kleanthes, and Phoevos Panagiotidis. 2004. "Demonstrative Doubling". University of Maryland Working Papers in Linguistics 13:110–132.

Gutiérrez-Rexach, Javier. 1999. "The formal semantics of clitic doubling". Journal of Semantics 16:315–380.

Hacquard, Valentine. 2011. "Modality", volume Semantics: An International Handbook of Natural Language Meaning of HSK 33.2, 1484–1515. Berlin: Mouton de Gruyter.

Haegeman, Liliane. 1994. Introduction to Government and Binding Theory. Number 1 in Blackwell Textbooks in Linguistics. Cambridge, MA: Blackwell Publishers, second edition.

Halle, Morris, and Alec Marantz. 1993. "Distributed morphology and the pieces of inflection", volume The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger, 111–176. Cambridge, MA: MIT Press.

Halle, Morris, and Alec Marantz. 1994. "Some Key Features of Distributed Morphology*". MIT Working Papers in Linguistics 21 275–288.

Halpert, Claire. 2012. Argument licensing and agreement in Zulu. Doctoral Dissertation, MIT.

Harizanov, Boris. 2014a. "Clitic doubling at the syntax-morphophonology interface: A-movement and morphological merger in Bulgarian". Natural Language and Linguistic Theory 32:1033–1088.

Harizanov, Boris. 2014b. On the Mapping from Syntax to Morphophonology. Doctoral Dissertation, University of California, Santa Cruz.

Harizanov, Boris. 2016. "Head movement to specifier positions in Bulgarian participle fronting". Presented at 90th Annual Meeting of the Linguistic Society of America. Handout available at https://stanford.edu//*sim*bharizan/pdfs/Harizanov_2016_LSA handout.pdf.

Harizanov, Boris, and Vera Gribanova. 2017. Whither Head Movement? Ms.

Haspelmath, Martin. 1999. "External possession in a European areal perspective", volume External possession, 109–135. Amsterdam: John Benjamins.

Heath, Jeffrey. 1977. "Choctaw cases". In Proceedings of the third annual meeting of the Berkeley Linguistics Society, ed. K. Whistler, 204–213. Berkeley, CA: Berkeley Linguistics Society.

Heck, Fabian, and Malte Zimmermann. 2004. "Remarks on the Phasal Status of DP". Ms., Universität Tübingen, Humboldt-Universität Berlin.

Heim, Irene. 1982. The Semantics of Definite and Indefinite Noun Phrases. Doctoral Dissertation, UMass Amherst.

Henderson, Robert. 2012. "Morphological alternations at the intonational phrase edge". Natural Language and Linguistic Theory 30:741–787.

Hermon, Gabriella. 1984. Syntactic modularity. Dordrecht: Foris.

Hiraiwa, Ken. 2002. "multiple agree". GLOW-TiLT workshop, Utrecht.

Hiraiwa, Ken. 2005. Dimensions of Symmetry in Syntax: Agreement and Clausal Architecture. Doctoral Dissertation, MIT.

Hock, Hans Henrich, and Brian D. Joseph. 1996. Language history, language change, and language relationship: An introduction to historical and comparative linguistics. Walter de Gruyter.

Hoeksema, Jarich. 1995. "Preposition Stranding and Resumptivity in West Germanic", volume Studies in Comparative Germanic Syntax, 95–118. Kluwer Academic Publishers.

Hollenbach, Barbara. 2015. "A preliminary reconstruction of Mixtec Pronouns". Available at http://barbaraelenahollenbach.com/PDFs/PRONRECN.pdf.

Holmberg, Anders. 1986. Word order and syntactic features in Scandinavian languages. Doctoral Dissertation, Universitet Stockholms.

Holmberg, Anders, and Urpo Nikanne. 2008. "subject doubling in finnish: The role of deficient pronouns", 325–349. Amsterdam: Brill.

Hopper, Paul, and Elizabeth Traugott. 1993. Grammaticization. Cambridge, England: Cambridge University Press.

Horvath, Julia. 2010. ""Discourse features", syntactic displacement and the status of contrast". Lingua 120:1346–1369.

Huang, C-T. James. 1982. Logical Relations in Chinese and the Theory of Grammar. Doctoral Dissertation.

Huang, C.-T James, Y.-H Audrey Li, and Yafei Li. 2009. The Syntax of Chinese. Cambridge Syntax Guides. Cambridge University Press.

Jackendoff, Ray. 1977. X-bar syntax: A study of phrase structure. Cambridge, MA: MIT Press.

Jaeggli, Osvaldo. 1982. Topics in Romance Syntax. Dordrecht, Holland: Foris Publications.

Jäger, Gerhard. 2001. "Topic-Comment Structure and the Contrast Between Stage Level and Individual Level Predicates". Journal of Semantics 18:83–126.

Jelinek, Eloise. 1984. "Empty Categories, Case, and Configurationality". Natural Language and Linguistic Theory 2:39–76.

Johnson, Audrey. 1988. "A Syntactic Sketch of Jamiltepec Mixtec", volume 3 of Studies in the Syntax of Mixtecan Languages, 11–150. Summer Institute of Linguistics of the University of Texas at Arlington.

Josserand, Judy. 1983. Mixtec dialect history. Doctoral Dissertation, Tulane University.

Josserand, Judy Kathryn, Marcus Winter, and Nicholas Hopkins, ed. 1984. Essays in otomanguean culture history, volume 31 of Publications in Anthropology. Nashville, Tennessee: Vanderbilt University.

Kallulli, Dalina. 2000. "Direct object clitic doubling in Albanian and Greek", volume Clitic Phenomena in European Languages of Linguistik Aktuell, 209–248. John Benjamins.

Kallulli, Dalina. 2016. "Clitic doubling as Differential Object Marking". Rivista di Grammatica Generativa 38:161–171.

Kandybowicz, Jason. 2006. Conditions on multiple spell-out and the syntax-phonology interface. Doctoral Dissertation, UCLA.

Kaufman, Terrence, and John Justeson. 2009. "Historical Linguistics and Pre-Columbian Mesoamerica". Ancient Mesoamerica 20:221–231.

Kayne, Richard. 1975. French Syntax: The Transformational Cycle. Cambridge, MA: MIT Press.

Kiss, Katalin. 2002. The Syntax of Hungarian. Cambridge Syntax Guides. Cambridge, England: Cambridge University Press.

Koopman, Hilda. 1984. The syntax of verbs: From verb movement rules in the kru languages to universal grammar, volume 15 of Studies in Generative Grammar. Dordrecht, Holland: Foris Publications.

Kramer, Ruth. 2009. Definite markers, phi-features, and agreement: A morphosyntactic investigation of the Amharic DP. Doctoral Dissertation, UC Santa Cruz.

Kramer, Ruth. 2014. "Clitic doubling or object agreement: the view from Amharic". Natural Language and Linguistic Theory 32:593–634.

Krapova, Iliyana. 1998. "Subjunctive Complements, Null Subjects and Case Checking in Bulgarian". University of Venice Working Papers in Linguistics 8.

Krapova, Iliyana. 2001. Comparative Syntax of Balkan Languages, chapter "Subjunctives in Bulgarian and Modern Greek", 105–126. Oxford University Press.

Kratzer, Angelika. 1995. "Stage-Level and Individual-Level Predicates", volume The Generic Book, 125–175. Chicago: Chicago University Press.

Krifka, Manfred. 2008. "Basic notions of information structure". Acta Linguistica Hungarica 55:243–276.

Kuiper, Albertha, and Joy Oram. 1991. "a syntactic sketch of diuxi-tilantongo mixtec", volume 3 of Studies in the Syntax of Mixtecan Languages, 179–408. Summer Institute of Linguistics of the University of Texas at Arlington.

Kuno, Susumu. 1976. "Subject raising", volume Symtax and semantics 5: Japanese generative grammar, 17–49. New York: Academic Press.

Kuroda, S.-Y. 1972. "The Categorical and the Thetic Judgment: Evidence from Japanese Syntax". Foundations of Language 9:153–185.

Ladusaw, Bill. 1994. "Thetic and Categorical, Stage and Individual, Weak and Strong". In SALT IV, ed. M. Harvey and L. Santelmann, 220–229. Ithaca, NY: Cornell University.

Lambrecht, Knud. 1981. Topic, Antitopic and Verb Agreement in Non-Standard French. Amsterdam: John Benjamins.

Landau, Idan. 2000. Elements of control: Structure and meaning in infinitival constructions. Studies in Natural Language and Linguistic Theory. Kluwer Academic Publishers.

Landau, Idan. 2004. "The Scale of Finiteness and the Calculus of Control". Natural Language and Linguistic Theory 22:811–877.

Landau, Idan. 2006. "Chain Resolution in Hebrew V(P)-Fronting". Syntax 9:32–66.

Landau, Idan. 2013. Control in Generative Grammar: A Research Companion. Cambridge University Press.

Lasnik, Howard. 1976. "Remarks on coreference". Linguistic Analysis 2:1–22.

Lasnik, Howard, and Tim Stowell. 1991. "Weakest Crossover". Linguistic Inquiry 22:687–720.

Lastra, Yolanda. 1992. El Otomí de Toluca. Universidad Nacional Autónoma de México.

Lastra, Yolanda. 1997. El Otomí de Ixtenco. Universidad Nacional Autónoma de México.

Legate, Julie Anne. 2005. "Phases and Cyclic Agreement". MIT Working Papers in Linguistics 49:147–156.

Legate, Julie Anne. 2011. "Warlpiri *Wh*-Scope Marking". Syntax 14:97–121.

Lehmann, Christian. 1988. Agreement in natural language: Approaches, theory, description, chapter "On the function of agreement", 55–65. Stanford, CA: CSLI Publications.

Li, Charles, and Sandra Thompson. 1976. "subject and topic: A new typology of language", volume Subject and Topic, chapter 15, 457–491. San Francisco: Academic Press, Inc.

Lillehaugen, Brook. 2003. "The Acquisition of Body Part Prepositions in Valley Zapotec". In Proceedings of the Conference on Indigenous Languages of Latin America I. University of Texas at Austin.

Lødrup, Helge. 1991. "The Norwegian Pseudopassive in Lexical Theory". Working Papers in Scandinavian Syntax 47:118–129.

López, Luis. 2009. A derivational syntax for information structure. Oxford: Oxford University Press.

Macaulay, Monica. 1996. A grammar of Chalcatongo Mixtec. Doctoral Dissertation, University of California, Berkeley.

Macaulay, Monica. 2005. "the syntax of chalcatongo mixtec: Preverbal and postverbal", volume Verb First: On the Syntax of Verb Initial Languages, 341–366. Linguistik Aktuell.

Macaulay, Monica, and Joseph Salmons. 1995. "The Phonology of Glottalization in Mixtec". International Journal of American Linguistics 61:38–61.

Mahajan, Anoop. 1997. "Rightward scrambling", volume Rightward movement, 185–213. Amsterdam: John Benjamins.

Maling, Joan, and Annie Zaenen. 1985. "Preposition-Stranding and Passive". Syntax and Semantics 24:153–164.

Manetta, Emily. 2006. Peripheries in Kashmiri and Hindi-Urdu. Doctoral Dissertation, University of California, Santa Cruz.

Manetta, Emily. 2010. "*wh*-expletives in hindi-urdu: the vp phase". Linguistic Inquiry 41.

Marantz, Alec. 1991. Proceedings of ESCOL '91, chapter "Case and Licensing", 234–253. Cornell Linguistics Club.

Marlett, Stephen. 1992. "Nasalization in Mixtec Languages". International Journal of American Linguistics 58:425–435.

Martinović, Martina. 2015. Feature Geometry and Head-Splitting: Evidence from the morphosyntax of the Wolof clausal periphery. Doctoral Dissertation, University of Chicago.

Marušič, Franc, Andrew Nevins, and Amanda Saksida. 2007. "Linear agreement in Slovene conjuncts". In Formal Approaches to Slavic Linguistics 15 (The Toronto Meeting), ed. R. Compton, M. Goledzinowska, and U. Savchenko, 210–227. Ann Arbor, MI: Michigan Slavic Publications.

Matthewson, Lisa. 2004. "On the Methodology of Semantic Fieldwork". International Journal of American Linguistics 70:369–415.

Matushansky, Ora. 2006. "Head-movement in linguistic theory". Linguistic Inquiry 37.

McCloskey, Jim. 1977. "An Acceptable Ambiguity in Modern Irish". Linguistic Inquiry 8:604–609.

McCloskey, Jim. 1986. "Inflection and Conjunction in Modern Irish". NLLT 4:246–281.

McCloskey, Jim. 1990. "resumptive pronouns, a'-binding and levels of representation in irish", volume Syntax of the modern Celtic languages of Syntax and Semantics, 199–248. New York and San Diego: Academic Press.

McCloskey, Jim. 2002. "resumption, successive cyclicity, and the locality of operations", volume Derivation and Explanation, 184–226. Blackwell Publishers.

McCloskey, Jim. To appear. "Resumption". The Wiley Blackwell Companion to Syntax. Available at http://ohlone.ucsc.edu/∼jim/PDF/syncom2.pdf: Wiley-Blackwell, 2nd edition edition.

McCloskey, Jim, and Ken Hale. 1984. "On the Syntax of Person-Number Inflection in Modern Irish". Natural Language and Linguistic Theory 1:487–533.

McDaniel, Dana. 1986. Conditions on Wh-chains. Doctoral Dissertation, City University of New York, CUNY, New York.

McDaniel, Dana. 1989. "Partial and multiple Wh-movement". Natural Language and Linguistic Theory 565–604.

McGinnis, Martha. 2008. "Applicatives". Language and Linguistics Compass 2:1225–1245.

McNally, Louise. 1998. "On Recent Formal Analyses of Topic", volume *The Tbilisi Symposium on Language, Logic, and Computation: Selected Papers*, 147–160. Stanford, CA: CSLI Publications.

Meacham, Michael. 1991. "The phonetics of tone in Chalcatongo Mixtec couplets". Occassional Papers in Linguistics: Papers from the American Indian Languages Conferences, Held at the University of Californa, Santa Cruz, July and August 1991 156–167.

Merlan, Francesca. 1985. "split intransitivity: Functional oppositions in intransitive inflection", volume Grammar inside and outside the clause, 324–362. Cambridge, MA: Cambridge University Press.

Mikkelsen, Line. 2015. VP anaphora and verb-second order in Danish. Journal of Linguistics 51:595–643.

Miller, Philip, and Ivan Sag. 1997. "French Clitic Movement without Clitics or Movement". Natural Language and Linguistic Theory 15:573–639.

Mithun, Marianne. 1991. Active/agentive Case Marking and Its Motivations. Language 67:510–546.

Miyagawa, Shigeru. 2010. Why agree? why move? unifying agreement-based and discourse configurational languages. Number 54 in Linguistic Inquiry Monograph. Cambridge, MA: MIT Press.

Miyagawa, Shigeru. 2017. Agreement Beyond Phi. Number 75 in Linguistic Inquiry Monograph. Cambridge, MA: MIT Press.

Mock, Carol. 1982. "Los Casos Morfosintacticos del Chocho". Anales de Antropologí (Instituto de investigaciónes Antropologicas, UNAM) 19:345–378.

Mohanan, Tara. 1995. Argument structure in Hindi. Stanford, CA: CSLI Publications.

Morimoto, Yukiko. 2006. "Agreement properties and word order in comparative Bantu", volume Papers in Bantu grammar and description of (ZAS Papers in Linguistics 43), 161–187. Berlin: Zentrum für Allgemeine Sprachwissenschaft.

Munn, Alan. 1993. "Topics in the Syntax and Semantics of Coordinate Structures". Doctoral Dissertation, University of Maryland.

Munn, Alan. 1999. "First conjunct agreement: Against a clausal analysis". Linguistic Inquiry 30:643–668.

Myhill, John. 1985. "Pragmatic and categorical correlates of VS word order". Lingua 66:177–200.

Nespor, Marina, and Irene Vogel. 1986. Prosodic Phonology. Dordrecht: Foris Publications.

Nevins, Andrew. 2011. "Multiple agree with clitics: person complementarity vs. omnivorous number". Natural Language and Linguistic Theory 29:939–971.

Nichols, Johanna. 1986. "Head-Marking and Dependent-Marking Grammar". Language 62:56–119.

North, Joanne, and Jäna Shields. 1977. Studies in Otomanguean Phonology, chapter "Silacayoapan Mixtec Phonology", 21–34. 54. Summer Institute of Linguistics of the University of Texas at Arlington.

Nunes, Jairo, and Ronice Müler de Quadros. 2004. "Phonetic Realization of Multiple Copies in American and Brazilian Sign Languages". Paper presented at the Theoretical Issues in Sign Language Research 8. Barcelona.

Ostrove, Jason. 2017a. "Severing PRO From its Silence: Control Constructions in San Martín Peras Mixtec". Talk at the 22nd Workshop on Structure and Constituency in Languages of the Americas, Univeristy of British Columbia. Handout available at https://people.ucsc.edu/∼jostrove/wscla2017.pdf, April 22nd 2017.

Ostrove, Jason. 2017b. "The Morphology-Prosody Interface in San Martín Peras Mixtec". Presentation at the Chicago Linguistics Society 53. Handout available at https://people.ucsc.edu/∼jostrove/cls2017.pdf, May 25th 2017.

Palancar, E., J. Amith, and R. Castillo García. 2015. "Verbal inflection in Yoloxóchitl Mixtec", volume Tone and Inflection. Available at: https://hal.archives-ouvertes.fr/hal-01099329.

Palancar, Enrique. 2009. Gramática y textos del hñöñhö: Otomí de san ildefonso tultepec, querétaro, volume Volumen 1: Gramática. Mexico City: Universidad Autónoma de Quetétaro.

Pearson, Hazel. 2013. The Sense of Self: Topics in the Semantics of De Se Expressions. Doctoral Dissertation, Harvard University.

Pearson, Hazel. 2015. "The interpretation of the logophoric pronoun in Ewe". Natural Language Semantics 23:77–118.

Perlmutter, David. 1978. "Impersonal Passives and the Unaccusative Hypothesis". In Proceedings of the Annual Meeting of the Berkeley Linguistics Society, volume 38, 157–189. Linguistic Society of America.

Pesetsky, David. 1987. "Wh-in-Situ: Movement and Unselective Binding", volume The Representation of (In)definiteness. Cambridge, MAh: MIT Press.

Pesetsky, David, and Esther Torrego. 2007. "the syntax of valuation and the interpretability of features", volume Phasal and Clausal Architecture: Syntactic Derivation and Interpretation, 262–294. Amsterdam: John Benjamins.

Pike, Eunice, and John Cowan. 1967. "Huajuapan Mixtec Phonology and Morphophonemics". Anthropological Linguistics 9:1–15.

Pike, Kenneth. 1948. Tone Languages. University of Michigan Press.

du Plessis, Hans. 1977. "Wh Movement in Afrikaans". Linguistic Inquiry 8:723–726.

Poletto, Cecilia, and Jean-Yves Pollock. 2004. "On *wh*-clitics and *wh*-doubling in French and some North Eastern Italian Dialects". Probus 241–272.

Polinsky, Maria. 2016. Deconstructing ergativity: Two types of ergative languages and their features. Oxford University Press.

Polinsky, Maria, and Eric Potsdam. 2001. "Long-Distance Agreement and Topic in Tsez". Natural Language and Linguistic Theory 19:583–646.

Pollock, Jean-Yves. 1989. "Verb Movement, Universal Grammar, and the Structure of IP". Linguistic Inquiry 20:365–424.

Postal, Paul. 1971. Cross-Over Phenomena. Holt, Rinehart and Winston.

Preminger, Omer. 2009. "Breaking Agreements: distinguishing agreement and clitic-doubling by their failures". Linguistic Inquiry 40:619–666.

Preminger, Omer. 2014. Agreement and Its Failures. Number 68 in Linguistic Inquiry Monograph. Cambridge, MA: MIT Press.

Preminger, Omer. to appear. "What the PCC tells us about "abstract" agreement, head movement, and locality". Glossa Available at: http://ling.auf.net/lingbuzz/003221.

Reglero, Lara. 2007. "Wh-in-situ interrogatives in Spanish". Probus 19:267–297.

Reglero, Lara, and Emma Ticio. 2015. "no superiority, no intervention effects: The spanish puzzle", volume Hispanic Linguistics at the Crossroads: Theoretical linguistics, language acquisition, and language contact of Issues in Hispanic and Lusophone Linguistics, 3–24. John Benjamins Publishing Company.

Reinhart, Tanya. 1981. "Pragmatics and Linguistics: An Analysis of Sentece Topics". Philosophica 27:53–94.

Reinhart, Tanya. 1983. Anaphora and semantic interpretation.. Chicago, IL: University of Chicago Press.

Rezac, Milan. 2003. "The fine structure of Cyclic Agree". Syntax 6:156–182.

Rezac, Milan. 2004. Elements of Cyclic Syntax: Agree and Merge. Doctoral Dissertation, University of Toronto.

Rezac, Milan. 2010. "ϕ-Agree versus ϕ-feature movement: Evidence from floating quantifiers". Linguistic Inquiry 41:496–508.

Rizzi, Luigi. 1997. "The Fine Structure of the Left Periphery" "The Fine Structure of the Left Periphery", volume Elements of Grammar. Kluwer Publications.

Roberts, Ian. 2010. Agreement and head movement: Clitics, incorporation, and defective goals. Number 59 in Linguistic Inquiry Monograph. Cambridge, MA: MIT Press.

Roberts, Ian, and Anna Roussou. 2003. Syntactic Change: A Minimalist Approach to Grammaticization. Cambridge University Press.

Rooth, Mats. 1992. "A Theory of Focus Interpretation". Natural Language Semantics 1:75–116.

Ross, J. 1967. Constraints on Variables in Syntax. Doctoral Dissertation, MIT.

Ruangjaroon, Sugunya. 2002. "Thai Wh in-situ". Talk presented at NWCL, Simon Fraser University, April 6-7.

Samko, Bern. 2016. Syntax & Information Structure: The Grammar of English Inversions. Doctoral Dissertation, UC Santa Cruz.

Scancarelli, 1987. 1987. Grammatical relations and verb agreement in Cherokee. Doctoral Dissertation, UCLA.

Selkirk, Elizabeth. 1984. Phonology and Syntax: The Relation Between Sound and Structure. MIT Press.

Selkirk, Elizabeth. 1995. "The Prosodic Structure of Function Words."', volume *Signal to Syntax*.

Selkirk, Elizabeth. 2009. "On clause and intonational phrase in Japanese: The syntactic grounding of prosodic constituent structure". *Gengo Kenkyu* (Journal of the Linguistics Society of Japan) Special Issue on Linguistic Interfaces ed. by Haruo Kubozono:1–39.

Selkirk, Elizabeth. 2011. "The Syntax-Phonology Interface", volume The Handbook of Phonological Theory. Oxford: Blackwell, 2nd edition edition.

Shields, Jäna. 1988. "A Syntactic Sketch of Silacayoapan Mixtec", volume 1 of Studies in the Syntax of Mixtecan Languages. Summer Institute of Linguistics of the University of Texas at Arlington.

Shlonsky, Ur. 1992. "Resumptive pronouns as a last resort". Linguistic Inquiry 23:443–468.

Smith, Thomas, and Fermin Tapia. 2002. "Amuzgo como lengua activa", volume Del Cora al Maya Yucateco: Estudios lingüisticos sobre algunas lenguas indigenas mexicanas. Mexico City: Universidad Nacional Autónoma de México.

Smith-Stark, T. 1988. "'Pied-Piping' con inversion en preguntas parciales". Ms. Centro de estudios lingüisticos y literarios, Colegio de México y Seminario de lenguas indígenas .

Soltan, Usama. 2006. Agreement systems, chapter "Standard Arabic subject-verb agreement asymmetry revisited in an Agree-based minimalist syntax", 239–265. Amsterdam: John Benjamins.

Sportiche, Dominique. 1996. "Clitic Constructions", volume Phrase Structure and The Lexicon, 213–276. IULC Press.

Stalnaker, Robert. 2002. "Common Ground". Linguistics and Philosophy 25:701–721.

Strawson, P. F. 1964. "Identifying reference and truth-values". Theoria 30:96–118.

Stump, Gregory, and Ramawatar Yadav. 1988. "Maithili verb agreement and the Control Agreement Principle". In Papers from the 24th Regional Meeting of the Chicago Linguistics Society, Part 2: Parasession on Agreement in Grammatical Theory, 304–321.

Suñer, Margarita. 1976. "Demythologizing the Impersonal "Se" in Spanish". Hispania 59:268–275.

Suñer, Margarita. 1988. "The Role of Agreement in Clitic-doubled Constructions". Natural Language and Linguistic Theory 391–434.

Suñer, Margarita. 2000. "Object-shift: Comparing a Romance language to Germanic". Probus 261–289.

Suárez, José. 1983. The Mesoamerican Indian Languages. Cambridge, England:: Cambridge University Press.

Svenonius, Peter. 2004. "On the Edge", volume *Peripheries: Syntactic Edges and their Effects*, 261–287. Kluwer.

Tanaka, Hidekazu. 2002. "Raising to Object out of CP". Linguistic Inquiry 33:637–652.

Terzi, Arhonto. 1992. PRO in Finite Clauses: A Study of the Inflectional Heads of the Balkan Languages. Doctoral Dissertation, CUNY Graduate Center, New York, New York.

Thráinsson, Höskuldur. 2001. "Object shift and scrambling", 148–202. Malden: Blackwell Publishers.

Thráinsson, Höskuldur. 2007. The Syntax of Icelandic. Number 5 in Cambridge Syntax Guides. Cambridge University Press.

Tomlin, Russell, Linda Forrest, Ming Ming Pu, and Myung Hee Kim. 1997. "Discourse Semantics", volume 1 of Discourse as Structure and Process, chapter 3, 63–112. SAGE Publications Ltd.

Tonhauser, Judith, and Lisa Matthewson. 2015. "Empirical evidence in research on meaning". URL `Available at http://ling.auf.net/lingbuzz/002595`.

Toosarvandani, Maziar. 2015. "Vocabulary Insertion and Locality: Verb Suppletion in Northern Paiute". In North East Linguistic Society, volume 46, 247–256. Available at: http://people.ucsc.edu/∼mtoosarv/papers/nels-46-paper.pdf.

Toribio, Almeida Jacqueline. 2000. "Setting parametric limits on dialectal variation in Spanish". Lingua 10:315–341.

Torrego, Esther. 1988. "Pronouns and determiners: A DP analysis of Spanish nominals". Ms.

Torrego, Esther. 1992. "Case and agreement structure". Ms.

Toyoshima, Takashi. 2001. "Head-to-spec movement", volume The minimalist parameter, 115–136. John Benjamins.

Truswell, Robert. 2009. "preposition-stranding, passivisation, and extraction from adjuncts in germanic", volume Linguistic Variation Yearbook. Amsterdam: John Benjamins.

Ortiz de Urbina, J. 1989. Parameters in the grammar of basque: A gb approach to baque syntax. Foris.

Ortiz de Urbina, J. 1993. Generative Studies in Basque Linguistics, chapter "Feature Percolation and Clausal Pied-Piping", 189–219. John Benjamins.

Uriagereka, Juan. 1988. On government. Doctoral Dissertation, University of Connecticut.

Uriagereka, Juan. 1995. "Aspects of the Syntax of Clitic Placement in Western Romance". Linguistic Inquiry 26:79–123.

Uribe-Etxebarria, Myriam. 2002. "In situ questions and masked movement". Linguistic Variation Yearbook 259–303.

van Urk, Coppe. 2015. A uniform syntax for phrasal movement: A case study fo Dinka Bor. Doctoral Dissertation, MIT.

van Urk, Coppe. To Appear. "Pronoun copying in Dinka Bor and the Copy Theory of Movement". Natural Language and Linguistic Theory .

van Valin, Robert. 1990. "Semantic Parameters of Split Intransitivity". Language 66:221–260.

Vallduví, Enric. 1992. The informational component. New York: Garland Press.

Vallduví, Enric, and Elisabet Engdahl. 1996. "The linguistic realization of information packaging". Linguistics 34:459–519.

Vázquez Álvarez, Juan Jesús, and Roberto Zavala Maldonado. 2013. "La estructura argumental preferida en el chol, una lengua agentiva". In Memorias del VI Congreso de Idiomas Indígenas de Latinoamérica. UT Austin.

Veselinova, Ljuba. 2006. Suppletion in verbal paradigms. John Benjamins Publishing Company.

Vicente, Luis Alberto. 2007. The syntax of heads and phrases: A study of verb (phrase) fronting. Doctoral Dissertation, Univeriseit Leiden.

Vikner, Sten. 1990. Verb movement and the licensing of NP positions in the Germanic languages. Doctoral Dissertation, Université de Genéve.

Vikner, Sten. 1994. "Scandinavian Object Shift and West Germanic Scrambling", 487–517. Mouton de Gruyter.

Vikner, Sten. 2005. "Object Shift", volume The Blackwell Companion to Syntax, chapter 46, 392–436. Blackwell Publishers.

Wasow, Thomas. 1972. Anaphoric Relations in English. Doctoral Dissertation, MIT.

Wichmann, Sø ren. 2007. "Case relations in a head-marking language: verb-marked casesin Tlapanec" .

Yuan, Michelle. 2017a. "Diagnosing object agreement vs. clitic doubling: Evidence from Inuktitut". Available at https://ling.auf.net/lingbuzz/003733.

Yuan, Michelle. 2017b. "Why ABS objects are exceptional in Inuktitut: Clitic doubling and cross-dialectal variation". Talk given at the Workshop on the Structure and Constituncy of the Language of the Americas. Handout available at: http://web.mit.edu/yuanm/www/Yuan_camcos_handout.pdf.

Zeller, Jochen. 2006. "Raising out of finite CP in Nguni: the case of *fanele*". Souther African Linguistics and Applied Language Studues 24:255–275.

Zentz, Jason. 2016. Forming *Wh*-Questions in Shona: A Comparative Bantu Perspective. Doctoral Dissertation, Yale University.

Zwicky, A., and G. Pullum. 1986. "The Principle of Phonology-Free Syntax: Introductory Remarks". OSU Working Papers 32:63–91.

Zylstra, C. 1991. Studies in the Syntax of Mixtecan Languages, volume 3, chapter "A Syntactic Sketch of Alantlatzala Mixtec", 1–179. Summer Institute of Linguistics of the University of Texas at Arlington.

Zyman, Erik. 2017. "P'urhepecha hyperraising to object: An argument for purely altruistic movement". Proceedings of the Linguistics Society of America 2:1–15.

Zyman, Erik. 2018. On the Driving Force for Syntactic Movement. Doctoral Dissertation, UC Santa Cruz.