A New Approach to Copy Reflexives

Maxwell Champlin

December, 2024

A Bachelor's Thesis Submitted in Partial Completion of the Degree in Linguistics
Haverford College

HAVERFORD
COLLEGE

Contents

Abstract

This thesis examines copy reflexive constructions which are found in many unrelated languages and appear to violate Condition C of Government and Binding Theory. I show that the copy constructions are a pronominal expression, ω, which is able to explain their distribution and predict their features. There are large gaps in the literature relating to the question of copy constructions; most authors focus on specific languages instead of looking at it as a larger problem. Authors within the Minimalist Program have approached this cross linguistic problem before, but their findings are explained only by theory internal concepts of the Minimalist Program. This thesis shows that previous challenges presented by previous Generative approaches to this issue can be rectified using the pronominal ω. ω is an addition to the tripartite splitting of pronouns into pro-NP, pro-ϕP, and pro-DP. I show that ω must be a fourth pronominal through its distribution and unique feature of co-reference of the highest C-commanding antecedent within its binding domain. I then address the issues of how ω works within the phonology of the languages it is present in, as well I examine the issue of quantification, which has long puzzled researchers in this area of study.

Acknowledgements

I would like to acknowledge everyone who helped me to write this thesis. Starting with Dr. Amanda Payne, who provided invaluable advising to me and helped turn this from an idea in my head into a thesis. I would also like to thank Dr. Brook Lillehaugen for inspiring my interest in this subject some three years ago, her teaching and work on Zapotec has provided this work with perspective on Zapotec Linguistics as a field and a deep relationship with those who speak it. Coco Liu, a peer of mine has been a wonderful person to work through the ideas within this thesis with, her support and challenging questions have allowed for this to shine. And finally, I would like to thank my entire thesis cohort from Haverford and Bryn Mawr Colleges for providing feedback, common cause, and support during the writing of this thesis. Without any of them, this work would not be possible.

1 Introduction

We will address the question of how to handle data wherein an R-expression (DP, some NPs, Proper names) is able to be copied in order to form part of or all of a reflexive construction. On the surface this behavior is a violation of Condition-C of the theory of binding presented in Chomsky (1986), where we have the three following rules which bind the relationships between pronouns and R-expressions.

(1) a. An anaphor is bound in a local domain.
b. A pronominal is free in a local domain.
c. An R-expression is free.

From (1) we will be most concerned with Condition-C. Because as we can see in Zapotec (2) and Thai (3), Zapotec and Thai should be violating Condition-C by having R-expressions, which are bound.

(2) R-yu'lààa'z ra bxuuhahz ra bxuuhahz
HAB-like PL priest PL priest
'The priests like themselves.' Black (1994)

(3) John konnuad John.
John shaved John
'John shaved himself.' Lasnik and Stowell (1991)

Many languages that use pronouns in reflexives will use something like the English, German and Polish examples below:

(4) John shaved himself.

(5) Max rasiert sich
Max shaves REFL
'Max shaves.' De Alencar et al. (2005)

(6) Widziała siebie w lustrze
she.saw self.ACC in mirror
'she saw herself in the mirror.' Haspelmath (2019)

The examples above have a reflexive pronoun, either glossed as a pure reflexive or as 'self'. These work like other pronouns, in that they can be bound by another pronoun like in (6) or by a name like in (5) or (4). These, naturally have no issue with Condition-C.

Another approach is to change the verb itself to mark that something is reflexive. In the Hebrew example in (8) we can see that the verb 'made up' is "detransivitized" instead of receiving an object that is known to be reflexive. We can see this in (7) and (8), where we see that the addition of the middle voice morpheme ('MID') creates the reflexive reading.

(7) Jitsxak iper et tomi
Yitzhak made.up.INTS ACC Tommy
'Yitzhak applied makeup to Tommy.' Kastner (2017)

(8) Jitsxak hitaper
Yitzhak made.up.INTNS.MID
'Yitzhak put on make-up.' Kastner (2017)

As we begin to think about how we will countenance the Zapotec and Thai data in (2) and (3) we must consider what strategy of reflexive constructions the data looks most similar to. The first thing that we assume is that the bound R-expression takes the place where we would traditionally see a pronoun in languages that use a bound pronoun for reflexives, like in English in (9),

(9) Brook sees herself

From this surface similarity we will see that these bound R-expressions are a type of pronoun, not an R-expression. Notably, these pronouns have a complex phonological representation which we will discuss further in section 5.

1.1 Copy Reflexives

What we will look at next is data where this phenomena appears. We will begin by looking at it in Thai, where we can see that the copy is used as the subject of an embedded clause. Then we will move to Zapotec (San Lucas Quiavini Zapotec) where we will see this phenomena as another embedded subject, and one that has a sloppy reading from a clause eliding the copied nominal. Then we will see the copy being used as a direct object of a transitive verb in Hmong. And finally we will see it in Vietnamese where the copy is of an embedded clause, not the highest possible projection as we have seen before.

Thai:

(10) Nít$_i$ phûut wâa Nít$_i$sabaaj
Nit say COMP Nit comfortable
'Nit says that she's comfortable.' Chaipet and Jenks (2021)

Zapotec:

(11) Zi'cygàa' nih cay-uhny Gye'eihlly zèèiny b-ìi'lly-ga' Gye'eihlly zë'cy cahgza' Li'eb
while that PROG-do Mike work PERF-sing-also Mike likewise Felipe
'While Mike was working, he sang, and so did Felipe.' Lee (2003)

Hmong:

(12) Pov$_i$ yeej qhuas Pov$_i$
Pao always praise Pao
'Pao always praises himself.' Mortensen (2004)

Vietnamese:

(13) Luna$_1$ nói la Ginny$_2$ trách Ginny$_2$
Luna say that Ginny criticize Ginny
'Luna said that Ginny criticized Ginny.' Ivan and Bui (2019)

These data show us that the copying construction that we are looking at can take a wide variety of forms, and that they are able to be in a number of different constructions across the languages where copy constructions are licensed. What we however, must still look at is how it varies by language as it is naturally not identical across all of the languages that we have seen and will be focusing on.

1.2 Differences in what is allowed by language

Each of the languages that exhibits this feature does so in its own unique way, in the following we will examine how restrictions show up in each language. We will look at a few of the larger differences between the languages so that we can better understand the core of the phenomena that we are working with, instead of focusing on it in one language. We will begin by looking at bare nouns across the four main languages we will be considering: Thai, Zapotec, Hmong and Vietnamese. We will then look at what is copied, be it a whole DP or just the head of the phrase, and then we will consider the distribution of the strict and sloppy readings allowed by each of the languages in regards to scope, and finally we will look at quantification.

With this view of the overall variation between the languages we will be able to see the copy as it varies over four languages which will allow us to see what the copy is structurally, and how its representation will vary across languages without causing harm to the validity of the theory that we propose to handle this data.

1.2.1 Bare nouns

In Hmong we find that we cannot copy a bare noun, however in Zapotec we see that we can copy bare nouns. In the following examples, (14) from Zapotec and (15) from Hmong, let us see how the languages differ.

(14) B-yennlààa'z bxuuhahz ny-ahcnèe bxuuhahz Gye'eihlly
PERF-forget priest SUBJ-help priest Mike
'The priest forgot to help Mike.' Lee (2003)

(15) Dlev$_i$ yeej tum Dlev$_{*i/j}$
dog always bit dog.
'Dogs always bit dogs.'
*'Dogs always bit themselves' Mortensen (2004)

In (15), we see that there are two potential readings for this phrase, the first, which is grammatical is a reading where the two instances of *dog* are referring to different sets of entities. The second reading is that which has the copying construction, which is ungrammatical in Hmong with bare nouns.

Why could this be? How can it be that the same structure, a pronoun without its own phonological representation, can appear so differently? We must consider the nature of the nominal domain in these two languages. In Hmong, a classifier language, bare nouns denote a set of entities with x feature Bisang (1993). Classifiers are then used to individuate the nouns into denoting specific entities. In (16) we can see this in action.

(16) quas-dlev$_i$ pum quas-dlev$_{i/j}$
IND-dog see IND-dog
'(The/a) dog sees itself.'
'(The/a) dog sees a dog.' Mortensen (2004)

Here we can see that once there is a non-set-denoting nominal structure the copying construction is allowed, and both possible readings of this phrase are valid in Hmong. This shows us that the different types of nominal set-denotation are important for what we will expect the copying construction to copy. In Hmong where bare nouns are set denoting they are not allowed to be copied, but in Zapotec, where they represent an entity they can be copied.

Unfortunately, data is lacking for Thai that shows the copying construction in concert with bare nouns, but from data like (17) we can see that it is a language with similar noun structure as Hmong in that the *policeman* is set denoting and is referring to all entities with *policeman* properties, the listener then interprets this to be one of the many optional readings we can see in (17).

(17) Nít yàak phóp tamrùat
Nit want meet policeman
'Nit wants to meet a policeman/policemen.'
'Nit wants to meet the policeman/policemen.' Jenks (2011)

As such we would expect that the copying construction will only appear with classified nouns or entity denoting nominal expressions, like in Hmong. Vietnamese is similar as well, we expect it to have the same distribution as Thai and Hmong while Zapotec, with its different nominal structure will differ from the rest.

1.2.2 Full DP Copying

We will next be considering what we will do about copying the full nominal expression. In Zapotec and Hmong we are able to copy an entire nominal expression, which we will see in (18) and (19), but in Thai we are not able to do so, instead we are only able to copy the head of the nominal expression, like we can see in (20) and (21).

(18) R-yu'làà'z ra bxuuhahz ra bxuuhahz
HAB-like PL priest PL priest
'The priests like themselves.' Black (1994)

(19) [nam dlev]$_i$ pum [nam dlev]$_{i/j}$
great dog see great dog
'The Ol' Dog sees itself'.
'The Ol' Dog sees the (other) Ol' Dog.' Mortensen (2004)

(20) aajan Sid book waa aajan mai waang phrungii
teacher Sid tell COMP teacher not free tomorrow
'Teacher Sid said that he isn't free tomorrow' van Blankenstein (2021)

(21) *aajan Sid book waa Sid mai waang phrungii
teacher Sid tell COMP Sid not free tomorrow
'Teacher Sid said that he isn't free tomorrow' van Blankenstein (2021)

This data is more complicated than that of bare nouns, and is something that we will address further in section 5, but the base outline that we will see is that while the mechanism is the same, there is a difference in what the copying construction is allowed to copy in a C-commanding antecedent and in Thai, it can only take the highest projection (e.g. specifier) of that syntactic item.

1.2.3 Strict and Sloppy Reading

Across the languages we will be working with the exact readings of certain phrases with ellipsis are variable. In SLQZ there is only one possible meaning of a sentence like (22), that the subject of the second clause is performing an action on ω as in they are doing a reflexive action, this is a sloppy reading. In Thai, it is possible to have a sentence that can be interpreted with the sloppy reading and with the strict reading (see (23)). In Hmong this is operating similarly, as the sloppy reading is only is preferred and the strict reading is possible, although not preferred (see (24)).

(22) B-gwi'ih Gye'eihly lohoh Gye'eihly zë'cy cahgza' Li'eb
PERF-look Mike at Mike likewise Felipe
'Mike looked at himself, and Felipe did too' (i.e Felipe looked at himself/*Mike)
Black (1994)

(23) john koonnuat khong John lae Peter ko muankan
John shave of John and peter the same
'John shaved himself and Peter did too'
'John shaved himself and Peter shaved John' Black (1994)

(24) Pov$_i$ yeej qhuas Pov$_i$; Maiv$_j$ los kuj ua le hab
Pao always praise Pao May TOP also do as too
'Pao always praises himself and so does May (praise$_{i/j}$) Mortensen (2004)

In order to address why this is the case and what it means for the copy construction we will need to first more about the extensions of the proposed solution to the issue of copying constructions.

1.2.4 Quantification

In all of the languages we are working with quantification and conjunction are not allowed when used with the copying construction. We will see in (25) that it is not allowed in Zapotec. As well, in (27) we will see that conjunction phrases are not allowed either.

(25) *B-guhty cho'nn ra bxuuhahz ch'nn ra bxuuhahz.
PERF-kill three PL priest three PL priest
'Three priests kill themselves' Lee (2003)

In SLQZ, the strategy to handle the ungrammaticality of directly quantifying the copied phrase is to raise the subject to a topic position and use a pronoun with appropriate ϕ features, which then undergoes the copying construction, as we can see in (26). This is not surprising since Zapotec nouns are not set denoting, so a quantification of a referent item would not be allowed. However, this behavior carries into all of the languages we will see so there is a more complex feature of the nature of the copy construction which prevents this, we will see this in section 6.

(26) Cho'nn ra bxuuhahz b-guhty-rih la'arih.
three PL priest PERF-kill 3PL.DIST 3PL.DIST
'three priests kill themselves' Lee (2003)

(27) *Pov qhuas Pov thiab Maiv
Pov praise Pov and Maiv
'Pov Praises himself and Maiv' Boeckx et al. (2007)

In Hmong, we can see that the use of the copy in a conjunction phrase is not allowed. We will see that this has been a large question for other researchers (see 2.4), and one that we will address in more detail, as it requires that we first discuss the proposal to be made about this behavior, in section 6.2.

1.2.5 Table of differences

Below we can see a table of differences for the various languages we are working with, as well as English for comparison. This allows us to see that we are working with the copying construction not as a tool that will slot into each language identically, but instead as a construct that the languages have but are utilizing with different constraints. Importantly, we see clear focus points for what is not allowed (strict readings and quantification). From this we can build a stronger ω that aligns with the workings of each language as well as possible.

Restriction	SLQZ	Thai	Hmong	Vietnamese	English
Bare nouns allowed?	✓	×	×	?	×
Full DP copied?	✓	×	✓	✓	×
Strict reading allowed?	×	✓	?	?	×
Quantification allowed?	×	×	×	×	×

In order to handle these data I propose a new pronoun, called ω, which is phonologically underspecified and takes its representation from the highest C-commanding antecedent. ω is able to handle the variation across languages we see in (27) and the distribution of copy reflexives as embedded subjects and direct objects.

2 Earlier Analyses

Earlier analyses of the problem of how to handle these constructions focus on the idea of the R-expression as a variable. They have however been limited to observing a single language. In the following section we will begin by looking at how each language has been approached in the literature. At the end of this section we will look at other attempts to create a unified mechanism for this phenomena.

2.1 San Lucas Quiavini Zapotec

In Lee (2003), which analyzes San Lucas Quiavini Zapotec (SLQZ) the analysis relies on constructions like (28), below.

(28) R-yu'lààa'z-ëng la'anng.
HAB-like-3SG.PROX 3SG.PROX
'He/she likes himself/herself' Lee (2003)

When Lee analyzes the data here she saw that the second instance of the pronoun must be identical[1] to that of the first instance of it. This leads Lee to see that the second copy is a variable which takes its phonological realization from that which it is copying.

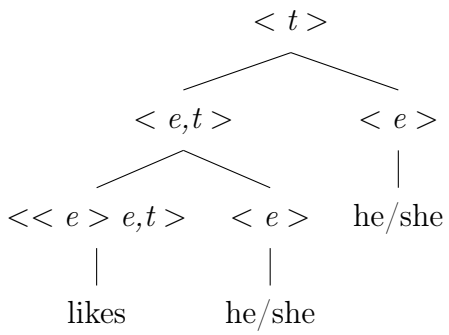
Lee then explains this mechanism with semantics, in the lambda denotation of the sentence in (28) we have:

[1]Phonology still applies, in this case 'la' is used to mark obligately bound morphemes when they are free.

(29) [[He/she likes himself/herself]] := [$\lambda x \in D.\lambda x \in D.x$ likes x] Where x is the pronoun.

In (29), which is the representation of (28), we can see that there is an X, as well as a second X, such that X *likes* X. This shows us that in the structure of a reflexive like this, two elements are created by the verb, from there the two objects which interact with the verb are the same *X*, which then shows us that in these we are not dissimilar from traditional transitive verbs, the key difference is in the fact that the two objects are the same referent, which is the core of reflexivity.[2]. In (30) we can see how Lee lays out a tree of the semantic types of these objects. This is important because it informs her argument about a type clash in sentences with conjunction.

(30)



Using the semantics of this structure Lee is then able to explain that reflexives that are quantified are not allowed due to a type clash. There is also the case of the strict and sloppy readings of constructions with two referents in them. Below we see an example:

(31) Zi'cygàa' nih cay-uhny Gye'eihlly zèèiny b-ìi'lly-ga' Gye'eihlly zë'cy cahgza' Li'eb.
while that PROG-do Mike work PERF-sing-also Mike likewise Felipe
'While Mike was working, he sang, and so did Felipe' Lee (2003)

In (31), Lee shows that the sloppy reading (Felipe is also singing while working), is an outcome of the variable analysis because if the phrase truly has two R-expressions we would expect that the sloppy reading would only apply to the first clause (i.e., Felipe did work while Mike sang). As this is not the case, Lee is able to use this data in support of the variable reading. Because a second full realization of *Mike* would cause two issues, the first being that we have two different subjects within embedded clauses, and that *Felipe* would either be referencing *singing* or *working*, not both like the sloppy reading we see in SLQZ.

2.2 Hmong

In Mortensen (2004) a very similar proposal to Lee's variable analysis is made. There is however a key difference; Mortensen creates a two tier structure of reflexive variables, *ana*, which a pronoun taking an invariant reflexive meaning as well as neither having person

[2]This will be useful when we consider whether non overt reflexives are ω

nor number ϕ features. However, Mortensen notices some problems with this approach in sentences like (32) and (33).

(32) *[tug twg]$_i$ los Pov$_i$ tsi nyam ana_i^{Pov}
CLF which TOP Pao NEG like *[ana]*
intended: 'No matter who he is Pao would dislike himself' Mortensen (2004)

(33) [tug twg]$_i$ los Pov$_{*i/j}$ tsi nyam nwg
CLF which TOP Pao NEG like 3SG
'Pao would dislike anyone' Mortensen (2004)

Mortensen[3] sees that the 'which', the CP which is attempting to take on the role of the subject of the sentence restricts the use of the *ana*. This is not a constraint which has been noted in SLQZ so we will have to find a way to create a framework that allows for both of these languages. Mortensen then explains the issue in (32) by arguing that *ana* is not licensed when there is another pronoun that is able to take its place. We will examine the work of the CP which is commanding the subject of the sentence as a potential barrier in the surfacing of the copy, which is not a path Mortensen took.

As well in his analysis Mortensen noted that there is a restriction in Hmong of the variable taking a bare pronoun. This is not a constraint in SLQZ and we will look at how the set denotation of NPs and DPs shapes the ω and where it can be used.

2.3 Thai

Of all of the languages that exhibit these copy reflexives Thai has had the most written about it. Unfortunately much of these works have contradicted each other both on matters of theory but also on matters of grammaticality. From Lasnik (1989), Lee used the example of (34) to show that copying constructions are active in Thai, but in Narhara (1993) it is contended that (34) is not grammatical to any of her language sources.

(34) John konnuad John
John shaved John
'John shaved himself' Lee (2003)

However, a sentence like (35) is grammatical in all of literature available. In order to countenance this we will spend a section on Thai after we have established ω more clearly.

(35) Aajarn kid waa puak rua chob aajarn
teacher think that all we like teacher
'The teacher$_i$ thinks that we like him$_i$' Lee (2003)

[3]All examples are directly as they appear in the original. In (33), the 'Pov' above 'ana' means that 'Pov' is the phonological form of 'ana'.

Lee, whose work has primarily focused on SLQZ, also addressed Thai, and claimed that (34) is grammatical, she then uses this to build up her theory of a null variable that is taking the place of the pronoun in reflexive constructions. Others like Larson (2006) and Chaipet and Jenks (2021) see the phenomena as a case of names acting as complex indices. This idea of complex indices is applied by Chaipet and Jenks (2021) to create a system of complex indices, as we will see in §2.2 These approaches see the apparent discordance of the Thai reflexives as evidence that Condition C is active, and that there are complex semantic workings within the use of a proper name in Thai reflexives that does not apply in other cases, like (36) where we can see that *pirate* does not have the same properties as *teacher* in Thai.

(36) coonsalàt$_1$ phûut wâa coonsalàt$_{*1/2}$ sabaaj
pirate$_1$ said COMP pirate$_{*1/2}$ comfortable
Pirates said that pirates are comfortable
∗'The pirate said that he/she's comfortable.' Chaipet and Jenks (2021)

This is not a perfect solution, but it does raise an issue that we will address later. Namely, that the way a language defines its nouns as sets or entities will determine how they are worked on by the copying construction. We will see how this is not as as simple as either Lee (2003) or Chaipet and Jenks (2021) treat it, and that there is in fact a different mechanism at play in Thai, namely one that focuses on the denotation of nouns, where *pirate* denotes a set, and *teacher* denotes an entity.

2.4 Copy Theory of Movement

The Minimalist Program, and its corollary, the Copy Theory of movement also provide a unified account of the ω phenomena in Boeckx et al. (2007). The authors lay out the idea that the behavior in languages like San Lucas Quiaviní Zapotec and Hmong are evidence of a movement analysis of control and reflexivization. They also argue that the phonetic realization of copies generated by the movement, which we term as ω, are regulated by linearization and morphological requirements.

The key thing that this approach does is present a unified account for the ω phenomena, it does not attempt to come up with a language specific system like the other approaches we have seen thus far. Within that account it is critical to reduce the antecedent-anaphor and control relations as instances of movement understood as the composite of Copy and Merge. This means that they are using a version of the semantic mapping we have seen before to say that we know there is an argument in all reflexive constructions. What they do with this is to break down simple reflexive and co-referential phrases into having an overt copy of the original subject in what we can think of as the deep structure. In (37) we can see what this looks like for basic English phrases, with the *a* sentence being the PF and the *b* sentence being the LF.

(37) a. John saw himself
b. John saw John

(38) a. John wants to eat
b. John wants John to eat

This then allows Boeckx et al. (2007) to build a framework within the Minimalist Program that explains how this idea in (38) and (37) can be applied to languages like SLQZ and Hmong. As they layout the workings of the movement taking place to account for this behavior they lay out a proposed structure for the syntax of these phrases.

(39) a. John likes himself
b. [$_{TP}$ John [$_{T'}$ T [$_{vP}$ John [$_{VP}$ likes John]]]]

Boeckx et al. (2007) see the reflexive behavior we have been looking at as a movement to the θ-position. Where in (38) we are seeing *John* merges with *likes* and thereby *John* obtains the 'likee' role. They are also able to explain the inability for the *John* to move outside of the local binding domain by locating this type of reflexive within the A-movement domain. In a sentence like (40), *John* cannot cross *Mary* as that would violate the shortest move condition.

(40) [$_{TP}$ John [$_{T'}$ T[vP John said[CP that[TP Mary likes John]]]]]

This condition also helps to explain why we do not see strict readings of these ω constructions, as the shortest move condition allows for the sloppy readings that we see in SLQZ and Hmong in (41) and (42).

(41) R-ralloh Gye'eihlly r-yu'lààa'z Lia Paamm Gye'eihlly
HAB-think Mike HAB-like FEM Pam Mike
'Mike$_i$ thinks that Pam likes him$_i$' Lee (2003)

(42) Pov xaav has tas Maiv Nyam Pov
Pao thinks say that May like Pao
'Pao$_i$ thinks that May likes him$_i$ Mortensen (2004)

Boeckx et al. (2007) also expect that this will not allow for copy-reflexives in coordinate structures ("and") to be grammtical, something which they claim, the pronoun-esque approaches in Mortensen and Lee should incorrectly predict. We will see in §6 how ω handles this data, seen in (43) and (44) from Hmong.

(43) *Pov qhuas Maiv thiab Pov.
Pov praise Maiv and Pov
'Pov praises Maiv and himself' Mortensen (2004)

(44) *Pov nyiam Pov thiab Maiv.
Pov likes Pov and Maiv
'Pov likes Maiv and himself' Mortensen (2004)

Boeckx et al. (2007) see this data as unexpected for Mortensen and Lee because neither of these accounts would expect the pronominal type system they are using to have a clash from a simple conjunction. They then compare this to English data like (45) which they see as contradicting this. We will see in 6, as well, that this data is not universally accepted by native speakers of English.

(45) a. ??Olivia praised herself and Amanda.

b. ??Olivia praised Amanda and herself.

This data is then used to advance their idea that the identical antecedent requirement from Lee is not a workable solution. The first part of this point is that without the ability to work in conjoined phrases, we would expect a very strict interpretation of identical antecedent and the existence of the conjunction-blocking of the reflexive copy would not allow for the ability of pronouns to co-reference with a higher antecedent like we can see in (46).

(46) R-yu'lààa'z me's nih r-umbèe' Lia Paamm la'ang
HAB-like teacher REL HAB-know FEM Pam 3SG.PROX
'The teacher who knows Pam$_i$ likes her$_i$.' Lee (2003)

It is unfortunate, but we do not know if there is an allowed j reading on the pronoun in (46) which would provide conclusive data to this point. However, since this is grammatical, there is a clear problem with Lee's approach to the reflexive copying as the system in Lee (2003) is seen by the theory in Boeckx et al. (2007) to predict that (46) would be ungrammatical, as this would be too long of a distance for a copy construction as Lee (2003) laid it out. The ω approach does not have this problem, which we will discuss in 6. Next Boeckx et al. (2007) look at case, which could present a challenge for their analysis, as the copy reflexive is case-less. They handle this by adding a -self, item to the movement of the noun. We can see what this looks like in (47).

(47) [$_{\text{TP}}$ John[T' T[vP John[VP likes John-self]]]]

The insertion of the *-self* is then able to block the Case requirements which would otherwise invalidate the path of minimalism that Boeckx et al. (2007) have been taking. This is a weak point in their analysis as the addition of the *-self* does not create an anaphoric construction by default, it simply removes the need for case assignment. This is supported by the claim that *-self* is not anaphoric because it is merged with *John* and *John* is then free to move to Spec of vP and check its case against the T', thus meaning that *John* carries the case, not the *-self*. From here, they lay out an idea where the *John-self* item is pronounced as him in english as it goes through Spell Out. This data seems to not align particularly well with the Zapotec data in (46), as we would not expect for there to be two competing paths for this to realized in SLQZ, it would instead be left as either a pronoun like in (46) or an overt copy like (41). To reinforce this point the authors gathers Chinese data which is part of a closed class of nouns, that allows for a *x-self* construction to be valid in Chinese. We can see this in (48).

(48) mama hen xihuan mama-ziji
mother very likes mother-self
'Mom likes herself' Boeckx et al. (2007)

This is the first categorical issue with the work, as it cannot adequately explain optionality, and the very limited use of *x-self* constructions in Chinese, as can be seen in Boeckx et al. (2007), which notes that some Chinese speakers only accept the *x-self* with a few familial terms.

Then Boeckx et al. (2007) move onto control structures. The first key to their analysis is that a second type of use for the copy reflexive is identified; the use of the reflexive copy in place of a PRO. We can see this in (49).

(49) Zi'cygàa' nih cay-uhny Gye'eihlly zèèiny b-ìi'lly-ga' Gye'eihlly zë'cy cahgza'
while that PROG-do Mike work PERF-sing-also Mike likewise
Li'eb.
Felipe
'While Mike was working, he sang, and so did Felipe' (Lee, 2003)

As well we can see the structure of (49) in (50).

(50) Mike sang while [Mike working], and so did Felipe.

Boeckx et al. (2007) see this as vindication for their theory of A-movement, where movement happens to a θ-position, this PRO is then actually a trace (copy) left over by A-movement as *Mike* moves up from the bottom of the structure to its resting place at the top of the structure. Boeckx et al. (2007) then move onto morphemes which block the copy reflexives in SLQZ and Hmong. We will address these in 5 and 6. Boeckx et al. (2007) see the quantificational blocking of the copy reflexives as a feature of the movement inherent to them under the Minimalist Program, while Lee and Mortensen see them as a unique outcome of quantifying these "quasi-pronouns" that their proposals advocate for. We will see these as a result of a feature that removes the voicing of ω. This data is then used to flesh out a theory of minimalism relating to the nature of linear control chains, which while an important aspect of their work is no longer focused on understanding the copy reflexives that we are focusing on.

Overall, the work in Boeckx et al. (2007) seeks to explain the use of the copy reflexives within the Minimalist Program, and they are able to explain many aspects of its use within the Minimalist Program. However, their analysis does not cover everything that we will be covering; it does not mention cases like Thai, where only certain parts of a copy may be articulated. As well there are issues in their *-self* analysis of reflexivity as it does not explain why multiple options exist in SLQZ for certain types of reflexives. While this article is important to the Minimalist Program, it does not adequately explain the phenomena that we are looking at.

3 Pronouns

In this section we will look at theories of pronouns as presented by Déchaine and Wiltschko (2002), which will form the basis of the ω analysis. From the three pronoun types; Pro-DP, Pro-NP, and Pro-ϕP which are presented in that article we will build towards adding a fourth, ω, and exploring why it cannot be a Pro-ϕP which it closely resembles. We will also look at the internal structure of the ω and explore why we see head only copying in Thai and not in other languages.

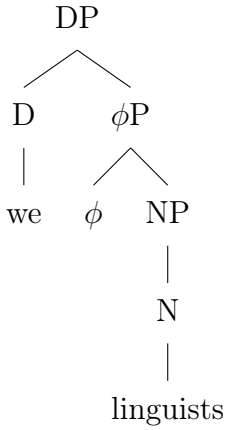
3.1 Earlier Approaches

As we have seen above, pronouns are not simply lexical items, they are complex structures. And we must understand their internal structures before we can approach variables in reflexives. We will build upon the theory of the tripartite pronoun of Déchaine and Wiltschko (2002), which states that there are three types of pronouns: Pro-DP, Pro-NP, and Pro-ϕP.

3.1.1 Pro-DP

To start with Pro-DP, we will expect them to take on the syntactic and morphological complexity as they have a sub-NP, which allows for the addition of additional layers to the syntax of these pronouns. As well, Pro-DPs should take the binding structure of R-expressions, because of this we expect Pro-DPs to be restricted to the argument position and have an embedded ϕP where they receive their ϕ features.

(51)



In 3.1.1 we can see the example sentence 'we linguists' has an intervening ϕP between the DP and the NP. This allows for 'we' to assign the needed ϕ features onto 'linguists'. We can also see this in other languages, like Halkomelem[4], where the Pro-Dp structure can explain and predict an interesting pronoun system.

(52) Tl'ó-cha-l-su qwemcíwe-t thú-tl'ò q'ami
then-FUT-1SG-so hug-TRANS DET.FEM-3SG girl

[4]Halkomelem is a Central Coast Salish language.

'then I'm going to hug that girl' Galloway (1993)

Within the DP in (52), 'thú-tl'ò q'ami' the D - 'thú' is able to take the feminine 3rd singular in its ϕ head position, while the NP, 'girl' fills out the structure. From this, we should expect that the Pro-DP will take the place of a full DP phrase, but that importantly, the rest of the pronoun is a ϕP.

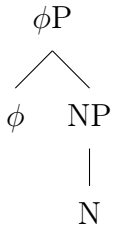
3.1.2 Pro-ϕ

Next we will look at the pro-ϕP. In (53) we can see that the pro-ϕP is the same as the interior of the Pro-DP structure. As well in (53) we can see what this looks like in practice from Halkomelem, in which the phonologically null 3.SG.SUBJ is able to be referenced as the subject of *leave* because it has matching ϕ features, which happen to be represented as a null in this language.

(53) Tsut-$\emptyset_i$ m qwetséts-$\emptyset_i$ newî7-s
say-3SG.SUBJ PAST leave-3SG.SUBJ EMPH-3
'He$_i$ said that **he**$_i$ left.' Lai (1998)

From (53) we can see that there are two null subjects

(54)



The ϕP approach allows for pronouns to be reduced to a bundle of features, and is used by Chaipet and Jenks (2021) as a foundational argument for their case of names as complex indices in Thai. In this case 3SG SUBJECT. This is the most important feature of the pro-ϕP because it is what separates it from something more filled out than the pro-DP, and can include an NP under it. In (55) we can see data from QZ that challenge this paradigm, and in (56) we can see data from Vietnamese that shows us it is still needed.

(55) R-yu'laaa'z-eng la'anng chiru' ze'cy cahgza' Gye'eihlly.
PERF-look-3SG.PROX 3SG.PROX also likewise Mike
'He/she likes himself/herself, and Mike does to' (i.e. Mike likes himself) Lee (2003)

(56) Snape$_1$ nói là Lockhart$_2$ bâu cho nó$_{1/2}$
Snape say that Lockhart vote for 3SG
'Snape said that Lockhart voted for him/himself.' Ivan and Bui (2019)

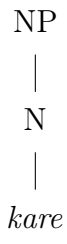
In (55) we can see that it is possible for an elided clause to take its reading from a pro-ϕ, but without referring to the entity inferred by the ϕ-features. This then shows us that this cannot be a pro-ϕ in SLQZ. As well in (56) we can see that in Vietnamese it is possible for a single pronoun to hold multiple meanings, this is unexpected by other pronoun systems than Pro-ϕ, which is able to explain this by simply seeing that the 3SG pronoun is simply matching ϕ features in both R-expressions.

3.1.3 Pro-NP

Next is the Pro-NP, which is replacing a simple NP expression. The Pro-NPs are expected to have the syntax of nouns, as such it is not surprising that they can also be modified by adjectives and other elements under an NP. In (57), we can see the example of *kare* from Japanese which is presented as the prototypical pro-NP in Déchaine and Wiltschko (2002).

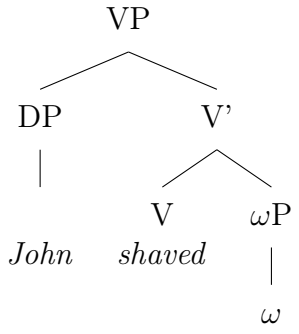
(57) tiisai kare
small he
'he who is small' Déchaine and Wiltschko (2002)

This then leads us to the structure of the Pro-NP. It is remarkably simple compared to the other structures we have seen thus-far, but it is important for the workings of numerous parts of the ω.

(58)


4 New Approach, ω

Building on Lee (2003) and Déchaine and Wiltschko (2002) I propose that there is a phonologically null pronoun in the syntax, ω, which holds no independent phonological realization and takes its realization from its C-Commanding antecedent. In the Thai from (3), we can see this in the tree below.

(59)


ω is key to understanding how reflexives work in the languages with copy reflexives. As we can see in (60) all reflexives have an object, but it is not always pronounced.

(60) [[He/she likes himself/herself]] := [$\lambda x \in D.\lambda x \in D.x$ likes x]

In some languages, we will see a Pro-DP being copied, but in others we might see a verb receiving a reflexive marker one of the numerous other approaches that languages take for reflexive constructions. We will pay particular attention to languages like Hebrew who "de-transivitize" verbs to make them reflexive, as we will see there is actually an unvoiced ω in their structure.

4.0.1 Binding is still active

One item that needs to be addressed for ω to be valid is that of Binding Theory, and its conditions from (1). ω relies on the assumption that all three conditions are still active in the languages that it is present in, particularly condition C, which is the primary condition that copy constructions seem to violate. In this following section I will briefly provide data to show that all of the conditions are still operating in the languages with ω.

To begin with we will see Hmong, we will see in (61), that Condition C is still active, as R-expressions cannot be bound by a pronoun.

(61) Nwg$_i$ yeej qhuas Txiv$_{*i/j}$
3SG always praise father
'He always praises Father'
*'Father always praises himself" Mortensen (2004)

This shows us evidence that Condition C is active in Hmong, and we will need an explanation for the copy constructions that works within the paradigm of the conditions of Binding. Next, we will see from SLQZ that it, as well cannot violate condition-C by binding R-expressions with pronouns, as we will see in (62),

(62) B-gwi'ih-ëng lohoh Gye'eihlly
PERF-look-3SG.PROX at Mike
'He looked at Mike$_{j/*i}$' Lee (2003)

That being said, neither SLQZ or Hmong can have an R-expression bound by another, even if they are meant to mean the same person, as we can see in Hmong in (64). This is particularly interesting as men in Hmong culture receive two names, and they cannot be used in concert in copy constructions, that is, we can see that ω does not refer to the individual, rather to the grammatical element above it. In (64), let us assume that;

(63) 'Kou's old-name is Chu-Ndzai.´ Mortensen (2004)

From (63) we can see how ω is certainly a grammaticized element in Hmong.

(64) Kub$_i$ yeej thua Tshuv-Ntxaij$_{*i/j}$
Kou always criticize Chu-Ndzai
'Kou always criticizes Chu-Ndzai
$*$'Kou always criticizes himself.' Mortensen (2004)

And next, to look at Thai in (65) we will see that the conditions are active.

(65) khon nán$_1$ phûut wâa Nít$_{*1/2}$ sabaaj
person that say COMP Nit comfortable
'That person$_1$ said that Nit$_{*1/2}$'s comfortable.' Ivan and Bui (2019)

Finally, Vietnamese is a problem, as it is argued that it does not obey the conditions of binding theory. It appears that it may violate all three of the conditions of binding theory, but as Ivan and Bui (2019) lays out, they are still at play, just with a more complex set of competition features for the interactions between long and short distance interpretations for pronouns in Vietnamese.

4.1 What is ω?

ω is a nearly empty set of features, it does not carry any ϕ features, it is instead only co-referential with a C-commanding antecedent. Because of this flexibility it can appear in a few different positions such as reflexive pronoun, or as a subject, because of its ability to reach Spell Out without its own lexical form. The key feature of ω is that it is obligatorily co-referent with a C-commanding antecedent, which is how it is able to take its reading as a reflexive, we will discuss this more in 4.3. ω is, as a bundle of features, able to take on quite a diverse set of uses, depending on the language. For example, it can go beyond a mere reflexive. In QZ, there must be an overt subject, in the embedded clause in (66) there is a repeat of the name Mike, which should be a violation of Condition-C, however, because ω is not actually the same as the original instance of 'Mike', instead it is merely a pronoun with a feature that makes it realize itself as a phonetic copy of 'Mike' in Spell out.

(66) Zi'cygàa' nih cay-uhny Gye'eihlly zèèiny b-ìi'lly-ga' Gye'eihlly zë'cy cahgza'
while that PROG-do Mike work PERF-sing-also Mike likewise
Li'eb.
Felipe
'While Mike was working, he sang, and so did Felipe' (Lee, 2003)

(67) val thii tua Val ʔ-eeng
Val hit RF Val RF
'Val (the speaker) hit Val-self' Narahara (1995)

In (67), we can see a case where there are are items below the ω. This then raises the question of what is happening under the ω. Since the ω is simply a pronoun it is then able

to take on the post-nominal-items, like the reflexive marker in Vietnamese.

4.1.1 ω in Direct Reflexives

We will begin by looking at ω in direct reflexives, where it is the direct object of a C-commanding antecedent. In (68) and (69) we can see examples of this in (68):

(68) R-yu'lààa'z ra bxuuhahz ra bxuuhahz
HAB-like PL priest ω
'The priests like themselves' Black (1994)

(69) John konnuad John.
John shaved ω
'John shaved himself' Lasnik and Stowell (1991)

In the prior two examples we can see that ω is being used directly as the object of a C-commanding Antecedent R-expression. As well, we can see that as we discussed in 1.2.1, the use of *priests* in SLQZ is equivalent to the R-expression in (69), from Thai because of the different set denotations of nominal expressions in the two languages. As for direct ω objects, they are not as common as the indirect use of ω, they are debated in Thai (Narahara (1995)) and seem to be facial violations of Condition C. This is not the case, they are still ω in the languages that they are present in, as we can see in (70) and (71).

(70) Pov_i yeej qhuas Pov_i Maiv_j los kuj ua le hab.
Pao always praise ω May TOP also do as too
'Pao always praises himself, and so does May.' (i.e. May praises herself) Mortensen (2004)

(71) B-gwi'ih Gye'eihlly lohoh Gye'eihlly ze'cy cahgza' Li'eb.
PERF-look Mike at ω likewise Felipe
'Mike looked at himself, and Felipe did too'
∗'Mike looked at himself, and Felipe looked at Mike' Lee (2003)

In (71) and (70) we can see some instances of how the use of ω as the direct object of a phrase is referring to the co-reference property of ω. As well, looking at this data we would expect that ω is reflexive. This is not the case, it is merely denoting that it is co-referent with a C-commanding antecedent. This is discussed in depth in §4.3.

4.1.2 ω in Indirect Phrases

When ω is in use in Indirect phrases, in which it takes on the role as a bound object in the subject position of a C-Commanded clause. We can see this in (72) from Thai and (73) from Zapotec.

(72) Nít$_i$ phûut wâa Nít$_i$ sabaaj
Nit say COMP Nit comfortable
'Nit said that she's comfortable' Chaipet and Jenks (2021)

(73) W-eey Benit mëlbyuuu ne y-ged Benit lo x-mig Benit Jasint
COMP-take Benito fish that POT-give Benito face POSS-friend Benito Jacinto
'Benito took a fish, which he gave to his friend Jacinto' Lee (2003)

When it is in this position ω is in the subject position of a subjacent clause to the main clause of the phrase, this is interesting to us as it is not the traditional role of the the reflexive, this provides us more evidence that it is not carrying a reflexive feature.

The first clue that this is an ω is that the Thai restriction on the use of the heads of phrases as the bound copy in the second subject instance as we can see in (74) and (75).

(74) Aajan Sid book waa aajan mai waang phrungii
Teacher Sid tell COMP teacher not free tomorrow
'Teacher Sid said that he isn't free tomorrow,' van Blankenstein (2021)

(75) *Aajan Sid book waa Sid mai waang phrungii
Teacher Sid tell COMP Sid not free tomorrow
'Teacher Sid said that he isn't free tomorrow.' van Blankenstein (2021)

This shows us that ω still has constraints upon it in these constructions, when it is the head of its own subjacent clause, this shows us that the C-command upon ω is happening in Thai, as we can see that there are constraints on the ω as it goes through spell out in (75) and (74), we will see in 5.2. As well we can look at binding theory for our next clue as to the use of ω in these constructions. According to condition-C an R-expression is free, and as we can see in the English gloss of (75), it must use a pronoun with matching ϕ features as the subject, as saying something like (76) is ungrammatical.

(76) Lucia$_i$ thinks Lucia$_{*i/j}$ is smart.

This, in English is a violation of condition C because the R-expression 'Lucia' is truly an R-expression. In the Zapotec and Thai we can see that in order for this to be grammatical we need the R-expression in the embedded clause to either not be an R-expression, which ω fulfills clearly, or the languages must not abide by the conditions of binding theory, which we do not have evidence for beyond the cases that we can see now are ω.

4.2 How is ω different from ϕ

Let us return to Jambi Malay from (111), where we can see a most interesting difference in pronouns. There is the option for ω and for a pro-ϕ.

(77) Dio$_1$ cinto dio$_{1/2}$
he love he
'He loves him' or: 'he loves himself.' Cole et al. (2015)

In (77), we can see that there is a distinction between the ϕ reading of the pronoun and the ω reading. Because the key difference between the two types of pronouns is that there is a bundle of ϕ features that comes with a pronoun, however, when the ω represents it does not have ϕ features, it is a copy, or a pure reflexive. So in (77), the first reading is that of an ω while the second, where the ϕ features are carried over is a pro-ϕ. This then, helps us to distinguish that, there is in fact an ω which is operating in the structure of Jambi Malay, which allows for these two possible readings.

How then do we find a difference between the two types of feature bundles? The most important piece is that the ω does not carry ϕ features as we saw above in the Jambi Malay. However, there are a number of languages that use so-called reflexive particles to achieve reflexivity. We are then faced with a question, do these meet our definition for an ω? Let us look at English and Polish to see differences in the distribution of pronouns.

(78) widziała siebie w lustrze
she.saw self.ACC in mirror
'she saw herself in the mirror.' Haspelmath (2019)

(79) Coco saw herself in the mirror.

In the Polish in (78), we can see that there is no gender or number marking on the reflexive *siebie*, the most common ϕ features like we can see in English in (79). In Polish, instead the reflexive marker *self* is marked with case, which is of paramount importance when it comes to Slavic languages, and is an important feature that we see marked in a large percentage of the worlds languages. By comparison, the languages that use ω do not mark it with case. It can however be marked as being present. In the Vietnamese in (80) we can see that ω is marked to show that it does not have ϕ features.

(80) val thii tua Val ʔ-eeng
Val hit RF Val RF
'Val (the speaker) hit Val-self' Narahara (1995)

We can see that the two RFs in this phrase are being used to mark that it is an ω, and not a pronoun with ϕ features. The question that comes next is why Vietnamese has such a complex system to mark ω, this is because Vietnamese has a very free binding system, which we can see in (81) and (82).

(81) Ginny$_1$ nói với Luna$_2$ về mình$_{1/*2/speaker}$
Ginny talk with Luna about SELF
'Ginny talked with Luna about herself/me' Ivan and Bui (2019)

(82) Ginny$_1$ nói vói Luna$_2$ vê nó$_{1/2/*speaker}$
Ginny talk with Luna about 3SG
'Ginny talked with Luna about her/herself' Ivan and Bui (2019)

This data shows us that Vietnamese pronouns and reflexives have a very wide range of possible interpretations and as such it is very important that ω be marked, lest it be confused for another option that Vietnamese has for binding constructions. The use of the marker becomes even more apparent if we consider a case where ω is taking the representation of a pronoun and may need to be marked to show that it must be reflexive.

4.3 Is ω reflexive?

ω is not reflexive. It is merely co-referential with its highest C-commanding antecedent. While it may seem natural for ω to be overtly reflexive with a [+ reflexive] feature, it is counterproductive to the point of the pronoun and leaves open much room for issue without clarifying the nature of ω in a useful way. The first thing that we would consider if ω were to be overtly reflexive is Case. For it to have the [+reflexive] feature it would also be invariantly accusative as it then must be the object of an overt subject. This is of course not the case as we can see in (83), from SLQZ.

(83) Zi'cygàa' nih cay-uhny Gye'eihlly zèèiny b-ìi'lly-ga' Gye'eihlly zë'cy cahgza'
while that PROG-do Mike work PERF-sing-also Mike likewise
Li'eb.
Felipe
'While Mike was working, he sang, and so did Felipe' (Lee, 2003)

This has the second instance of *Mike*, the ω representation in the subject position of a clause without an accusative argument. For this reason ω must not be overtly reflexive. The next question that this raises, is without it being an overt reflexive, how do we handle issues like (84) where a sloppy reading proliferates, would it not be the case that ω means that *Felipe* would be looking at *Mike* since ω is referencing *Mike*?

(84) B-gwi'ih Gye'eihly lohoh Gye'eihly zë'cy cahgza' Li'eb
PERF-look Mike at Mike likewise Felipe
'Mike looked at himself, and Felipe did too' (i.e Felipe looked at himself/*Mike)
Black (1994)

No. it is not the case because when we look at what is elided as the object of *Felipe* we see that it is *looked at* ω. Because of this, and the *likewise*, limiting the elided ω to the binding domain with *Felipe* at its top, it is then an obvious conclusion that ω does not need to be reflective to create the sloppy reading. This also leaves us open for an explanation of the preference, but not requirement, that phrases like (84), in languages like Hmong to take the sloppy reading over the strict reading. The explanation being that the interceding

element that would limit the ω to the binding domain to that with the head of the embedded clause at its top is not forcing the restriction, rather merely implying it. This can be seen in (85).

(85) Pov$_i$ yeej qhuas Pov$_i$; Maiv$_j$ los kuj ua le hab
Pao always praise Pao May TOP also do as too
'Pao always praises himself and so does May (praise i/j) Mortensen (2004)

From this we can see that ω is not overtly reflexive, or assigned a specific Case, it is merely a pronoun without phonological representation that takes its PF from a antecedent C commanding Nominal Expression.

5 Phonology

In this section we will be covering the phonology of ω and how it interacts with that of the languages that exhibit the ω in their grammar. We will start by detailing my proposed mechanism for the process of articulation, and where ω fits into that schema, and then we will look at a general course of how a language could choose to use ω, then finally we apply that information to Thai, SLQZ, Hmong, and Vietnamese.

5.1 From Semantics to Speech

The first thing that I want to cover is how we go from the mind, to the speech act. There is not yet a fully settled approach for how ω will work in the phonology of the languages it works in. We will therefore lay out two different ideas of why ω is reduplicated. The first is that when it is inserted into a linear model of the semantics to phonetics pathway there is no provided phonological data in the ω "package" in the syntax, from there it takes on the phonological representation of its C commanding antecedent. The other proposal we will consider is one of reduplication. Since the ω lacks a phonological realization and is yet spoken as a copy of another item in the syntax, we can instead look at this as a case of reduplication. I will present the arguments and the extensions for how each will handle the data that we are working with below.

5.2 Linear Model

In the linear model we be working with assumption that the order of operations for production goes something like this, in (86) below.

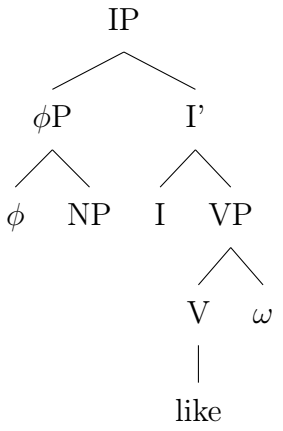
(86) [Semantics $\rightarrow$ Syntax $\rightarrow$ Phonology $\rightarrow$ Phonetics]

So as we can see, we have an underlying semantic representation. Say for example the sentence 'he/she likes himself/herself'

(87) [[He/she likes himself/herself]] := [$\lambda x \in D.\lambda x \in D.x$ likes x]

From here we apply this to the syntax, in English we get something like this,

(88)



In (88) we can see that, while there is still structure to the syntax, it is lacking a representation in the phonology, so when we move to the phonology the grammar looks up from ω for the highest C-commanding object, traditionally the object of the main clause of the binding domain of the ω. What we are yet to find is the limits of the binding domain of the ω. We can see this in the Zapotec in (89) that the ω is in fact taking its representation from the highest C commanding antecedent in its binding domain. But, what we do not have data for is a case where a non subject item in the syntax is raised above the verb to a focus or topic position, this would be something that would help to clarify the nature of the phonology of the ω. However, as for the affirmative argument for this system, which also helps to clarify the nature of the phonology we can see that in (89).

(89) R-yu'lààa'z-ëng la'anng.
HAB-like-3SG.PROX 3SG.PROX
'He/she likes himself/herself' Lee (2003)

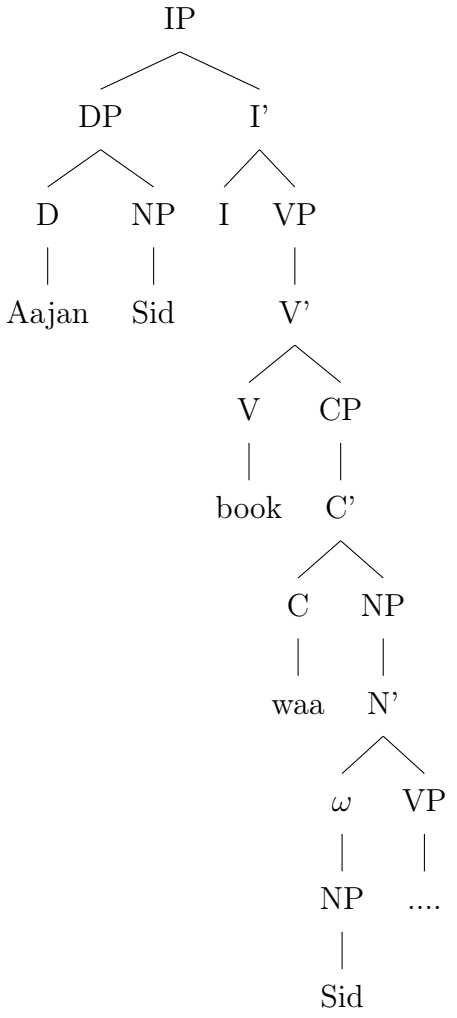
Here, the *ëng* and *la'anng* are the same. In Zapotec there is quite a robust system of phonology, which changes the form of pronouns, sometimes massively. This lets us locate that the copying is happening at the level of the syntax, since the phonological changes that are then applied in the phonology of Zapotec. This mechanism would then suggest a very simple copying operation, and this is true, except for Thai. In that language only the head of a phrase can be copied by the ω.

(90) Aajan Sid book waa aajan mai waang phrungii
teacher Sid tell COMP teacher not free tomorrow
'Teacher Sid said that he isn't free tomorrow' van Blankenstein (2021)

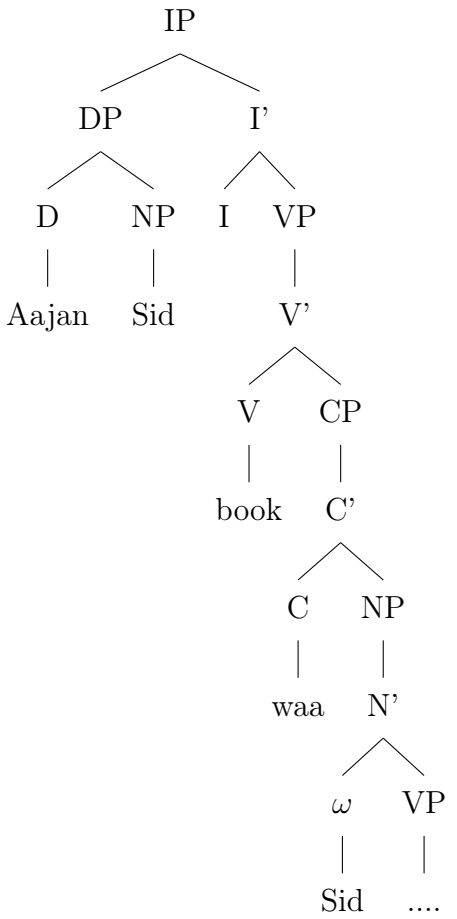
(91) *Aajan Sid book waa Sid mai waang phrungii
teacher Sid tell COMP Sid not free tomorrow
'Teacher Sid said that he isn't free tomorrow' van Blankenstein (2021)

Here we can see that in Thai, we can only use the ω on the head of the phrase. Below in (92) and (93), we can see the structure of these two sentences.

(92)



(93)

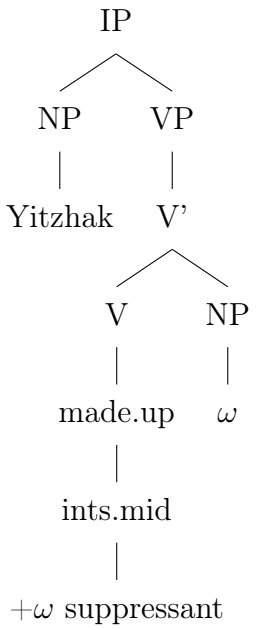


Here we can see that there is an interesting move at play, the ω in Thai, is not only taking the highest C-commanding antecedent as its representation for ω, but it is also restricted to taking the highest projection of that projection. In this case, that is the DP specifier, *aajan*. By this mechanism we could also expect to see a mechanism whereby ω can be repressed, and not spoken at all. This is a way whereby we can explain the interesting case of Hebrew. In (94) we can see a verb which is "detransivitzied" to mark that it is reflexive.

(94) Jitsxak hitaper.
Yitzhak made.up.INTNS.MID
'Yitzhak put on make-up' (*'Yitzhak got make up applied to him), Kastner (2017)

When we look at the structure of this phrase in (95) we can see that it is actually holding an ω, as the semantic representation of this type of phrase in (87).

(95)



Through this we can see that in Hebrew there is an active ω, it is merely suppressed by the verb, which is already receiving a voiced marking that there is an object, this does not however, represent the object itself in the language, as it is part of the verb in a way that other objects in Hebrew are not allowed to be.

In this last section we have seen a potential "linear" model for addressing the nature of the phonology of ω. This theory is not without flaws, the largest being its insistence on a truly linear structure of grammar that requires syntax being fully formed before phonology, this would fail to handle languages like Basque, where phonological Focus impacts the syntax, like in (96) and (97) where the phonological Focus must be in the same position, so the objects of the sentence are reordered in the syntax to handle this. The linear model would not predict this. So while it is a useful tool to consider how ω works we must look at more widely accepted theories, such as the Y model which will allow to base the Phonology of ω in well backed principles.

(96) [Ainarak ardoa erosi zuen]$_F$
Ainara wine buy AUX
'Ainara bought wine.' Irurtzun (2009)

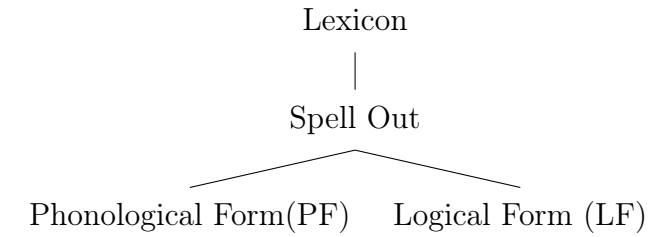
(97) Ardoa [Ainarak]$_F$ erosi zuen
wine Ainara buy AUX
'Ainara bought wine' Irurtzun (2009)

5.3 Y Model

We will now look at the Y model of grammar. The fundamentals of the Y Model can be seen in (98), where we can see that the path for the creation of an utterance involves the

lexicon, going through spell out where Syntax applies, and from there it branches into the path of Phonology, in the PF, and the Logical Form, LF, which is where Semantics applies to the utterance.

(98)



This structure easily handles ω at the syntax level as it is a phrase level object and is then able to progress to the PF where it is interpreted as taking on the base form of the highest C-commanding Antecedent in its binding domain. Or in a language like Hebrew in (113), it is suppressed from receiving a form by the syntax. The largest difference from the linear model we discussed earlier is that the semantic representation of language is relegated to a post-syntax position rather than being a pre-syntax operation[5]. This then leaves us with the question of Thai, in phrases like (90) and (91) where we have accounted for the ungrammaticality of the Thai phrases by claiming that in Thai ω must take only the highest projection of its highest C-commanding antecedent in the binding domain, this is also able to apply to the Y model as we can see in (92) and (93), as it is handled in the syntax as it limits at what level the ω object is placed above an item in a nominal phrase.

As the syntactic object created from the lexicon is sent to the PF we are able to generate sentences like (89), as the phonology is able to put the bound morpheme that the pronoun represented by ω needs to exist in a free form phonologically.

Fundamentally, for our purpose there are but few differences between the Y Model and the Linear Model, using the linear model as a way to examine the Syntax of the ω we are then able to see that the phonology works out identically for both as we are working with a formed syntactic representation. the differences of note lay entirely within semantics, which may effect the issue of quantification, but this difference is not relevant as it will invalidate the quantification of ω in the same way, merely at a different place in the structure.

6 Quantification and Conjunction

In this section we will see why ω cannot be quantified, conjoined, and why it can be suppressed by various elements of a phrase. We will see that ω is easily blocked by features on the various elements of the phrase which render it superfluous in the syntax. Since we will follow a principle of least computation for the amount of data that needs to be overtly spoken, Kul (2007). We will begin by looking at quantification proper.

[5]I would like to express my personal disagreement with this way of approach syntax and semantics. The nature of the lexicon in the Y model is perhaps too dependent on the words themselves being relevant rather than the event arguments we use language to describe. It is my opinion that the linear model is a place where further study is needed as the relegation of semantics to the question of Quantification leaves much of its explanatory value in the representation of event arguments and the fundamental structure of events pre-syntax undervalued in the current literature.

6.1 Quantification

We will begin by looking at Quantification. In (99) we can see that in Zapotec, the strategy for quantified expressions is to raise the main subject to the topic position and then use a Pro-ϕ with appropriate ϕ features which is then used in conjunct with an ω to form a quantified reflexive.

(99) B-guhty cho'nn bxuuhahz cho'nn ra bxuuhahz
PERF-kill three PL priest PL priest
'three priests killed themselves' Lee (2003)

As well, in Hmong we are not allowed to have bare nouns as our ω copies, as we can see in (100).

(100) Dlev pum dlev
dog see dog
? 'Dogs see dogs.'
?? '(a) dog sees itself.' (Mortensen, 2004)

Why would this be a similar phenomena to that of the quantifier in (99)? The answer is in the idea of set denotation. When we have a quantified set of nouns like we do in (99) we are looking at sets, rather than entities. Something the ω must do is denote a specific set of individuals not an open ended one. In SLQZ we have seen that it is possible to quantify a plural like in (101), how can that be if we restrict ω to only quantifying individuals?

(101) R-yu'lààa'z ra bxuuhahz ra bxuuhahz
HAB-like PL priest PL priest
'The priests like themselves' (Black, 1994)

In SLQZ, the plural morpheme is able to be added because bare nouns in SLQZ are individual denoting instead of set denoting. This then means that the addition of the plurality marker does not interfere, or block, the ω because it is a set of plural entities, as well it is entirely within the domain of the DP, acting as the specifier to the DP, while *priests* as the NP in the phrase. But, when we attempt to use other items of quantification than the plural morpheme we are not able to do so as we can see in (102), where even though this is a somewhat defined set, the use of the Quantifier Phrase instead of a simple plurality marker has caused issues with the semantic mapping of the ω onto the lower position, but it is not a lethal complication, merely one that a speaker would not generate.

(102) ??R-a txup tson wnaa r-ka txup tson wnaa gyus
HAB-go two three woman HAB-buy two three women pot
'A few women went to buy a pot' Black (1994)

As well, in (102) we can see that there is quantifier in *two three*, but we are lacking a

determiner, but in SLQZ when one attempts to use a quantifier instead of the DP level plurality marker it is not allowed, and instead must be used in a focus position. We can see this in (103).

(103) Cho'nn ra bxuuhahz b-guhty-rih la'arih.
three PL priest PERF-kill 3PL.DIST 3PL.DIST
'Three priests kill themselves Lee (2003)

This then shows us, as there is both an overt QP *three* and that the plural marking is in the DP position, as it is able to exist without a quantifier as we can see in (101), this then supports out current theory that the ω is able to take on items located in the DP domain, but not those above it, like QPs or CPs. We can also see this data working with CPs in the *wh*-movement context in Hmong, where ω is allowed in some contexts.

(104) [tug twg]$_i$ los nwg$_i$ nyam nwg$_i$
CLF which TOP 3SG like ω
'Anybody would like themselves' Mortensen (2004)

In (104) we can see that Hmong must have a raised, and separate CP for ω to be licensed, ω cannot take from items in a CP to find its reading, that is too high in the phrase to be licensed. As well, we will see shortly that there are a number of other ways that ω can be blocked from having a phonological representation at all, beyond what we have seen in Hebrew.

In this last section we have seen how basic quantification works with ω, in the following we will look at how ω handles conjunction with another item via an *and*. In the following we will see that conjunctions are not allowed, and that they are part of a broader trend of restricting ω.

6.2 Conjunction

Let us begin by looking at data from Hmong and SLQZ where we can see that it is not allowed in either language, in a wide variety of situations.

(105) *Pov qhuas Maiv thiab Pov
Pov praise Maiv and Pov
'Pov Praises Maiv and himself.' Boeckx et al. (2007)

(106) *Pov qhuas Pov thiab Maiv
Pov praise Pov and Pov
'Pov Praises himself and Maiv.' Boeckx et al. (2007)

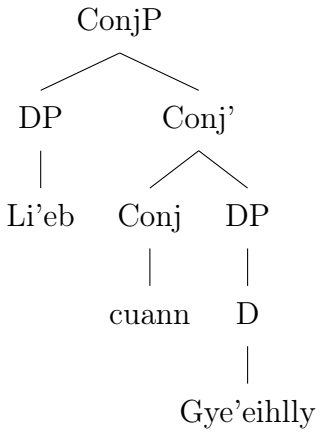
(107) *R-yu'lààa'z Li'eb cuann Gye'eihlly Li'eb cuann Gye'eihlly
HAV-like Felipe and Mike Felipe and Mike

'Mike and Felipe like themselves.' Lee (2003)

In the previous examples we can see just how impossible it is to have ω in a conjoined sentence. Why could this be? We will explore a few different options: an inability of ω to cross the *and*, and take on the phonological reading, a feature of the *and* which blocks the ability to take on the reading, as Lee (2003) proposed a semantic type clash as well as island violations as a potential explainer.

First, we will begin with the idea that *and* is somehow a level in the syntax that the ω cannot pass through. Let us begin this with a Tree for reference, in (108).

(108)



From this structure we can see that the Conjunction phrase on the top of the structure may pose an issue for ω as it attempts to chose the target for copying, having two DPs as options within the phrase could simply be too much for it to sort through without causing undue computation for the speaker. What we must now consider is if there are any issues that could arise from the idea of the ConjP blocking ω.

The first question that we will address is which phrases might block ω and via what mechanism. Currently we are working from the idea there is an inherent computational issue in the ConjP that leads to the ungrammaticality that we see. This then could perhaps be derived from the issues we saw with quantification. This then, would leave us with an undefined, and without far more speakers of at least two languages to run data by, unverifiable claims about types of phrases which do not license ω.

Next, we will consider the idea of a feature on the head of the ConjP which can block the voicing of ω. As we briefly discussed in an earlier section, there are a set of items which can block the overt pronunciation of ω, we have seen it in modern Hebrew, with the intransitive morpheme, which blocks the pronunciation of ω, as it is then superfluous to the sentence. We will now see this again in SLQZ, with *-ni* and referential possessive.

(109) B-to'oh Gye'eihlly x:-ca'rr-ni
PERF-sell Mike GEN-car-REFL.POSS
'Mike sold his own car.' Lee (2003)

This data allows us to see that, where we would expect an ω, it is instead replaced by *-ni*,

this is even more clearly demonstrated in (110), which shares very many similarities with (11).

(110) Bùunny nih r-umbèe Li'eb g-uhcnèe behts-ni'
man REL HAB-know Felipe PERF-help brother-REFL.POSS
'The man$_i$ who knows Felipe helped his$_{i/*j}$ brother' Lee (2003)

This allows us to see that there is an interesting case of blocked ω happening in these sentences. Since we must work with Hebrew, and this case in mind, what then prevents us from assuming that *and* simply has this feature? The key difference between *and* and the data that we have in (110) and (109) as well as in (113) is that there is always a marking that ω exists in the things that block it from surfacing. There is a clear reason why it is expected under the Grice's Maxims, particularly quantity, that we would not want to pronounce both the ω and the items that we have seen block it. No such clarity of reason applies to this case with *and* as it does not need to be seen as superfluous with the ω when it is used, as both will still be contributing to the meaning of the phrase. Therefore, we cannot follow the path of there being a feature in the *and* that blocks ω.

7 Conclusion

We have seen that ω is a feature of 5 unrelated and typologically distinct languages' grammars, making clear that it is a feature within Universal Grammar and is a pronoun that is a logical extension of the syntactic structures within various languages. ω is a phonologically underspecified pronoun which receives its phonological representation as the base form of the highest C-commanding antecedent within its binding domain. Within it, ω does not have any ϕ features, instead it is merely co-referent with its highest C-commanding antecedent, as we saw in 4.2. As well, because of this ω is able to exist as the direct object of a phrase (see 4.1.1) as well as the subject of an embedded phrase within an utterance (see 4.1.2). This then makes clear for us that ω is not carrying any ϕ features, which other scholars have claimed. We can see in 4.2 that there are languages where ω and pro-ϕ phrases are overlapping, we can see this in (111) from Jambi Malay, which shows clear evidence that in this language ω and Pro-ϕ phrases are licensable.

(111) Dio$_1$ cinto dio$_{1/2}$
he love he
'He loves him' or: 'he loves himself' Cole et al. (2015)

After we have seen that it is certain to exist, we must now consider how it is created. I propose a novel method to conceptualize the process of taking the semantics of an ω bearing phrase and turning it into syntax and then phonology via spell out. The core of this argument is the idea that at spell out the underspecified phonology in the ω is able to reach up the tree of its C-commanding antecedents and pull an element from that as its base phonological form. The pulling of a single element allows us to explain cases like Thai, where only

the highest element in the highest C commanding antecedent can become the phonological representation of ω (see 5). This novel approach changes the orthodoxy of the most common approach, the Y-model, but the core of the argument works just as well in the Y-model as in a linear path like has been laid out in 5.2.

We then look at the idea of Quantification, for ω cannot be quantified, nor can it be conjoined. To start with quantification, we see that ω cannot be qauntified because it is a semantic type clash. This is because ω requires its C-commanding antecedent to not be set-denoting. This helps us to explain cases like (112), from Hmong, where an equivalent would be licensable in Zapotec, this is because in Hmong bare nouns are set denoting and creates a situation where an ω reading is not licensed, but a reading with the second instance of the highest element is its own independent representation and denotes its own set.

(112) Dlev$_i$ yeej tum Dlev$_{*i/j}$
dog always bit dog.
'Dogs always bit dogs.'
$*$'Dogs always bit themselves' Mortensen (2004)

We then move onto to considering the case of conjunction, which ω cannot be part of. Previous analyses of the ω phenomena have had difficulty with this as they are hard to account for without the features of pronouns. Namely that there is an island violation when the languages with ω attempt to reach up to the C-commanding antecedent to locate a desired phonological reading, this is in part because of the head of the phrase not being a nominal, instead it is a Conjunction Phrase with a nominal as a specifier. This does not lend itself to ω.

The final item of ω that we have considered is what are the extensions of it. We have seen that it is likely to exist in a language like Hebrew in (113), where there are reflexives that are created by suppressing ω using a "intransitive" marker on a transitive verb that tells the listener that it is being used reflexively.

(113) Jitsxak hitaper.
Yitzhak made.up.INTNS.MID
'Yitzhak put on make-up' (*'Yitzhak got make up applied to him), Kastner (2017)

We know ω is still present in this because of the way these words work when they are not modified with the "intransitive" morpheme. This tells us that semantically and lexically, we still need something in the place that ω is because the verb still requires an object as it has the same transitive meaning, ω is merely suppressed from reading upwards by the intransitive morpheme, which tells the listener that it is reflexive and makes an overt ω superfluous.

8 Further Research

There is however more research to be done. Particularly focusing on the exact workings of it in Vietnamese and Thai, as the data for these languages can be sparse, particularly there

are few documented examples of ω copying a pronouns in these languages, which would help us to understand the phonology at play more.

As well, we predict that ω will not be able to copy complex constructions like a VP or ConjP as its phonological reading. We have evidence that this is the case for the Conjunction phrase, but I have not included data on complex subjects like a VP or event argument as the subject of a phrase. This requires further research and data gathering from the languages in which ω is present, particularly non Zapotec languages like Hmong or Vietnamese as the literature is lacking in these areas.

Further research into the role of ω in embedded subject clauses is needed, particularly with a focus on the issue of null subjects. The issue of how ω and pro interact has not been something that this thesis has been able to examine, as such further research is needed. However, from what we have seen from Thai, Vietnamese, SLQZ and Hmong, it is clear that ω represents a distinct pronoun within these languages.

References

Bisang, W. (1993). Classifiers, quantifiers and class nouns in hmong. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language"*, 17(1):1–51.

Black, C. A. (1994). *Quiegolani Zapotec syntax*. University of California, Santa Cruz.

Boeckx, C., Hornstein, N., and Nunes, J. (2007). Overt copies in reflexive and control structures: A movement analysis. *University of Maryland working papers in linguistics*, 15:1–46.

Chaipet, K. and Jenks, P. (2021). Names as complex indices: On apparent condition c violations in thai. In *NELS*, volume 51.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Praeger.

Cole, P., Hermon, G., and Yanti (2015). Grammar of binding in the languages of the world: Innate or learned? *Cognition*, 141:138–160.

De Alencar, L. F., Kelling, C., Butt, M., and King, T. H. (2005). Are reflexive constructions transitive or intransitive? evidence from german and romance. In *Proceedings of the LFG05 Conference*, pages 1–20. CSLI Publications Stanford.

Déchaine, R.-M. and Wiltschko, M. (2002). Decomposing pronouns. *Linguistic inquiry*, 33(3):409–442.

Galloway, B. D. (1993). *A Grammar of Upriver Halkomelem*. PhD thesis, University of California, Berkeley.

Haspelmath, M. (2019). Comparing reflexive constructions in the world's languages. *Ms. MPI-SHH Jena & Leipzig University, Germany*.

Irurtzun, A. (2009). Why y: on the centrality of syntax in the architecture of grammar. *Catalan journal of linguistics*, pages 141–160.

Ivan, R. and Bui, T. (2019). Vietnamese anaphora: Binding principles and the lack thereof.

Jenks, P. S. E. (2011). *The hidden structure of Thai noun phrases*. PhD thesis, Harvard University.

Kastner, I. (2017). Reflexive verbs in hebrew: Deep unaccusativity meets lexical semantics. *Glossa: a journal of general linguistics*, 2(1).

Kul, M. (2007). The principle of least effort within the hierarchy of linguistic preferences: External evidence from english. *Unpublished PhD dissertation, Adam Mickiewicz University, Poznań*.

Lai, I.-J. S. (1998). *The grammar and acquisition of Secwepemctsín independent pronouns*. Text, University of British Columbia.

Larson, M. (2006). The thais that bind: Principle c and bound expressions in thai. In *PROCEEDINGS-NELS*, volume 36, page 427.

Lasnik, H. and Stowell, T. (1991). Weakest crossover. *Linguistic inquiry*, 22(4):687–720.

Lee, F. (2003). Anaphoric r–expressions as bound variables. *Syntax*, 6(1):84–114.

Mortensen, D. R. (2004). Two types of variable elements in hmong anaphora. *University of California, Berkeley*.

Narahara, T. (1995). Alternatives to reflexives in thai and vietnamese: Binding theory and language variations. In *Papers from the Third Annual Meeting of the Southeast Asian Linguistics Society, Tempe, Arizona: Arizona State University, Program for Southeast Asian Studies*.

van Blankenstein, Z. (2021). Condition c violations in thai and st'át'imcets. *Working Papers of the Linguistics Circle*, 31(1):48–57.