UC Santa Cruz
UC Santa Cruz Electronic Theses and Dissertations

Title

The Distribution AND Identity of Bare Indefinites in San Martín Peras Mixtec

Permalink

https://escholarship.org/uc/item/41d1t3hk

Author

Tuffs, David William

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

SANTA CRUZ

THE DISTRIBUTION AND IDENTITY OF BARE INDEFINITES

IN SAN MARTÍN PERAS MIXTEC

A thesis submitted in partial satisfaction
of the requirements for the degree of
MASTER OF ARTS

in

LINGUISTICS

by

David Tuffs

March 2024

The Thesis of David Tuffs
is approved:

_____
Professor Ivy Sichel, Chair

_____
Professor Roumyana
Pancheva

_____
Professor Maziar
Toosarvandani

_____
Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

ABSTRACT

THE DISTRIBUTION AND IDENTITY OF BARE INDEFINITES IN SAN MARTÍN PERAS MIXTEC

David Tuffs

San Martín Peras Mixtec has two elements which function as indefinites without quantifiers – however, despite superficial similarities, the identities and distributions of these indefinites differ. *Na* ("someone" or "them") is a clitic, while *ña'a* ("something," "thing," or "things") is not. Both can be described as bare nouns that are number neutral. Through elicitation sessions with a native speaker, we can make some conclusions about the distribution of *na* in particular. *Na* is generally prohibited clause-initially, but it can raise to this position when it receives an additional tone. Conversely, it cannot occur after other raised elements. This paper proposes that the facts of *na* can be explained as effects of an intonational phrase which maps to the clause, which requires that its first element be a full prosodic word.

1 Introduction[1]

The identity and behavior of clitics in Mixtec (and specifically the dialect of San Martín Peras Mixtec, or SMPM) have received considerable attention in recent years, though in different capacities. Eischens (2020) has proposed that the negation particle *ko* is a clitic which leans on elements to its right, while recent research in Ostrove (2023) has noted a set of clitic pronouns for local and non-local persons which lean to their left. This paper will focus on the identities of two indefinites which seem to display clitic-like properties. *Bare indefinites* comprise the indefinites *na* (someone) and *ña'a* (something) and both seem to show some level of restriction clause-initially. However, while we see a genuine ban on *na* at the start of clauses, there are instances in which *ña'a* can appear clause-initially, without additional material attached. In contrast with *ña'a*, our analysis will show that *na* behaves predictably as a clitic. To explain the patterns that *na* shows, we will use the intonational facts of the clause and detail how *na* is not satisfactory when it is clause-initial. In addition, we will lay out a more complex group of indefinites, based on *na* and *ña'a*, which can readily appear clause-initially.

In analyzing the distribution of bare indefinites, we will utilize two other areas of Mixtec which have also received scholarly attention, namely its liberal use of raising and its complex negative forms. Raising to focus has been discussed in

Hedding (2021) and Hedding (2022), which centered on the semantics of alternate

particles, but here it is useful as a method of moving a variety of nominals, clitic and

non-clitic, to various positions, both clause-initial and non-clause initial. Of course,

raising to focus is not the only way that indefinites can move to early positions - this

paper will also examine how raising to C in interrogative contexts interacts with how

na raises. SMPM's process of negation which involves *ko* and a rising tone, and

which is sensitive to verbal aspect, has been analyzed by Eischens (2020) from a

morphological perspective, but it is useful in our analysis too, as it provides structure

early in the clause with which can interact, and it allows us to see a more complete

picture of the pre-verbal positions that phrases can raise to.

While this paper will greatly utilize both the facts of negation and the facts of

raising, our goal is to describe the behavior of *na* and *ña'a*. Semantically, *na* can be

the indefinite "someone" or the third person plural pronoun "them". *Ña'a* on the other

hand does not have a definite form, instead manifesting as "something," "thing," or

"things," all of which have indefinite interpretations. Notably, despite this difference,

both *na* and *ña'a* are number-neutral, and do not display a consistent singular or

plural meaning. As a result, we can identify both elements as number-neutral bare

nouns, though the distributions of these bare nouns differ.

There are also two situations in which the rule against clause-initiality for *na*

breaks. First, *na* can occur clause-initially in certain interrogative or negative contexts

- we will explain this through the attachment of a second tone. Importantly, this

exception works within a full ban on clause-initiality when *na* is on its own - in

contrast *ña'a* appears clause-initially outside the cases where tone is involved.

Second, *na* may not follow another raised element, meaning that there are non-clause initial positions where it is also disallowed - we will ascribe this to the requirements of the intonational phrase.

Overall, the goal of this paper – beyond sketching the distribution of a set of indefinites – is to in part synthesize previous clitic analyses in SMPM, by analyzing how *na* interacts with *ko* directly, and how *na* and *ña'a* are similar to and different from clitic pronouns. This paper will also detail the extent to which *ña'a* shows restrictions clause-initially, and explain why we do not consider this universal. Section 2 of this paper will provide an overview of the relevant areas of San Martín Peras Mixtec - movement, indefinites, and negation. Section 3 will introduce the central issue and build upon the facts in Section 2, defining the types of DPs that *na* and *ña'a* are, and discussing the exceptions to the general pattern. Section 4 will introduce a clitic analysis for *na*, and explain how we can explain the exceptions. Section 5 will conclude and provide areas for further research.

1.1 Methods

The novel Mixtec data presented in this paper were collected remotely through elicitation sessions with a native speaker of SMPM remotely over Zoom. Tonal data is included in the examples for completeness, but is often influenced by tone sandhi processes which are not the focus of this paper. A more thorough account of tonal behavior in SMPM can be found in Eischens (2022). Sentences were elicited

3

from the native SMPM speaker in one of two ways. Some sentences were presented in Spanish for the purposes of a translation in Mixtec, while others were presented in Mixtec as test sentences, which were judged for grammaticality and meaning, and were often judged for well-formedness, similarity to other sentences, and possible contexts in which they could be used. The majority of sentences which I provided in Mixtec were repeated back to me by the speaker, and when possible, these are the sentences used for notes on tone. To most closely match the speaker's interpretation of sentences we will be glossing *na* as 'INDF +hum' to indicate the human indefinite 'someone' or as "3pl" to indicate the third person pronoun "them". *Ña'a* will be glossed as 'INDF -hum' to indicate the interpretation of 'something' and as 'thing' or 'things' where appropriate.

2 The Syntax of Mixtec

San Martin Peras Mixtec (SMPM in this paper) is a Mixtec language in the Mixtecan branch of Oto-Manguean languages, along with Triqui and Cuicatec. Mixtec is spoken by around half a million people in various regions of Mexico (Caponigro, et al. 2013), while SMPM in particular is spoken by around 12,000 people in the region surrounding the town of San Martín Peras in Oaxaca, and in Central California (Eischens, to appear). This section will discuss three topics in SMPM: word order, indefinites, and negation. The discussion of word order and its interaction with raising will be important to illustrate where bare indefinites and their derived forms can appear, and a distinction will be made with regard to this in section

4

3. Negation impacts the facts of both movement and indefinite forms, and interacts with the latter via negative concord.

2.1 Word order

Mixtec has a VSO word order, and when both subjects and objects appear after the verb, this order must be preserved. Mixtec, however, does diverge from this schema when one or more arguments raises above the verb. As shown in (1), only the subject-object reading is available when both DPs are in situ; the object-subject counterpart ("Pedro is hitting someone") is not present.

(1)   kānī   í=nā   pēdrò

        hit.CONT   one=INDF+HUM   Pedro

        "Someone is hitting Pedro"

        *"Pedro is hitting someone"

Word order in Mixtec is variable in cases of raising, which is highly productive and can be used to focus arguments and adjuncts in positions early in the sentence (Hedding 2022). We will describe raising of verbal arguments as a focus move similar to Hedding's, but with an additional position to which phrases can focus. We will thus be including two proposed positions for focus raising, one which occurs before negation (*ko*), and one which occurs after. The sentences below show the schema for focus raising which we will be using.

Pre-negation raising to focus

(2)   **pedro$_i$**   [$_{NEG}$ ko   [$_v$nomi   __$_i$ =na]]

       Pedro   NEG   hug.CONT   INDF+hum

       "Pedro is not hugging anyone."

Post-negation raising to focus

(3)   ni-in=na$_i$ [$_{NEG}$ ko   **pedro$_j$**   [$_v$nomi __$_i$ __$_j$]]

       not-one=INDF+hum   NEG   pedro   hug.CONT

       "No one is hugging Pedro."

When negation is not present, both raising positions manifest simply as a pre-verbal argument.

(4)   **pedro**   [$_v$kani   in=na

       pedro   hit.CONT   one=INDF+hum

       "Pedro is hitting someone"

We will claim that instances of raising to focus are not contingent on the clitic status of the moving objects, and instead can apply to DPs across the board.

While it is useful to follow Hedding's approach, the question of whether or not this is focus movement is necessary to address for our final proposal that *na* is a clitic. Since focus appears to be the most promising explanation for this movement, we will need to align this choice with the overall clitic patterns described in

Cardinaletti and Starke (1999) which detail a general dispreference for clitics with semantic focus. Importantly for us, this is not a universal rule, as is the case in French, where, according to Cardinaletti and Starke, 'deficient pronouns,' their term for the group which includes clitics, can be semantically focused, both with stress on the pronoun and with flat intonation.

Multiple DPs can move to a preverbal position in positive sentences, as shown in (5). In examples like (5), we can assume that one DP has moved to the higher focus position and one has moved to the lower one.

(5)   **í̵=nà**   **í̵**   **ñā̄ʔā̀**   kánī

        one=INDF+hum   one   INDF-hum   hit.CONT

        "Someone is hitting something"

We will see later that DPs which raise to higher positions likely raise outside the scope of the intonational phrase, which maps to a CP (Ishihara 2022).

2.2 Indefinites

Mixtec has a set of words which we will define as bare indefinites. This group comprises the elements *na* and *ña'a*. While we will ultimately find that these elements exhibit conflicting behaviors, this is a good place to start, as they can have

indefinite meanings and can be used to form larger indefinites. These words have a few important features to note. First, they have a similar broad distribution - they can appear in situ or raised, and they can appear in positive and negative sentences, though only *ña'a* can raise pre-verbally in positive contexts.

In situ (positive)

(6) í=nà kánì **ñā̄ʔā̀**

one=INDF+hum hit.CONT INDF-hum

"Someone is hitting something"

(7) kani pedro **=na**

hit.CONT pedro =INDF+hum

"Pedro is hitting someone"

In-situ (negative)

(8) ni-í=nà ko kani **ñā̄ʔā̀**

not-one=INDF+hum NEG hit.CONT INDF-hum

"No one is hitting anything"

(9) kánī **=nà** Márìá

hit.CONT =INDF+hum Maria

"Maria is hitting someone"

Raised (positive)

- Only attested with *ña'a*

(10) ña'a   kani   Pedro

      thing   hit.CONT   Pedro

      "Pedro is hitting things"

Raised (negative)

(11) kǒ=**nā**   kánī   =nâ

      NEG=INDF+hum   hit.CONT   INDF+hum

      "No one is hitting anyone"

(12) ni-�ī   ñā̠ʔã̠   ko   **ñā̠ʔà̠**   kani

      not-one   INDF-hum   NEG   INDF-hum   hit.CONT

      "Nothing is hitting anything"

Bare indefinites often occur without quantifiers, but they can also occur in constructions with them. The quantifier *in,* meaning 'one,' forms positive indefinites, which are restricted to positive sentences (13) and prohibited in negative sentences (14) [2]

(13) **ɪ́=nà**   kānì   ɪ́   ñā̠ʔã̠

      one=INDF+hum   hit.CONT   one   INDF-hum

      "Someone is hitting something"

---

[2] Unlike bare indefinites, which can appear on their own within a sentence, *in* needs something (virtually always a bare indefinite) to follow it.

(14)   *kò   kánì   pēdɾò **í=nà**

NEG   hit.CONT   Pedro one=INDF+hum

 Intended: "Pedro is not hitting anyone"

Ña'a also co-occurs with *in*, forming *in-ña'a*, is formed the same way, also utilizing the word for 'one'. As seen in (13) above (in the in-situ position), it is able to appear in positive sentences, and like *ni-in-na*, it is prohibited in negative sentences as well, as in (15).

(15)   *kō   kìshà   =nà   í   ñā²ā̀

NEG   hit.CONT   =INDF+hum one   INDF-hum

"Noone is hitting anything."

These two indefinites also have negative versions. *In-na* has *ni-in-na* for 'no one', while *in-ña'a* has two: *ni-ña'a* or *ni-in-ña'a* for 'nothing'.

(16)   **nī-í̄-nà**   kò   kánī   pέdɾō

not-one=INDF+hum NEG   hit.CONT   Pedro

"No one is hitting Pedro"

(17)   **ni   í̄   ñā̄²ā̀**   ko   ñā̄²ā̀   kani

not   one   INDF-hum   NEG   INDF-hum   hit.CONT

"Nothing is hitting anything"

(18) **ni**   **ñã̄ˀã̃**   kò   kánī   nī-í̄=nà

not   INDF-hum   NEG   hit.CONT   not-one=INDF+hum

"No one is hitting anything"

The following table shows the selection of indefinites which we will be using

in this paper. The presence of quantifiers on *na* and *ña'a* means that they become

isolated to either a positive or negative context.

	nà, ña'a	*in=nà, in=ña'ã*	*ni-in=nà, ni-(in)=ña'ã*
In positive sentences?	✓	✓	
In negative sentences?	✓		✓

Figure 1: A table showing three types of indefinites present in SMPM

The element *ni* which occurs at the start of every negative indefinite can be

shown to have intrinsic negative meaning. These negative indefinites only appear in

negative contexts: sentences which also have sentential negation. They do not appear

in positive constructions, nor do they appear in characteristic environments for

negative polarity items, such as questions and conditional sentences, and as such *ni*

has been labeled a negative element in other analyses, which we will follow here

(Eischens 2020, Israel 2011). In the conditional and interrogative examples below,

(19) and (20), *ni-in-na* is not grammatical.

11

Conditional Sentences

(19)   *ātồ   **nī- í:=nà**   =ná   kānì pēdɾò,   ⁿtà ījīrà

if   not-one=INDF+hum   =INDF+hum   hit.CONT? Pedro   cry.POT?

Intended: "If someone hits Pedro, he is going to cry"

Interrogative Sentences

(20)   ***nĩ-ĩ:=nã**   kánì   péðɾò

not-one.INDF+hum   hit.CONT   Pedro

Intended: "Is Pedro hitting anyone?"

2.3 Negation

Mixtec generates sentential negation through the element *ko* (Eischens 2020). While negative indefinites are limited to negative contexts, they do not contribute an independent negative force to sentences in the way that *ko* does. As a result, Mixtec has been defined as a negative concord language (Eischens 2020). Negative concord in Mixtec does not require that indefinites be negative in negative contexts (*na* and *ña'a* are allowed in them, despite their lack of negative morphology), but it does prohibit them from contributing to their own negative meaning, even if they begin with a negative element. (Eischens 2020) Thus a sentence like the one below has one semantic negation present, rather than double negation, despite the presence of multiple negative elements.

(21) nī-ǐ=na   kō   ní   ǐ   nā'ǎ   kânī

not-one=INDF+hum NEG   not   one   INDF-hum   hit.CONT

"No one is hitting anything."

Now that the interaction between sentential negation and the indefinites discussed in the previous section has been clarified, we will spend some time fleshing out the former. Sentential negation in SMPM can be expressed in multiple ways, using phonetic elements and tone. Negation can appear as the element *ko* which precedes verbs obligatorily in continuative aspects and optionally in completive aspects (Eischens, to appear).

(22) **kō**   kánī   pɛdɾō nī-ǐ=nā

NEG   hit.CONT   Pedro not-one-INDF+hum

"Pedro is not hitting anyone"

(23) **(kò)**   ní-ʃāʃǐ   nā ⁿtsīkǎ

(NEG) *ni*-eat.COMP   3f   banana

"Noone ate a banana"

Mixtec uses five tones ("low, middle, high, rising and falling", per Eischens (2020)) for a variety of grammatical functions, including verbal aspect. Sentential negation can also be expressed as a tone, as is visible in potential and some completive constructions (those in which *ko* is absent).

(24)   **nǐ-ʃakù**   Eischens (to appear)

NEG-laugh.COMP

"didn't laugh"

(25)   **kʷǎkù**   Eischens (to appear)

laugh.NEG.POT

"will not laugh"

The following table lays out how negation behaves in different aspects

	Continuative	Completive	Potential
Rising Tone?	Yes	Yes	Yes
Ko?	Yes	Sometimes	No
Tone falls on	*ko*	*Ko* (spreads to *ni*) *ni* (if *ko* is absent)	First syllable of the verb

Figure 2: A table comparing the form of negation in different aspects

Negative nominals like *ko=na* utilize the marker of sentential negation, but carry the meaning of *no one* or *nothing.* The *ko* in negative nominals is described by Eischens to be an instance of the same element used for sentential negation. It is not possible for both a negative nominal using *ko* and sentential negation using *ko* to appear in the same sentence, as this would require duplication of the sentential negation particle, which Mixtec does not have the capacity to do.(Eischens 2020) (26) illustrates this.

14

(26) ***kō** ñā yìβī **kò** kánì pēdɾò Eischens (2020)

NEG 3f person NEG hit.CONT Pedro

Intended: "No one is hitting Pedro."

To bring things back to the raising facts discussed in 2.1, multiple elements

can raise, but only one may follow *ko* (see example 27). Multiple DPs raising

between the NegP and the verb is not possible. This means that only one raised

position is available after negation, but the ability for DPs to appear before and after

ko means that this is not a position that every DP must raise through - otherwise

traces would prevent filling post-*ko* raising positions if pre-*ko* raising positions are

filled (see example 20 for an instance of this).

(27) *[~NEG~kō ní-î́=nà ní ñā²ǎ̀ [~v~kānì]]

NEG not-one=INDF+hum not INDF-hum hit.CONT

Intended: "No one is hitting anything."

2.3.1 *ko* as a clitic

Ko has been described as being able to attach to elements to its right, as well

as to pass its tone to such elements, which have led to an analysis of it as a clitic by

Eischens. In that analysis, *ko* can attach to verbs and raised DPs, and to adverbs as a

rising tone (Eischens, to appear). While it still contributes sentential negation, it can

adjoin to DPs can create negative indefinites such as *ko=na*, which can be interpreted

on their own as 'no one,' as an example. Eischens' analysis of *ko* as a clitic has dealt with its morphological idiosyncrasies among verbal aspects by relegating them to the morphology. Indeed, *ko* does not seem to be an affix to the verb, as raised DPs would not be able to come between the two if this was the case (and orders like *ko=na kani* would not be attested).

While our analysis does not require *ko* to be a clitic, it also does not forbid it from being one. Despite the fact that *na* attaching to *ko* would lead to two clitics essentially leaning on each other, there are other similar proposals in other languages, such as an analysis of Alsea presented in Bošković (2001). Bošković notes examples like *k=in* from Alsea, in which a proclitic and an enclitic element are able to form a prosodic word, despite neither being a prosodic word on its own. Another possibility is that *ko*, while not a full prosodic word, is large enough to serve as a host for the clitic *na*. *Na*, being an internal clitic in this case (following the description of internal clitics in Anderson 2005) would form a single novel prosodic word with *ko*. As a result, despite the expected facts that a clitic *na* would not align with clitic *ko*, there are options which allow the two to function together. The formation of a prosodic word via a host and a clitic is useful, as the prosodic word is the unit which we will propose must start an intonational phrase (a.k.a. a clause), so the fact that forms like *ko=na* occur clause-initially is expected in our analysis.

2.4 A framework for the syntactic structure of Mixtec sentences

Now that we have the individual ingredients of Mixtec syntax, it is beneficial

to describe how they piece together to form the structures in which raising and

na-cliticization occur. We will start from the verb phrase and work upwards. Mixtec

verb phrases are VSO, and we will follow Eischens (to appear) in assuming a vP

construction with a verb head that raises and a subject and object (in the specifier and

complement position of VP respectively) that remain in situ. This aligns with

structures proposed for other VSO languages (Carnie, et al. 2005). A verb phrase with

no focus raising would have the structure shown below, for the sentence "*kani in-na*

Pedro"
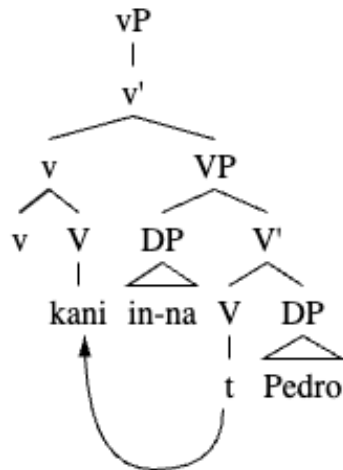
Figure 3: A tree for a VSO verb phrase in SMPM

We will place *ko* in a NegP outside of the verb phrase, as it precedes adverbs.

SMPM does not have overt tense or aspect material, which makes the exact

positioning of NegP hard. However, we can note that, as would be expected for

sentential negation, it occurs underneath CPs, such as question CPs in example (28) below.[3] Following a combination of Eischens (2018) and Alqassas (2019), we will write *ko* as a NegP as a way of integrating it into the structure of the sentence, and a way of situating it before the verb and before focus positions. As Mixtec expresses aspect on the verbs, there is a lack of overt Aspect heads in the data, so the position of NegP with relation to Aspect is uncertain - the trees in this paper (such as (31)) place it above Aspect, but the inverse could be true, and we would not expect this to change the facts.

(28) jo   ko   kani   Pedro

      who   NEG   hit.CONT   Pedro

      "Who is not hitting Pedro?"

The focus position following *ko* will also be outside of the verb phrase, as indefinites cannot raise between adverbs and verbs (as shown in (29) and (30)). We will not follow a possible alternate theory in which raising DPs land in the specifier position of vP, as this is often associated with specific theta roles, and Mixtec can readily raise agents and themes into this position (Carnie et al. 2005).

(29) kǒ=nā   kāmà   kánì   pedro

      NEG=INDF+hum   quickly   hit.CONT   pedro

---

[3] We will use interrogative CPs to make this point rather than regular complementizers as SMPM does not have an overt version of 'that.' However, the non-interrogative complementizer facts, while they do not inform the analysis much, do not contradict it either.
(i)   kúntā'ìní maria kǒ=nā   kānì   pedro
        know   maria NEG=INDF+hum hit.CONT pedro
        "Maria knows that Pedro is not hitting anyone"

"Pedro is not hitting anyone quickly"

(30)　　*(kǒ)　　　　kāmá　　　　=nà　　　　kàní　　　　pedro

　　　　(NEG)　　　quickly　　　=INDF+hum　hit.CONT　　pedro

Intended: "Pedro is not hitting anyone / is hitting someone quickly"

(31) shows where this leaves us - Focus and Negation are located outside of the verb phrase, but within the same clause as the verb phrase. VSO constructions take place entirely within the VP and vP.

(31)
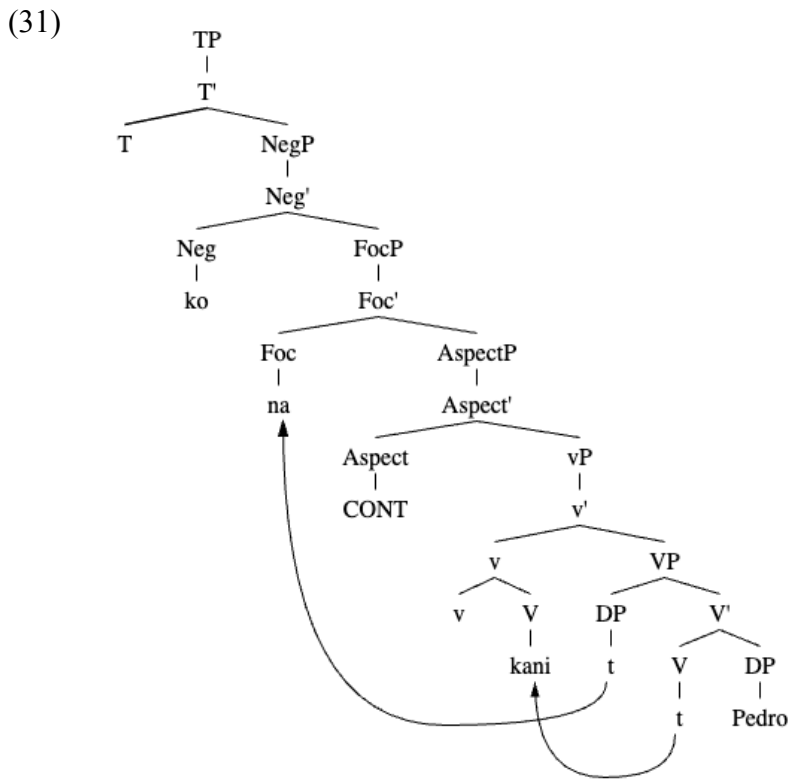
Our syntactic structure must also make an effort to account for the negation facts of Mixtec. As discussed earlier, Mixtec forms negation through a combination of *ko* and a rising tone. *Ko* is obligatory in the continuative aspect, optional in the completive, and absent in the potential. We will account for these facts by proposing

two levels of negation: a higher tonal negation which is obligatory in all negative contexts, and an optional lower negation containing *ko* when applicable. Splitting the two allows us to generalize the tonal facts of negation and separate them from the idiosyncratic behavior of *ko*. Negative tone maps onto the element following it. When that element is *ko*, *ko* receives the tone[4]. When that element is a DP, the DP receives the tone[5] (this will be discussed in section 4). And when that element is a verb, the verb receives the tone.

Ko receives the negative tone

(32) **kǒ** =nā   kāmà   kánì   pedro

NEG =INDF+hum   quickly   hit.CONT   pedro

"Pedro is not hitting anyone quickly"

DP receives the negative tone

(33) **nà**   **ní-ʃã̄ʃì**   ⁿt͡síkā

NEG=INDF+hum   NEG-eat.COMP   banana

"Somebody didn't eat a banana"

---

[4] In this example, *ko* bears the full tone due to its context - however, *ko* tends to spread this tone onto the following elements.

[5] As with *ko*, we see a rising tone spread across two syllables. Rather than a rising tone on ni, as seen in (37), we see a low tone on *na* followed by a high tone on *ni*. The low tone followed by the high tone can be seen as an instance of a rising tone. This is similar to the process shown in (21), in which a rising tone on *ko* is spread between *ko* and the first syllable of the verb.

The V receives the negative tone (either on the completive morpheme *ni* or on the verb itself)

(34) **nǐ**–tsʲaˀjĭ   (Eischens to appear)

  NEG-rot.COMP

  "Didn't rot"

(35) **tsʲǎˀjĭ**   (Eischens to appear)

  NEG-rot.COMP

  "Will not rot"

The trees below show the three situations: in each of them, LH maps onto the tone following it - further tone spreading may occur, but it does not affect this step.
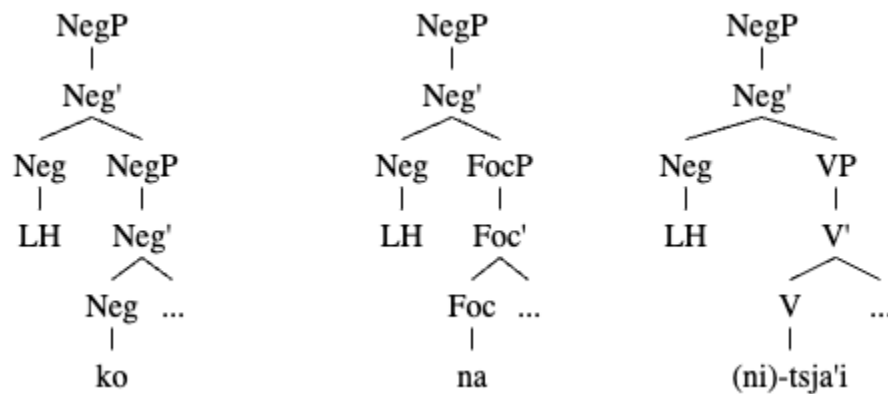
Figure 4: Three trees showing the location of the LH negation tone

3 Restrictions on Bare Indefinites

Negative sentences can help highlight the restriction on *na*. In negative

sentences, larger indefinites such as *ni-in-na* can raise to two pre-verbal positions:

before *ko* (37), and after *ko* (36). The situations in which full negative indefinites

raise to the former and latter positions overlap significantly, and there is no clear

semantic difference between the two sets - both positions appear to be equally valid

options for movement, as is remaining in situ. It is worth noting as well that when

negative indefinites raise before *ko*, they move outside the scope of negation, but the

meaning of the sentence does not change.

(36) kô **nì-í̋=nà** kānì pēdrò

NEG not-one=INDF+hum hit.CONT Pedro

"Pedro is not hitting anyone"

(37) **nī-í̋=nà** kò kánī pέdrō

not-one=INDF+hum NEG hit.CONT Pedro

"Pedro is not hitting anyone"

In contrast, *na* is restricted - it can raise (and are thus not restricted to in-situ

positions), but it cannot raise ahead of *ko*.

(38) kǒ=**nā** kánī =nâ

NEG=INDF+hum hit.CONT =INDF+hum

"Noone is hitting anyone"

22

(39) *=**nà**   kǒ   nà   kānì

       =INDF+hum   NEG   =INDF+hum   hit.CONT

     Intended: "No one is hitting anyone"

     The facts for *ña'a* are different. While we do see some examples in which

ña'a is ungrammatical clause-initially in negative sentences, in positive sentences,

this restriction does not hold (42). This is the case for singular and plural readings of

ña'a.

(40)   ko   **ñā̰̄ʔā̰**   káni   =nà

       NEG   INDF-hum   hit.CONT   =INDF+hum

       "No one is hitting anything."

(41)   *****ñā̰̄ʔā̰**   ko   káni   =nà

       INDEF-hum   NEG   hit.CONT   =INDF+hum

       Intended: "No one is hitting anything."

(42)   **ñā̰̄ʔā̰**   kani   Pedro

       thing   hit.CONT   Pedro

       "Pedro is hitting a thing"

(43)   **ñā̰̄ʔā̰**   kani   Pedro

       thing   hit.CONT   Pedro

       "Pedro is hitting things"

In contrast, *na* is even further restricted in positive sentences, and cannot occur pre-verbally at all. This contrasts with larger indefinites like *in-na*;

(44) kánì   pédɾò =**nà**   In situ

hit.CONT   Pedro =INDF+hum

"Pedro is hitting someone"

(45) *=**nà**   kānì   pēdɾò   Raised

=INDF+hum   hit.CONT   Pedro

Intended: "Pedro is hitting someone"

(46) kānī   **í**   =**nā**   pēdɾò   In situ

hit.COMP   one   =INDF+hum   Pedro

"Someone hit Pedro"

(47) **í**   =**nà**   kánì   pēdɾò   Raised

one   =INDF+hum   hit.CONT Pedro

"Someone is hitting Pedro"

Na can appear before verbs in negative sentences, so we cannot quantify this restriction in terms of their position relative to the verb. We can, however, note that both pre-*ko* positions in negative sentences and pre-verb positions in positive sentences are sentence-initial. Sentence-initiality (which we will specify is technically clause-initiality in the following section) will be the feature that makes *na* ungrammatical in these raised positions. In contrast, while we see *ña'a* restricted in

some clause-initial positions, clause-initiality is not enough to determine whether

ña'a is grammatical.

3.1 Evaluating a bare noun analysis using distributional patterns

     Na as an indefinite means the indefinite 'someone,' but it can also mean the

third person plural pronoun 'them'. (Quirk 1985) Similarly, *ña'a* means the indefinite

'something," but it can also mean the noun 'thing' with singular or plural meaning.

Thus there is the possibility that these elements are bare nouns which are number

neutral (with the possibility of singular and plural meanings). Bare nouns are

characterized by the lack of an overt determiner (such as *in* in this case) and the

ability to have definite and indefinite readings. (Dayal & Sa 2020). They have also

been shown to be number-neutral in other languages such as Turkish, with bare

singular nouns (Sag 2021). We will explore this here.

     The definite form of *na* has the same distribution as the indefinite form. It can

appear in situ, and in raised situations, but (like the indefinite) it cannot appear

clause-initially in positive or negative sentences.

(48)   kani   Pedro **=nà**   In situ (POS)

        hit.CONT   Pedro =3pl

        "Pedro is hitting them"

(49) *=**nà**   kani   Pedro   Raised (POS)

      =3pl   hit.CONT   Pedro

      Intended: "Pedro is hitting them"

(50) ko   kani   Pedro =**nà**   In situ (NEG)

      NEG   hit.CONT   Pedro =3pl

      "Pedro is not hitting them"

(51) ko   =**nà** kani   Pedro   Raised post-*ko*

      NEG   =3pl   hit.CONT   Pedro

      "Pedro is not hitting them"

(52) *****nà**   ko   kani   Pedro   Raised pre-*ko*

      3pl   NEG   hit.CONT   Pedro

      Intended: "Pedro is not hitting them"

Ña'a can also be interpreted as number-neutral, although so far only has indefinite readings (namely "something," "a thing," and "things"). However, we don't want to exclude the possibility that it can have definite meanings as well ("the thing" and "the things," e.g.). What is notable about it is that both singular and plural readings break the apparent restriction on clause-initiality. While there are restrictions on clause-initial positions, these cannot be described with the same rule as *na*. While it may be appealing to try and separate the "thing" and "things" meanings from the "something" meaning for *ña'a*, we currently do not have a metric with which to do

26

this. All current interpretations of *ña'a* are indefinite, and "something" can have singular or plural number attached to it, so there is currently no meaningful distinction we can draw.

(53) kani   Pedro ñã̄ˀã   In situ (POS)

  hit.CONT   Pedro thing

  "Pedro is hitting things"

(54) ko   ñã̄ˀã   kani   Pedro   Raised post-*ko*

  NEG   thing   hit.CONT   Pedro

  "Pedro is not hitting things"

(55) *ñã̄ˀã ko   kani   Pedro   Raised pre-*ko*

   thing NEG   hit.CONT   Pedro

   Intended: "Pedro is not hitting things"

(56) ñã̄ˀã   kani   Pedro   Raised (POS)

  thing   hit.CONT   Pedro

  "Pedro is hitting things"

So where does this leave us? We do see that *na* displays the exact same distribution whether it is indefinite or definite, and a singular or plural meaning seems possible for each point in the distribution. As a result, we have evidence that *na* may be a number-neutral bare noun. *Ña'a* can also be classified as a number-neutral bare noun, given how closely its different meanings pattern together, but it does differ

significantly from *ko*. First, it does not seem to have a pronoun attached to it, but rather manifests as a noun in many cases. Second, it can appear clause-initially. So while the identity of *na* and *ña'a* turns out to be quite similar, their facts nonetheless diverge here.

3.2 A ban on sentence-initiality is a ban on clause-initiality

The majority of the sentences we have been exploring up until this point have been simple sentences, and as such when an indefinite is clause-initial, it is sentence-initial and utterance-initial as well. However, *na* cannot appear in clause-initial positions in situations where it is not sentence- or utterance-initial. In (57) and (58), while *na* appears after other material, this is not enough to render the sentence grammatically correct.

(57) *kúntâʔīnī   māríâ [=**nà**   kō   kánì   pēdrò]

       know.3sg.CONT   Maria =INDF+hum NEG   hit.CONT   Pedro

       Intended: "Maria knows that Pedro is not hitting anyone."

(58) *íjò=nā   [=**nā**   kānì]

       is=INDF+hum   =INDF+hum hit.CONT

       Intended: "Someone is hitting someone"

3.3 An Exception: *Na* cannot follow raised DPs

Beyond its inability to appear initially in sentences, *na* may also not follow other raised DPs (63). This is unexpected, as up until this point any pronounced material, no matter its grammatical purpose, has been able to serve as a host for this element.

(59)   *Pedro   **=nà**   kani

         Pedro   =INDF+hum   hit.CONT

         Intended: "Pedro is not hitting anyone"

In addition, other raised DPs can precede a bare *na* when there is at least one element that is not raised. In (60), *ko* allows *na* to appear after it, independent of the raised DP that precedes both words.

(60)   nī-ín=nà   ko   nà   kánì

         not-one=INDF+hum   NEG   INDF +hum   hit.CONT

         "No one is hitting anyone."

We should note here that *ña'a* also displays some level of restriction in situations like this, though it is uncertain whether there are other examples which work better. Moreover, this does not negate what we have seen clause-initially in positive sentences.

(61)   *nī-í=na   **ñā'ằ**   kò   káni-nà

         not-one=INDF+hum   INDF-hum   NEG   hit.CONT

         Intended: "Noone is hitting anything""

29

3.4 An Exception: *Na* can appear clause-initially in certain circumstances

To wrap up our survey of the behavior of *na* we will touch on two situations which seem to violate its restriction on clause-initiality.

3.4.1 Sentences that utilize tonal sentential negation

The majority of negative example sentences in this paper have been in continuative aspect which consistently marks sentential negation with the particle *ko*. However, negation is represented only as a tone in potential aspect, as well as with a combination of tone and the *ni* morpheme in completive aspects. In contexts associated with tonal negation, *na* appears to raise to a sentence-initial position.

(62) nà   ní-ʃãʃì   ⁿtsíkā

NEG.INDF+hum   eat.COMP   banana

"Somebody didn't eat a banana"

3.4.2. Interrogative sentences

Negative sentences are not the only location in which we see an exception to the ban on sentence-initial *na*- questions, too, show this property. Questions in Mixtec also involve raising, but the specifics are different. Wh-questions in Mixtec, like negation, can be expressed with both a particle and tone. Particle wh-words like *jo* have a few key properties.

- The wh-word appears sentence-initially.

- When the wh-word appears with *na*, *na* may only appear directly afterwards. This movement is obligatory, unlike raising to focus.[6]

Example (63) demonstrates a grammatical wh-question in SMPM. We will posit that the obligatory movement of *na* to the CP is independent of its status as a clitic - however, since there is pronounced material to its left, *na* may appear in this position. (64) demonstrates an ungrammatical sentence in which *na* does not move to a CP position.

(63)　jó　　=nà　　　　　kánī　　　　pēdɾò

　　　　Q　　INDF+hum　　hit.CONT　　Pedro

　　　　"Who is hitting Pedro?"

(64)　*jó　　　　　kānì　　　=nā　　　pēdɾò

　　　　 Q　　　　　hit.CONT　=INDF+hum　Pedro

　　　　Intended: "Who is hitting Pedro?"

Similarly, *na* cannot be separated from *jo* by sentential negation - it must raise to a position directly after the question morpheme, and cannot be separated by other elements like *ko*.

(65)　jó　　=nà　　　　　kō　　kánī　　　　pēdɾò

　　　　Q　　=INDF+hum　　NEG　hit.CONT　　Pedro

　　　　"Who is hitting Pedro?"

---

[6] jo also appears in yes/no questions, leading to an analysis in which it is a broad question morpheme, rather than an analogue of 'who'. We will not go into too much detail about its precise meaning, but simply note that it can be used as both a Q particle and a human wh-word. It does not appear with questions asking 'what'.

(i)　　jó kánì　　pēdɾò
　　　　Q hit.CONT　Pedro
　　　　"Is Pedro hitting someone?"

(66)' *jó kō =nà   kánì   pēdrò

  Q   NEG =INDF+hum   hit.CONT   Pedro

  Intended: "Who is not hitting Pedro?"

Na may also not rise above *jo* to form a question. The sentence below violates
jo's requirement for *na* to come immediately after it, but it also violates *na*'s
restriction on appearing clause-initially

(67) *=nà   jo   nómi   pedro

  =INDF+hum   Q   hug.CONT   Pedro?

  Intended: "Who is hitting Pedro?"

  (68) and (69) demonstrate that, while *na* cannot raise to a left-periphery
position outside the question, other DPs can - again we see restrictions on *na*, similar
to the ones with *ko*.

(68) *nà   jo   =nà   káni

  INDF+hum   Q   =INDF+hum hit.CONT

  Intended: "Who is hitting someone?"

(69) ni   ñāˀằ   jo   =nà   ko   káni

  not INDF-hum   Q   =INDF+hum NEG hit.CONT

  "Who is not hitting anything?"

However, despite *jo*'s tight restrictions on where *na* can occur in its presence, it does not always occur in questions. *Jo* can be absent, leading to *na* occurring sentence-initially, with question-hood occurring as a tone.

In (70), we see a distinct rising tone on *na* which cannot arise from negation, as this rising tone appears on *ko*.

(70)   nâ   ko ní-ʃãʃì   ⁿts͡íkā

        Q.INDF+hum   eat.COMP   banana

        "Who didn't eat a banana"

Ña'a can also appear sentence-initially in questions, though this may be a consequence of its overall ability to appear clause-initially, rather than an exception to the rule in the way (70) is for *na*.

(71)   ñàʔã̀   kánì   pɛdɾò

        Q.INDF-hum hit.CONT   Pedro

        "What is Pedro hitting?"

We once again see an exception made in certain cases to allow bare *na* to appear sentence-initially, when its natural tone is superseded by a tone expressing interrogativity.

4 Analyzing *na*: is it a clitic?

4.1 Summarizing the facts

Before we get into our analysis of indefinites in Mixtec, which will seek to provide an explanatory theory for the data and posit the deeper structures and general processes involved, it is useful to compile and highlight the generalizations we can make from the data itself.

First is the general behavior of *na* and *ña'ã*, two elements which can function as indefinites and which overall display a similar pattern against clause-initiality, though the latter can indeed be clause-initial, breaking this rule. In other positions, including raised positions (such as the position following *ko*) both elements are unrestricted. This clause-initial position may also be sentence-initial, but it does not have to be. *Na* contrasts with other DPs, including other indefinites, which can occur clause-initially.

Neither *na* nor *ña'a* has a unique independent syntactic movement - they may raise to early positions in the sentence, but none of these movements are specific to *na* and *ña'ã*. The data have shown two processes that can move these indefinites: raising to focus and wh-raising. Both of these movements can occur with other DPs, such as names and more complex indefinites, neither of which have a restriction against clause-initiality. So to summarize, movement and positional restrictions on *na* and *ña'ã* are not related. (72) moves a larger indefinite, and (73) moves a name.

34

(72)   jo   **ni-in**   **=na**   kani   Pedro   wh-raising

      Q   not-one   =INDF+hum   hit.CONT   Pedro

      "Who is not hitting Pedro?"

(73)   **Pedro** ko   **ña yibi**   káni   focus raising

      Pedro   NEG   3f person   hit.CONT

      "Pedro is not hitting anyone"

While this fact is important for *na*, it is also what we would expect from *ña'a* if we are to call it a bare noun, as we would not expect a difference in movement between nouns and other DPs.

While the rule for *na* detailed above provides a general summary of the facts, it nevertheless comes with two exceptions. First, a potential analysis must be able to account for instances in which the it *does* occur clause-initially. This exception is seen both in cases of raising to focus and in wh-movement. When we raise *na* in a negative sentence in which negation is expressed via tone (such as in the completive when *ko* is not present), a rising tone seems to map onto the raised *na*, and it is able to occur clause-initially, without a change in meaning.

(74)   nà   ní-ʃãʃì $^{n}\widehat{ts}$íkā

      NEG.INDF+hum   eat.COMP   banana

      "Somebody didn't eat a banana"

Likewise, when *na* is raised to wh-positions in questions where interrogation is expressed through a falling tone, it may occur at the start of the clause.

(75) nâ   ko   ní-ʃãʃì   ⁿtsíkā

Q=INDF+hum   NEG   NEG-eat.COMP   banana

"Who didn't eat a banana"

It is worth mentioning that when (74) was presented without the falling tone on *na*, it was labeled ungrammatical (Figure 5, shown below) . This is because *ko* is present, so it receives the negative tonal features, rather than *na*. There is no position that *na* can raise to in negative sentences in which it receives tonal information over *ko*, and since there are no other floating tones available, it may not start the clause.

(76) *=nà   kǒ   ni-ʃãʃì   ntsika

=INDF+hum   NEG   NEG-eat.COMP   banana

 Intended: "No one ate a banana"

The syntax trees below show the situation in (78) and the situation in (80) respectively. In the first, *ko* is present, and *na*, occurring before it, must appear in some higher clause-initial position - one which is outside the scope of the tonal NegP – *na* does not receive a tone and the structure is disallowed. However, when *ko* is not present, *na* may appear in a clause-initial position, one which *is* within the scope of the tonal NegP, and which the floating tone can attach to. Note that the tonal NegP is obligatory, while the lower *ko* NegP is optional. Assume that the morpheme *ni* is located within the verb phrase. While the inability for phrases to raise between the

two NegPs is important for our analysis, the exact mechanisms which cause the restriction are outside the scope of this paper.
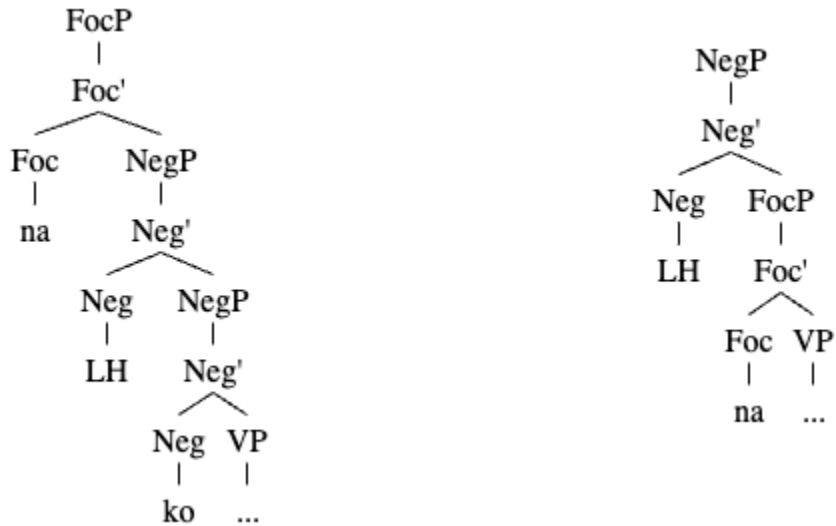

Figure 5: A tree showing a situation in which na *cannot receive tone, followed by one in which it can receive tone.*

So once again, tone allows us to circumvent the rule against clause-initiality, and as we will discuss in section 4, this appears to be the result of a second (floating) tone appearing on elements which otherwise only bear a single tone. We can also note that, despite the presence of two NegP's, there is no space between them to which a phrase can raise.

In the second exception, there are also instances where the *na cannot* appear in positions that are not clause-initial. These positions overall involve *na* in a raised position following another DP.

(77) *Pedro   **=nà**   kani

      Pedro   =INDF+hum hit.CONT

      Intended: "Pedro is not hitting anyone"

In contrast, multiple raised DPs can co-occur freely at the start of the clause without restrictions.

(78) in   =nà   in   na'a   kani

      one   =INDF+hum   one   INDF-hum   hit.CONT

      "Someone is hitting something"

(79) nī-ī́   =nà   ní   ñā́ʔā̀   kò   káni-nà

      not-one =INDF+hum   not   INDF-hum   NEG hit.CONT

      "Noone is hitting anything"

4.2 Defining *na*

     We must now turn our focus to the identity of *na*, namely whether it is a full word, a clitic, or something else. A clitic-based analysis is appealing, as we can characterize the absence of *na* and *ña'a* clause-initially as a consequence of the lack of a host (Spencer 2012). If it attaches to a host on its left, this position is unsatisfactory in clause-initial positions, as there is nothing in this position. This is in contrast with other locations, when it follows another element such as a verb, a DP, or

38

negation. As a result, we will need to test whether *na* and *ña'a* display other predicted properties of clitics. For our purposes, we will use Zwicky's 1977 typology for clitics, which differentiates between simple clitics and special clitics. Simple clitics have similar distributions to overall words, while special clitics have specific syntactic processes dedicated to them (Zwicky 1977). The behavior of *na* aligns it most closely with simple clitics, as both elements can appear in situ and only raise via processes which target all DPs, such as focusing or wh-raising. In addition, any restrictions on the distribution of *na* can be explained in terms of the phonological facts of their position - namely, whether there is a pronounced element to the left.

To determine whether *na* is indeed a clitic, we will have to look at it in other environments. So far it has been restricted in clause-initial positions, but we must also analyze its behavior in other places where clitics would be expected to be disallowed. Cardinaletti and Starke's 1999 analysis of clitics notes that they are disallowed in isolation in fragment answers and in coordinated contexts. Indeed, both of these facts follow for *na*.

(80) *jo-nà   kānì   Maria?   =nà!

       Q=INDF+hum   hit-CONT   Maria?   =INDF+hum

       Intended: "Who is Maria hitting? Someone!"

(81) *káni   pedro   ñā̧ʔā̧   ɾa   =nà

       hit.CONT pedro   INDF-hum   and   =INDF+hum

       Intended: "Pedro hit something and someone"

As a result, it is reasonable to define *na* as a clitic - not only does it seem to require a host when it is clause-initial, it is also prohibited in other positions where clitics are disallowed. However, in keeping with its ability to break from the behavior of *na*, *ña'a* is once again allowed in positions where *na* is not, and indeed where clitics are not.[7] *Ña'a*'s ability to occur in fragment answers supports the idea that it can occur clause-initially without being ungrammatical.

(82) nǎ   kānì   Maria?   Ña'ắ!

      Q=INDF-hum   hit.CONT   Maria   INDF-hum

      Intended: "What is Maria hitting? Something"

(83) káni   pedro   na yibi   ɾa   ña'a

      hit.CONT   pedro   3sg person   and   INDF-hum

      "Pedro hit someone and something"

Ostrove (2023)'s analysis notes that *na,* when it functions as a third-person pronoun, is a clitic. The data allows us to predict that definite *na* will be absent in

---

[7] The same facts are present for other meanings of *ña'a*, such as "things". However, it seems that for now "thing" may not be an available interpretation, and may only be expressed as *in-na*. This may turn out to be a non-issue, however, as in other clause-initial positions, there is no distinction based on number, and "something," which is still available as an interpretation, can have a singular meaning.

(i)   nǎ   kānì   Maria? Ña'ắ!
        Q.INDF-hum   hit.CONT   Maria   thing
        Intended: "What is Maria hitting? Things"

(ii)   kani   pedro na yibi   ɾa   ña'a
        hit.CONT pedro 3sg person   and   thing
        "Pedro hit someone and things"

40

isolation and in coordinate forms. Both of these patterns mirror what we see with indefinite *na*.

In sum, the data does not point to any situations where the definite and indefinite forms of *na* are expected to differ, so our proposal that *na* is a number-neutral bare noun remains unchallenged.[8]

So where does this leave us? Our clitic analysis is indeed borne out for *na*, and thus we have a way of explaining why it may not occur at the start of a clause. *Ña'a* on the other hand, continues to break the rule against occurring clause-initially, and does not behave as a clitic. While proposing requirements that *na* has on the position it moves to, it is more useful to utilize a different method for determining when and where *na* can appear: the requirements of the intonational phrase. Clauses can be mapped to intonational phrases which we can propose have a requirement that they must begin with a full prosodic word, barring *na* from these positions. This is independent of *na*'s requirement that a host appears to its left, but does not preclude this process from happening, and it allows us to more easily explain why elements outside of the clause and other raised DPs are not available for *na* to use.

---

[8] I have additional data in which *na* is disallowed in a context where it seems to mean "them". The sentence was provided as part of a larger discussion on *na* as a third person plural pronoun. However, I do not have explicit confirmation that the meaning of *na* in this instance is "them" even though I expect that this is the interpretation.

(i)   *jo-na   kānì   Maria?   =na!
        Q=INDF+hum   hit-CONT   Maria?   =them(?)
        Intended: "Who is Maria hitting? Them!"

4.3 *Na* can occur clause-initially with tone

We can hypothesize that *na* has a single tone associated with it, while prosodic words in Mixtec have a two-tone minimum. The nouns and verbs we have seen so far in Mixtec typically have two or more syllables, and each syllable can be proposed to carry its own tone. Clause-initial *na* typically does not meet this requirement, but it can meet this requirement through a combination of its lexical tone and the floating tone which expresses negation or forms wh-phrases. We can call these tones floating in these instances because there is no accompanying syllable in these cases to express question or negation. ("floating tone" 2014) The presence of two tones seems to be sufficient to allow *na* to be treated as a full word, even if only one tone gets expressed.

LH   L

na

Figure 6: Negation represented by a rising floating tone on na

HL   L

na

Figure 7: Question represented by a falling floating tone on na

4.4 Clause-initiality

The fact that *na* cannot appear clause-initially, even when material precedes it outside of the clause, is an important topic for our analysis of *na* to address.[9] The facts surrounding *na* and its restrictions are phonological in nature, but the clause boundary is a syntactic one, so we need a way to unite the two.

To do this, we will utilize Match Theory, which maps the structures and borders of syntactic phrases onto prosodic ones (Ishihara 2022). Within Match Theory, CP is mapped onto an Intonational Phrase (often written as ι), a prosodic phrase which often aligns itself with 'root sentences,' which often coincide with CP structures, though they are associated with many other elements which allow for pauses in sentences (Ishihara 2022, Nespor 2007). A typical example of a single intonational phrase corresponding to a CP is shown below.

(84)　　[ι The hamster eats seeds all day]　　　　　　　　　　(Nespor 2007)

As a result, we can map the intonational phrase boundary to clause boundaries seen previously. The effect, as seen below, is that whether or not a clause begins a sentence, the intonational phrase associated with it may not begin with an element smaller than a prosodic word – in other words the element must have at least two tones associated with it. *Na*'s apparent insensitivity to elements outside the clause is thus a consequence of those words' placement - any words outside the clause are

---

[9] A ban on clause-initiality is also seen in Romance clitics, which is known as the Tobler-Mussafia law. (Ingham 2014)

necessarily outside of the intonational phrase, and are thus unavailable to take up that initial position, and cannot satisfy the intonational phrases requirement.

Sentence-initially

(85)   *[ᵢ **nà**   kǒ   =nà   kānì]

        INDF+hum   NEG   =INDF+hum hit.CONT

        Intended: "No one is hitting anyone"

Non-sentence-initially

(86)   *kúntâʔīnī   māríâ   [ᵢ **nà**   kō   kánì   pēdrò]

        know.3sg.CONT Maria   INDF+hum   NEG   hit.CONT   Pedro

        Intended: "Maria knows that Pedro is not hitting anyone."

    Since we are not proposing a ban on *ña'a* clause-initially, we do not need to explain why *ña'a* can occur in isolation - this is an expected outcome when we consider it a non-clitic bare noun.

4.5 Restrictions on non-clause-initial positions

    The use of intonational phrases can also help us explain why *na* cannot follow other raised DPs. *Na* can follow negation, question morphemes, verbs, and in-situ DPs, but it cannot appear after a DP which raises. To explain this, we can propose that, while raised *na* falls inside the boundary, other indefinites which raise to the left periphery may fall outside of this scope. Looking again at the examples of this from

44

section 4.1, we can include the boundary of the intonational phrase to delineate how these sentences are structured. In (87) Pedro is outside of the intonational phrase.

(87)   *Pedro   [ι=na   kani

        Pedro   =INDF+hum   hit.CONT

        Intended: "Pedro is not hitting anyone"

This schema is useful for two reasons. First, the intonational phrase already appears as a boundary between *na* and material in other clauses, so a similar process which leaves *na* blind to material above it would be a good instance to utilize this boundary again. It also avoids the pitfalls that alternative explanations fall into, such as the one in the next paragraph.

We could also explain *na*'s inability to use higher DPs as a result of the order of movement. If we propose that *na* moves, intonational phrases check the nature of their first word, and then additional DPs move, we effectively create an ad hoc timing for the second step. Such an analysis would work as follows:

- Word order before movement

  [FOC____ [*FOC*____ [NEG ko [V kani … ni-in=na … na]

- Indefinite moves for focus purposes (the indefinite is clause-initial here)

  [FOC____ [*FOCna* [NEG ko [V kani… ni-in-na]

- The intonational phrase checks whether it begins with a prosodic word

  - In this case it does not

- Other DPs move

  - [$_{FOC}$ ni-in=na [$_{FOC}$ **na** [$_{NEG}$ ko [$_V$ kani]

Outside of solving this specific problem, there is no reason why intonational phrases should check their starts after one DP moves, but not additional ones. If, alternatively, certain DPs move outside of the intonational phrase, then the intonational phrase can check its start after all syntactic movements have occurred.

5 Conclusion

In this paper we have analyzed a relatively simple restriction on bare indefinites occurring clause-initially, which is made more complex by the differing behaviors of *na* and *ña'a* in other contexts, and by exceptions to the pattern when tone is involved. We have also focused on the identity of *na* and *ña'a*. *Na* is a clitic and can be labeled as a number neutral bare noun with singular indefinite meaning and plural definite meanings. *Ña'a* on the other hand neither behaves as a clitic nor as a bare noun, and shows some ability to appear clause-initially. As a result, we must limit a clause-initial ban to only *na* and work forward from there. Our analysis has also investigated the interaction of *na* with other clitics. The intonational phrase may begin with *ko*, which has previously been labeled as a clitic, so our grammar must have a way for ko to form part of a prosodic word in these cases. Within a clitic analysis, this can be achieved by attaching *ko* to a host verb, or by combining *ko* with a bare indefinite, and forming a prosodic word from two elements which are too small

to do this on their own. Our analysis does open up the possibility that *ko* is not a clitic, and is actually a prosodic word, which is an alternative explanation for why *na*, which has independent evidence of clitic-hood, is able to appear with it.

Our analysis has also made steps to broaden the framework of clitic pronouns. We have shown that the clitic *na* meaning "them" and the clitic *na* meaning "someone" have virtually identical distributions, and thus can be grouped together. This opens up the possibility that other third person pronouns are bare nouns as well, and pattern with their own indefinites. Similarly, we have shown that *ña'a* can be described as a bare noun, opening up possibilities for other bare nouns with indefinite meanings to occur.

This research raises interesting questions which can likely lead to future directions for research. This includes how an apparent two-tone minimum interacts with mora-hood for Mixtec words, and how adding a tone to form a prosodic word may be instantiated using a mora-based framework. In addition, the ability for bare indefinites to appear in positive and negative sentences, while other indefinites are confined to one polarity may lead to discoveries about the interaction of indefinites, negative concord, and clitic-hood. Future research will be necessary to explain the distribution of *ña'a*. It is possible that the position to which *ña'a* raises affects its grammaticality, as we often see it disallowed in pre-*ko* positions, but not in pre-verbal positions in positive sentences, which can be the lower focus position. We may also discover that there are situations where only a single or only a plural *ña'a* can occur, without the indefinite "something" also being present.

47

Bibliography

Anderson, Stephen R. (2005) Aspects of the Theory of Clitics. Oxford University.

Alqassas, Ahmad. "Locus of Negation in Syntactic Structure." A Multi-Locus
    Analysis of Arabic Negation: Micro-Variation in Southern Levantine, Gulf
    and Standard Arabic, vol. 1, Edinburgh University Press, 2019, pp. 17–65.

Bjorkman, Bronwyn and Zeijlstra, Hedde. (2014) Upward Agree is superior. Draft.

Bošković, Željko. (2001). On the nature of the syntax-phonology interface :
    cliticization and related phenomena (1st ed.). Elsevier.

Caponigro, Ivano; Torrence, Harold. (2013) Free Relative Clauses in Two Mixtec
    Languages. International Journal of American Linguistics, Vol. 79 (1).

Cardinaletti, Anna and Starke, Michal. "The typology of structural deficiency: A case
    study of the three classes of pronouns". 5 Clitics in the Languages of Europe:
    Vol 5 (1). De Gruyter Mouton, 1999, pp. 145-234.

Dayal, V., & Sa, Y. mur. (2020). Determiners and Bare Nouns. Annual Review of
    Linguistics, 6(1), 173–194.
    https://doi.org/10.1146/annurev-linguistics-011718-011958

Eischens, Ben. Classifying negated nominals across Mixtec. [Manuscript submitted
    for publication in the International Journal of American linguistics]

Eischens, Ben. (2020) "Decomposing negative indefinites in San Martín Peras
    Mixtec"

Eischens, Ben. (2022). "Tone, Phonation, and the Phonology-Phonetics Interface in
    San Martín Peras Mixtec. UC Santa Cruz. Retrieved from
    https://escholarship.org/uc/item/6sq4z7cm

"floating tone". (2014). In The Concise Oxford Dictionary of Linguistics (3rd ed.).
    Oxford University Press.

Hedding, Andrew A. (2021) Asymmetric pitch raising under focus in SMPM.
    Proceedings of the Workshop on the Structure and Constituency of Languages
    of the Americas 24.*University of British Columbia Working Papers in
    Linguistics*.

Hedding, Andrew A. (2022) How to move a focus: the syntax of alternative particles.
    Dissertation, University of California, Santa Cruz.

Ingham, Richard. (2014) Old French negation, the Tobler/Mussafia law, and V2. Lingua, 147, 25–39.

Ishihara, Shinichiro. (2022). Match Theory: An overview. Language and Linguistics Compass: Vol 16 (1)

Israel, M. (2011). The grammar of polarity : pragmatics, sensitivity, and the logic of scales. Cambridge University Press.

Nespor, Marina, and Irene Vogel. Prosodic Phonology : With a New Foreword. Mouton de Gruyter, 2007, https://doi.org/10.1515/9783110977790.

Ostrove, Jason (2023) Obligatorily Overt PRO in San Martín Peras Mixtec. Linguistic Inquiry. Advance publication.

"Oto-Manguean Language Family" Native Languages of the Americas. Native-languages.org

Sağ, Yagmur. (2022). Bare singulars and singularity in Turkish. Linguistics and Philosophy, 45(4), 741–793. https://doi.org/10.1007/s10988-021-09323-0

Spencer, Andrew. (2012) Clitics: An Introduction.

Quirk, Randolph (1985) *A Comprehensive grammar of the English language*. Longman.

Verb First : On the Syntax Verb Initial Languages. (2005) Edited by Andrew Carnie, et al., John Benjamins Publishing Company. ProQuest Ebook Central,

https://ebookcentral-proquest-com.oca.ucsc.edu/lib/ucsc/detail.action?docID=623268. Zwicky, Arnold M. (1977) On clitics. Ohio State University. Reproduced by the Indiana University Lingistics Club.