

Chapter 6

Constituency in Tù'un Ntá'ví (Mixtec) of San Martín Duraznos

Sandra Auderset^a, Carmen Hernández Martínez^b & Albert Ventayol-Boada^b

^aUniversity of Bern ^bUniversity of California, Santa Barbara

In this chapter we report the results of 27 constituency diagnostics applied to verbal predicate constructions in San Martín Duraznos Mixtec. We show that there are remarkably few convergences between diagnostics. We also discuss issues we encountered in establishing the verbal planar structure as they relate to competing analyses of morphemes.

1 Introduction

In this chapter we provide the first description of constituency in verbal predicate constructions in San Martín Duraznos (SMD) Mixtec. We follow the methodology laid out in Tallman (2020, 2021).

Constituency in Mixtec languages has previously been discussed by Macaulay (1993, 1996) for Chalcatongo Mixtec. Specifically, Macaulay focuses on describing the ordering of constituents in the language and offers a template for the positions of arguments, topic and focus constituents, and phrasal clitics. In fact, discussions on clitics in Chalcatongo Mixtec and the closely related variety of San Miguel El Grande have featured prominently in the literature on the morphology-syntax division (cf. Pike 1944, Pike 1945, Macaulay 1987, among others).

Pike (1945) argues that there is no global morphology-syntax distinction in San Miguel El Grande Mixtec. This claim is based on the observation that many bound forms can be synchronically analyzed as phonological reductions of full words. Furthermore, he notes that there is distributional overlap between bound



forms that are not reductions and bound forms in general, so that all bound forms could be analyzed as underlyingly derived from full words. Macaulay (1987) argues against such an analysis, claiming that it misses important distributional, semantic, and phonological differences between morphemes and syntactic constructions attested in Chalcatongo Mixtec. Rather, she posits that a distinction between affixes, clitics, and words is motivated and that clitics can be classified into the two types proposed by Zwicky (1977): ‘simple clitics’ and ‘special clitics/phrasal affixes’. A separate study on ‘clitics’ in SMD, however, found no support for this classification, but rather showed that there are more classes of morphemes and constructions (Auderset et al. 2021).

As for SMD Mixtec, there is no earlier descriptive work other than word lists collected by Josserand (1983) and Padgett (2017). This chapter is based on our ongoing collaborative documentation project and, thus, represents what we currently know about the language.

1.1 The language and its speakers

SMD is the Tù'un Ntá'ví variety spoken in the community of San Martín Durazos in Oaxaca (Mexico) and various diaspora communities located in the US, mainly along California's Central Coast. The Tù'un Ntá'ví (or Tù'un Sàvi) languages are part of the Mixtecan branch in the Otomanguean language family (Longacre 1957, Kaufman 1988). Across Mixtec, there is a high degree of diversification, and there is no agreement on how many varieties there are and where the boundaries among them lie (Josserand 1983, Campbell 2017). They tend to form dialect continua across the vast area they occupy, which covers most of western Oaxaca, parts of eastern Guerrero and some neighboring areas in Puebla. Varieties are often divided into three geographic areas: Mixteca Alta, Mixteca Baja, and Mixteca de la Costa. However, these do not reflect linguistic groupings. Linguistically, the most comprehensive study that analyzes variation across Mixtec was carried out by Josserand (1983). She surveyed 188 lexical items from 120 villages where Mixtec is spoken and, based on their phoneme inventories and isoglosses of sound changes, she proposed 12 major dialectal clusters. SMD belongs to the Southern Baja subgroup in her proposal.

The analysis presented here is based on approximately seven hours of naturalistic speech, along with many elicited sentences and native speaker judgements by one of the co-authors. Most of the naturalistic speech was recorded in San Martín Durazos, but some recordings were made in Ventura County, California, where a sizeable diaspora community has settled. The data will be archived with

ELAR (Auderset & Hernández Martínez 2022). The primary contact language is Spanish, although English is also used among speakers in the diaspora.

In what follows, we first present the verbal planar structure in §2 and elaborate on some difficulties and unresolved issues. We include a brief overview of important grammatical features of the language. We then discuss each diagnostic in turn, by providing a definition, justifying fractures, and presenting the domains identified with illustrative examples. We start with phonological domains in §3, then discuss indeterminate domains in §4, and finally address morphosyntactic diagnostics in §5. We summarize our findings and discuss their implications in §6.

2 The planar structure of the verbal complex

Table 3 presents the verbal planar structure of SMD. This is a maximally flat representation of all the elements that can occur in a clause with a verbal predicate. Note that the internal structure of other types of phrases, such as noun phrases (NPs) or prepositional phrases (PPs), is not represented.

Before discussing some problematic cases we encountered in establishing the planar structure, we introduce a few core grammatical elements of the verbal predicate clause and provide some background on the practical orthography.

2.1 Relevant grammatical features and background on the orthography

All examples in this chapter are provided in the practical orthography developed with the community. The orthography is largely phonemic and makes use of digraphs and trigraphs, with diacritics reserved for tone. SMD has a split into post-alveolar and alveolo-palatal consonants, so far unattested in other Mixtec varieties. This means that there are two series of fricatives and affricates: <sh, ch, nch> = [ʃ, tʃ, ⁿdʒ], but <x, tx, ntx> = [ç, tç, ⁿdʒ̥]. The glottal stop is represented as an apostrophe or saltillo. Nasalization of vowels is indicated by an <n> following the nasalized vowel. There are no final consonants and nasalization is contrastive only on final vowels, so <an#> is always [ã]. Long vowels are represented by doubling the vowel. There are three tonemes: high, which is marked with an acute accent; low, which is marked with a grave accent; and mid, which is unmarked. Every vowel is marked for tone (i.e. we do not posit toneless elements). Finally, in the practical orthography we use hyphens to visually separate certain bound elements like pronouns, as in *ve'-un* [house-2SG.NHON] ‘your (SG) house’. These hyphens do not indicate the type of morphological boundary.

SMD, like other Mixtec languages, is verb initial – that is, in a basic declarative clause the verb comes first, followed by the actor argument and then the under-goer (VS and VAO). It is obligatory for the S/A-argument to be present, either as an NP or a pronoun, to form a complete declarative clause, unless the verb is impersonal. SMD has two series of pronouns, which we will refer to as dependent and independent. Dependent pronouns are all mono-moraic and cannot appear as free forms, hence the term ‘dependent’. For first and second persons, they are restricted to S/A arguments, while no such restriction exists for third persons. Independent monomorphemic pronouns only exist for some first and second persons and are all bimoraic. The other independent pronouns, including all of the third persons, are combinations of the topicalizer *míi* and the corresponding dependent pronoun (cf. Table 1 covering first and second person). After the verb, independent pronouns can only be used as P arguments, although they can appear preverbally in focus position or as emphatic pronouns representing any grammatical role.

Table 1: First and second person pronouns

Gloss	Dependent	Independent
1SG	<i>i</i>	<i>yi'i</i>
1PL.INCL	<i>ò</i>	<i>míi-ó</i>
1PL.EXCL	<i>nti</i>	<i>nti'i</i>
2SG.NHON	<i>un</i>	<i>yò'ò</i>
2SG.HON	<i>ní</i>	<i>míi-ní</i>
2PL	<i>ntò</i>	<i>ntó'ó</i>

Verbs are obligatorily marked for aspect-mood, either with a tonal marker, a segmental marker, or a combination of both. Otomanguean languages are famous for their intricate systems of verbal inflectional classes including complex interactions of segmental and tonal marking. We will briefly outline the most important points here, since the SMD inflectional class system has not been previously described (apart from an overview provided in Auderset & Hernández Martínez 2019). SMD exhibits a somewhat simpler system than that of other Otomanguean languages, such as Chichimec (Palancar & Avelino 2019) and Cuicatec (Feist & Palancar 2016). Nevertheless, tonal inflection plays an important role in the verbal system and there are multiple inflectional classes.

Whereas the completive form is always marked with a preverbal element *i* or *ni*, the incompletive and potential forms are often only marked by tone, the

former showing a characteristic high tone on the first mora. We have identified 9 segmental and 7 tonal patterns, but not all combinations of segmental and tonal patterns are attested. The roughly 380 verbal paradigms analyzed so far fall into 28 classes.

Table 2: Verbs (position 17) showing tonal inflection of different inflectional classes

INCMPL	CMPL	POT	Gloss	Class
<i>káchí</i>	<i>ikachi</i>	<i>kachi</i>	'say'	INCMPL high
<i>xú'ní</i>	<i>ixu'ní</i>	<i>ku'ní</i>	'squeeze'	INCMPL high with stem alternation
<i>íin</i>	<i>ixòo</i>	<i>koo</i>	'live, stay'	INCMPL high with suppletion
<i>núná</i>	<i>inúnà</i>	<i>núnà</i>	'be open'	INCMPL with both morae high
<i>kai</i>	<i>ikài</i>	<i>kài</i>	'burn'	INCMPL mid
<i>kuà'àn</i>	<i>ixà'àn</i>	<i>kù'ùn</i>	'go'	segmental alternation only

We now discuss problematic cases that arose in establishing the planar structure: the additive *va*, adverbials, and the form *tiki* ~ *ti*. We then briefly reference the other positions, starting with positions before the verb core and then after. We also comment on how cognate forms are classified in descriptions of other Mixtec varieties, and on how these positions are represented in the practical orthography.

2.2 Issues in establishing the planar structure

The first issue in establishing the planar structure concerns the element *va*, which is glossed here as 'additive'.¹ It is very frequent in naturalistic speech and can appear multiple times in a clause, cf. (1b) and (1c). With verbs, it seems to indicate that the action has happened before or is a consequence of what was done before, as in (1a). With nouns, it appears as a linker in listings and otherwise indicates that there is more of something cf. (1c). With other adverbials, it also seems to mean 'more' e.g., in (1b). This element is a bound form – in other words, it can never appear by itself and is phonologically left-leaning. The difficulty in analyzing this element lies in assessing what it modifies in any given position it can appear in. This is especially pertinent when *va* appears after an adverbial, as it is often unclear whether *va* in these cases modifies the verb core or the adverb

¹In San Martín Peras Mixtec, the label given to the cognate element is 'sequential' and that might be just as appropriate for SMD.

Table 3: Verbal planar structure of SMD

Pos.	Type	Elements	Forms
(1)	slot	connectors, question marker	<i>an, ta, távà, chii</i> , etc.
(2)	slot	question words	<i>nishi, ntxáa</i> , etc.
(3)	slot	focus (S/A/P, OBL, etc.)	
(4)	slot	realis negation	<i>kòó</i>
(5)	zone	adverbials	<i>xàà, sa'a, và'a, vitxi</i> , etc.
(6)	slot	additive	<i>va</i>
(7)	slot	intensifier	<i>kuà'à, tóntó</i>
(8)	slot	intensifier; again	<i>ntxìvà'a, yáá; tiki ~ ti</i>
(9)	slot	intensifier; again	<i>tiki ~ ti; ntxìvà'a, yáá</i>
(10)	slot	additive	<i>va</i>
(11)	slot	modals	<i>nì, ná</i>
(12)	slot	completive; potential negation	<i>i; u/o ~ i</i>
(13)	slot	causative 'do'	<i>sá</i>
(14)	slot	POT; CMPL class markers	<i>ku; xì</i>
(15)	slot	iterative	<i>nta ~ nti</i>
(16)	slot	transitivizer 'put'	<i>chi</i>
(17)	slot	verb core	
(18)	slot	additive	<i>va</i>
(19)	slot	reciprocal	<i>ta'an</i>
(20)	slot	temporal adv.	<i>tuun, kû</i>
(21)	slot	intensifier	<i>tóntó, kuà'à</i>
(22)	slot	intensifier; again	<i>ntxìvà'a, yáá; tiki ~ ti</i>
(23)	slot	intensifier; again	<i>tiki ~ ti; ntxìvà'a, yáá</i>
(24)	slot	additive	<i>va</i>
(25)	slot	inside/being	<i>ini</i>
(26)	slot	S/A	
(27)	slot	P	
(28)	zone	OBL, PP, LOC, adv.	
(29)	slot	discourse markers	<i>ní, ví</i>

(which in turn modifies the verb core). Since a detailed study of the semantics of *va* lies outside the scope of this chapter, our analysis is preliminary. In the current study, we assume the following: (i) when *va* appears directly after the verb core, it modifies the verb and this position is thus included in the planar structure (at 18); (ii) when *va* appears after one or more adverbials, it modifies the verb and these positions are thus also included in the verbal planar structure (at 6, 10, and 24); and (iii) in all other cases, *va* does not modify the verb, thus these appearances are excluded from the planar structure.

- (1) a. *tatùun xàà xínì-va míi-ntí ña kò'va*
 like already INCMPL.know-ADD TOP-1PL.EXCL CLF.THING amount
chikàà-ntì
 POT.put(invisible)-1PL.EXCL
 'so we already know what amount to put in' SMD-0020-Huauzontle
- b. *sààn na kuntxati-ó iin rátó lo'o-va ini kasun*
 so MOD POT.wait-1PL.INCL one moment little-ADD inside POT.toast
kueé và'a-va
 slowly good-ADD
 'so we will wait a little moment longer so that it gets well toasted
 slowly' SMD-0020-Huauzontle
- c. *taa ñà xàà ntóvà kíi sévóyá-va*
 and CLF.THING already INCMPL.sprout COP onion-ADD
tùyá'à-va xàà ntóvà ntxi'i-va
 CLF.WOOD.chile-ADD already INCMPL.sprout POT.finish-ADD
tú-kán
 CLF.WOOD-DEM.PROX
 'and what is already sprouting here is onion and the chile plant here
 has already sprouted' SMD-0009-Jardin

The next issue we address concerns intensifiers and adverbials, which can appear before and after the verb core, but exhibit peculiar behavior with respect to ordering. SMD has a variety of intensifiers (we have identified six so far), some of which can combine with verbs. They are all translated as 'a lot, very much', but it is likely that there are slight semantic differences among them that we are not yet aware of. They can be grouped into two positions based on co-occurrence restrictions: if there is more than one intensifier, *kuà'à* and *tóntó*² have to appear

²This is clearly a loan from Spanish that has taken on a new function. It could be derived either from *tonto* 'dumb, foolish', which has also been borrowed as an intransitive verb 'to be stupid', or possibly from *tanto* 'so much', which is closer in meaning.

before either *yáá* or *ntxìvà'a*, as illustrated in (2). This leads us to add two slots to the verbal template.

- (2) a. *chíntxeé ta'an tóntó ntxìvà'a-na*
 INCMPL.help RECP INTENS INTENS-3PL.HUM
 'they really help each other a lot' elicited
- b. * *chíntxeé ta'an ntxìvà'a tóntó-na* elicited
- c. *itxààn sáchuun kuà'à yáá kì'vi-i*
 tomorrow POT.do.work INTENS INTENS sister[F]-1SG
 'tomorrow my sister is going to work a lot' elicited
- d. * *itxààn sáchuun yáá kuà'à kì'vi-i* elicited

When an intensifier and the adverb *tiki* 'again' combine, they exhibit variable ordering, but only if no other intensifiers or slots after the verb are present. If there are other elements present, the variable ordering is blocked, as in (3). Whether the elements appear before or after the verb has no effect on this constraint, cf. (4). This suggests that in longer constructions, fixed mini-constituents have formed, perhaps based on frequency of usage.

- (3) a. *chíntxeé ta'an ntxìvà'a tiki-na*
 INCMPL.help RECP INTENS INTENS-3PL.HUM
 'they again help each other a lot' elicited
- b. * *chíntxeé ta'an tiki ntxìvà'a-na* elicited
- c. *chíntxeé ta'an tóntó ntxìvà'a tiki-na*
 INCMPL.help RECP INTENS INTENS again-3PL.HUM
 'they again help each other a lot' elicited
- d. * *chíntxeé ta'an tiki tóntó ntxìvà'a-na* elicited
- (4) a. *itxààn sáchuun tiki yáá kì'vi-i*
 tomorrow POT.do.work again INTENS sister[F]-1SG
 'tomorrow my sister is going to work a lot again' elicited
- b. *itxààn sáchuun yáá tiki kì'vi-i* elicited
- c. *itxààn yáá tiki sáchuun kì'vi-i* elicited
- d. *itxààn tiki yáá sáchuun kì'vi-i* elicited

To adequately represent this in the planar structure, we set up three positions that are slots but can contain either an intensifier or an adverbial depending on the construction. These positions have to be repeated before the verb, since

these elements can also appear before the verb core, as mentioned above. The constraint on the ordering cannot be represented in the planar structure, but that is true for other co-occurrence constraints as well.

The third and final issue concerns the ordering of the already introduced additive *va* and *tiki* ~ *ti* 'again'. Based on examples like the one provided in (5), we had first analyzed them as variably ordering with respect to each other. However, it is more straightforward to analyze this as fixed ordering with *va* appearing in different slots, one directly after the verb and one after *tiki* ~ *ti*, since these slots are necessary anyway to account for other constructions. The same reasoning is applied to cases in which *va* and *tiki* appear before the verb core.

- (5) a. *kusi tiki va-ó*
 POT.sleep again ADD-1PL.INCL
 'We (incl.) will go to sleep again.' all elicited
- b. *kusi-va tiki-ó*
- c. *kusi-va-ti-ó*
- d. **kusi-ti-va-ó*
- e. **kusi-va-ó tiki*

2.3 Elaboration on the verbal planar structure and its positions

We now turn to the positions preceding the verb core. Position 16 contains the no longer productive element *chi*. Historically, it is derived from the verb *chi'i* 'sow', which in the past had a more general meaning 'put' (still present in other varieties of Mixtec). This more general meaning seems to be still present in most verbs formed with *chi*. Otherwise, it is difficult to pinpoint the exact function of *chi*. It combines with intransitive and transitive verbs, but also with nouns and adverbials. The result is always transitive, so we gloss this element as a transitivity marker.

In position 15, we find the iterative marker *nta* ~ *nti*. The allomorphy is neither phonologically nor semantically conditioned and often either allomorph can be used with the same verb base with no difference in meaning. This marker can co-occur with the transitivity marker *chi*.

In position 14 we find the mutually exclusive potential and completive markers *ku* and *xì*. The latter always co-occurs with the completive marker *i* or *nì*. These markers are only present with certain inflectional classes of verbs (hence the term 'class markers'). Other verb classes exhibit different marking for these categories.

The elements in positions 16 through 14 (or, rather, the elements in other Mixtec languages that are cognate with these) are usually described as derivational

prefixes and are written together with the verb core in descriptions of other Mixtec varieties (e.g., Macaulay 1996, Hollenbach & Erickson 2013). In the practical orthography of SMD we also opted to write these elements together with the verb.

Position 13 contains the productive causative marker *sá*, derived from the verb *sá'a* 'do, make'.

In position 12 we find the potential negation *i* and *o ~ u* and the general completive marker *ì*. These two elements can never co-occur, so it would also be possible to represent them in two adjacent slots (in either order). However, no evidence could ever be provided for favoring one order over the other; therefore we represent them together in one slot, since we have evidence for both of them that they are positioned between the modal markers and the causative. The potential negation can be marked either by *i* or *u~o* – these two markers are completely interchangeable for every verb. We have not yet determined the rules of the allomorphy for *u~o*. We hypothesize that historically the allomorphy was phonologically conditioned, such that verb cores with back vowels would have been marked with *o* and the rest with *u*. However, now we find exceptions to this rule, probably due to the lexicalization of certain combinations.

Position 11 consists of the elements *ná* and *nì*. We currently have only a limited understanding of their exact semantics and functions and we hope to investigate this issue more closely in the future. The element *ná* combines with the potential form of verbs and often appears in contexts of events that have not yet taken place but are desired to occur. This analysis fits well with what has been found for cognate forms in other Mixtec varieties, which have been described as marking deontic modality (Macaulay 1996: 76–78). It is thus quite probable that *ná* also has this function in SMD. The element *nì*, on the other hand, combines with the realis form of verbs, and it only occurs in completive contexts alternating with *ì*. Comparison with other Mixtec varieties is not as instructive in this case, because the completive is either marked with tone alone (e.g., San Martín Peras Mixtec), or only displays a marker *ni* (e.g., Chalcatongo Mixtec, cf. Macaulay 1996: 74–75). We take *ì* to be the basic, unmarked form, since it is more frequent and the one given in elicitation. We suspect that *nì* might mark deontic modality of past events, and so diachronically it might represent a combination of *ná* and *ì*. In the practical orthography, the modals are written as separate words; in Macaulay (1996)'s grammar they are written as a prefix (with a hyphen).

Positions 10 through 6 are filled out by the additive, the intensifiers, and the repetitive discussed above. Position 5 contains a zone with various adverbials, such as temporal ones like *vitxi* 'now, today' and *itxààn* 'tomorrow'; aspectual ones such as *xàà* 'already'; and adverbials expressing manner like *sa'a* 'like that', *và'a* 'good,

well', among others. They can variably order with one another with no difference in meaning or scope.

Directly preceding this zone is the realis negation marker *kòó* in position 4. The focus position in 3 can contain an NP expressing an argument, but also non-arguments of any kind, e.g., prepositional phrases. In position 2 we find content question words, such as *nishi* 'how', *ntxáa* 'where', *yoo* 'who', etc. The first position contains conjunctions and connectors of various types, as well as the polar question marker *an*. This concludes the discussion of the positions before the verb core; we now move on to the positions after the verb core that have not been discussed.

Between the additive (in 18) and the intensifier (in 21) discussed above, there are two additional slots: one for the reciprocal marker *ta'an* in position 19, and one for *tuun* 'always, habitually' in position 20. We suspect that other adverbial expressions might be able to appear in the latter position, but we have not been able to find specific examples.³

After positions 21 through 24, we find *ini* which can be translated as 'inner core, being (of a person)'. This element is often obligatory with verbs denoting mental or emotional states or processes, such as *ntiku'un ini* 'remember', *kutátxí ini* 'be sad', or *koto ini* 'look at somebody from askance'.

In positions 26 and 27 we find the arguments of the verb, expressed either as full noun phrases or as pronouns. Both are unmarked, but the S/A argument must come first, before the P argument. Furthermore, independent pronouns can only occur as P arguments after the verb.

After the arguments, position 28 contains a zone with optional prepositional phrases, locatives, oblique arguments, adverbials, etc. These can variably with one another, thus the designation as a zone. The last position 29 contains discourse markers such as *ví* 'certain' and *ní* 'affirmative'.⁴ To sum up, the verbal planar structure of SMD consists of 29 positions, 16 before the verb core and 12 after it.

It is instructive at this point to compare the planar structure for SMD with Macaulay (1996) proposed template for Chalcatongo Mixtec, the only other Mixtec variety for which constituency has been investigated. It should be noted, however, that this variety is spoken in the Mixteca Alta region and is not closely related to SMD. The template (based on hierarchical bracketing) includes a total of 12 positions, 7 before the verb and 4 after. We summarize her proposal

³For example, the semantically similar *taki* 'always' cannot appear in this position.

⁴Further research is needed to clarify the exact function of each of these markers. So the labels given here are preliminary.

below, combining the “basic sentence structure” with the “relative ordering of inflectional prefixes” (Macaulay 1996: 79, 146):

- (6) TP[[TOPIC] S'[[[NEG.FOC][FOC]] NEG=S[V'[(ADV) (TEMP-COMP-PL-)V (ADV)]
=ADD/RES=PRO (XP*)]]]]

The examples below show different elements of Macaulay’s template for Chalcatongo Mixtec. The examples in (7a) and (7b) show the preverbal positions of topic and focus, whereas (7c) and (7d) show the ordering of negation markers, adverbs, and the temporal and additive markers.

- (7) a. *roʔo tú=kúʔu=ro*
2SG NEG=be.sick=2SG
‘As for you, you aren’t sick.’ (Macaulay 1996: 106)
- b. *pero niasu xí x^wá tandaʔá=∅ či tándaʔá=∅ xí péðrú*
but NEG.FOC with Juan marry=3SG because marry=3SG with Pedro
‘But it isn’t Juan who she’s marrying, she’s marrying Pedro.’
(Macaulay 1996: 123)
- c. *sókó tú=šãã kúñú=∅*
well NEG=much deep=3SG
‘The well is not very deep.’ (Macaulay 1996: 120)
- d. *ni-žéé=ka=rí takú ásu róʔó*
COMPL-eat=ADD=1SG taco than 2SG
‘I ate more tacos than you did.’ (Macaulay 1996: 141)

Her template is similar to ours in that there are more preverbal positions than postverbal ones. The positions of the focus marker and the realis negative marker also correspond quite closely to our findings. It is also similar in that it recognizes that certain elements can appear either before and after verb, although she simply groups them together as adverbs. Chalcatongo Mixtec also has an additive marker, but it is represented only once in Macaulay’s template. It would be interesting to know whether its single occurrence in the template is due to differences between the markers or due to differences in the methodology of establishing templatic structures.

3 Phonological domains

In this section we discuss the diagnostics that identify phonological domains. Unlike what has been reported for other varieties of Mixtec (cf. Hunter & Pike 1969,

Daly 1973, Macaulay 1996, Hollenbach 2003, among many others) and other Otomanguan languages (cf. Campbell 2024, Gutiérrez & Uchihara 2024, Nakamoto 2024 [this volume]), SMD exhibits few tonal processes and few general phonological rules.

We identify three phonological processes that apply to the verb complex, of which two concern segments and one concerns tone. These are: vowel overwriting, bimoraicity, and tone sandhi of dependent pronouns. The first two must be fractured into a minimal and maximal domain to render consistent spans, resulting in a total of five diagnostics. Throughout this section we also provide IPA transcriptions for the examples. These are given in square brackets underneath the orthographic representation. For tone representation we chose numbers rather than bars for better readability. The low tone is represented by 1, the mid tone by 3, and the high tone by 5.

3.1 Bimoraicity constraint (12-18, 1-27; 17, 1-28)

Mixtec varieties are known for their preference for bimoraic “prosodic words” (cf. Pike 1948, Penner 2019 on Ixtayutla Mixtec, and Uchihara & Mendoza Ruíz 2022 on Alcozauca Mixtec, among others). This means that free forms have a strong tendency to be bimoraic – that is, to have two vowels.⁵ This is also the case in SMD, where lexical free forms minimally have the structure CVCV (e.g., *titi* ‘paper’), CVV (e.g., *nùù* ‘face’), CV?V (e.g., *tù'un* ‘word, language’), or VCV (e.g., *àsì* ‘tasty’).

There are two ways this general observations can be applied as a constituency diagnostic. It is important to note that the verb base, like any other lexical item, cannot be monomoraic, but rather has to be (at least) bimoraic. First, we can look at the smallest and largest spans that contains only monomoraic forms (excluding the verb base). These could be equated with larger “prosodic words”, given that these spans contain only one bimoraic element, the verb base. Second, we can look at the smallest and largest span overlapping the verb base that contains bimoraic forms. These could be interpreted as the verb phrase since these spans contain multiple bimoraic forms.

We start with the span overlapping the verb base that contains only monomoraic elements (apart from the verb base which cannot be monomoraic). Here we discuss both the minimal domain (i.e. the smallest span) and the maximal domain (i.e. the largest span). As outlined above, wordhood in Mixtec is often associated with bimoraicity and thus the minimal span should correspond to what is termed

⁵Long vowels count as bimoraic, i.e. as two vowels.

a “phonological/prosodic word” in other descriptions (Uchihara & Mendoza Ruíz 2022, Penner 2019).

In SMD, the minimal monomoraicity diagnostic identifies the span from 12 through 18. Apart from the verb base, this span includes all the elements usually classified as prefixes and written together with the verb, as well as the additive marker *va* when it appears directly after the verb core. The additive marker in position 10 cannot be included in this span, because – as mentioned in §2 – it is left-leaning and thus cannot appear in this position without a preceding bimoraic element. Despite being monomoraic in form, the modals in position (11) must be excluded as well, because they cannot appear without a preceding clause linker (e.g., a subordinator or conjunction). Note also that this minimal domain excludes pronouns, so it can only be applied with imperatives and impersonal verbs, since all other verbs require at least one argument to be present to form a complete utterance (see §4.1 for more details). An example is provided in 8.

- (8) *i-tàan-va*
 [i¹-tã:¹³-βa³]
 12-17-18
 CMPL-quake-ADD
 ‘It quaked (after having quaked before).’ elicited

The maximal interpretation of the monomoraicity diagnostic identifies the whole verbal planar structure to the exclusion of the last position in 28, which only contains bimoraic elements. An example is provided in (9) with a polar question and both A and P arguments realized as dependent pronouns.

- (9) *An i-tàshì-ùn-ñā?*
 [ã³ i¹-ta¹ʃi¹-ũ¹-ɲa³]
 1 12-17-26-27
 Q CMPL-crush-2SG.NHON-3.THING
 ‘Did you crush it?’ elicited

Next we will turn to the span overlapping the verb base in which all positions are filled with bimoraic elements. This diagnostic also has to be fractured into a minimal and maximal domain. The minimal domain is just the verb base in position 17 since, as mentioned above, it is always at least bimoraic. The maximal span covers the whole planar structure apart from the last slot (position 29) which contains monomoraic discourse markers – that is, the span runs from position 1 through 28.

Note that the maximal spans from both interpretations are almost identical. This, together with the fact that the minimal and maximal domains identify spans of vastly different sizes (1 and 7 vs. 27 and 28 positions), suggests that bimoraicity might not be an informative diagnostic for constituency in SMD.

3.2 Vowel overwriting after glottal stop (17-26; 6-29)

This diagnostic is based on a phonological process in which final vowels are replaced or overwritten by the initial vowel of the following element. More precisely, when an element of the structure $CV_i?V_i$ is followed by a vowel-initial monomoraic pronoun, the final vowel of that element is replaced with that of the pronoun. Whether or not the nasality of the overwritten vowel is preserved depends on the pronoun (cf. Table 4). The rule is formalized below:⁶

$$(10) \quad (X)CV_i?V_i+V_j \rightarrow (X)CV_i?V_j$$

Table 4: Vowel-initial dependent pronouns

Pronoun	Gloss	Nasality
<i>i</i>	1SG	preserves nasality of base
<i>ò</i>	1PL.INCL	does not preserve nasality of base
<i>un</i>	2SG.NHON	always nasal
<i>àn</i>	3SG.F	always nasal
<i>an</i>	3SG.THING	always nasal

Instead of making reference to final vowels, this process could alternatively be described as targeting rearticulated vowels around the glottal stop. Tonal processes targeting this same domain are attested in Huajuapán Mixtec (Pike & Cowan 1967).⁷ There are two reasons we do not adopt the rearticulation analysis. First, while in most cases the vowels around the glottal stop are identical, this is not always the case and with non-identical vowels it is difficult to imagine that we are dealing with rearticulation. Second, the descriptive facts remain the same whether we refer to the domain as “final vowel” or “rearticulated vowel”.

Examples (11a) and (11b) show the rule applying to a noun and a verb, respectively. (11c) shows that the process also applies when the vowels are non-identical

⁶X = additional syllable in trisyllabic words, either V or CV, e.g., *àsì'i* ‘wife’ or *txiya'à* ‘gallon (container)’.

⁷We thank Taylor Miller for pointing us to this alternative analysis.

(with a different pronoun to make the process more visible). In examples (11d) and (11e), we see that the rule does not apply when the glottal stop is followed by a consonant.

- (11) a. $yé'é + i \rightarrow yé'-i$
 $[ʒe^5ʔe^5] + [i^1] \rightarrow [ʒe^5ʔi^1]$
 door + 1SG \rightarrow 'my door'
- b. $inù'ùn + i \rightarrow inù'-ìn$
 $[i^1nu^1ʔũ^1] + [i^1] \rightarrow [i^1nu^1ʔi^1]$
 CMPL.go.home + 1SG \rightarrow 'I went home'
- c. $ntxè'i + un \rightarrow ntxè'-ùn$
 $[^ndʒe^1ʔi^1] + [ũ] \rightarrow [^ndʒe^1ʔũ^1]$
 clay + 2SG \rightarrow 'your (sg.) clay'
- d. $ko'nto + i \rightarrow ko'nto-i$
 $[ko^3ʔ^n do^3] + [i^1] \rightarrow [ko^3ʔ^n do^3 i^1]$
 bone + 1SG \rightarrow 'my bone'
- e. $xá'ntxá + i \rightarrow xá'ntxá-i$
 $[ca^5ʔ^n dʒa^5] + [i^1] \rightarrow [ca^5ʔ^n dʒa^5 i^1]$
 INCMPL.cut + 1SG \rightarrow 'I'm cutting (sth.)'

There is one exception to this process: the back vowel [o] at the end of the base will overwrite [u] of a monomoraic element. Examples (12a) and (12b) illustrate the different vowel overwriting for back vowels with a noun and a verb base, respectively.

- (12) a. $kò'ò[ko^1ʔo^1] + un \quad [ũ] \rightarrow kò'-òn [ko^1ʔ-ò^1]$
 plate + 2SG.NHON \rightarrow 'your plate'
- b. $ntó'o[^ndo^5ʔo^3] + un \quad [ũ] \rightarrow ntó'-ón \quad [^ndo^5ʔ-ò^5]$
 INCMPL.suffer + 2SG.NHON \rightarrow 'you are suffering'

Vowel overwriting is observed with vowel-initial dependent pronouns (cf. Table 4) in position 26 following a CV?V base, and with elements in position 22, such as the intensifier *ntxìvà'a* (cf. example 13b). Thus, the span from 17-26 provides positive evidence for this process, i.e. the minimal span.

Negative evidence, however, can only be found for slots/zones that contain elements of the relevant structure. Slots 13-16 can never provide any evidence for or against vowel overwriting: the elements found there do not contain a glottal stop, nor are any of the immediately following elements vowel-initial. Therefore we fractured the test so as to also include a maximal domain, to identify the span

in which there is no negative evidence for vowel overwriting. This identifies a much larger span, ranging from position 6 through 28. Negative evidence can be found in position 5, with the adverbial *sa'a* 'like this/that' never taking part in this process (cf. (13c), and after position 29 at the clause boundary.

- (13) a. *ta sáàn ì-sùvá'-ì* *ì-sísini-va-ì*
 [ta³ sã:⁵¹ i¹-su¹βa⁵ʔ-i¹ i¹-si⁵si⁵ni¹-βa³-i¹]
 1 5 11-17-26 11-17-18-26
 and then CMPL-prepare-1SG CMPL-have.breakfast-ADD-1SG
 'And then I prepared breakfast.' SMD-0009-Jardin
- b. *lo'o ntxivá'-ì* *ì-xì'ì* *nánà-ì* *tátà-ì*
 [lo³ʔo³ n³dzi¹βa¹ʔ-i¹ i¹-çi¹ʔi¹ na⁵na¹-i¹ ta⁵ta¹-i¹]
 17 22-26 12-17 26
 be.small INTENS-1SG CMPL-die mother-1SG father-1SG
 'I was very little when my mother and father died.' SMD-0059-Padres
- c. *ta sa'a sa'a ì-nto'-án*
 [ta³ sa³ʔa³ sa³ʔa³ i¹-n³do³ʔ-ã⁵]
 1 5 5 12-17-26
 and like.that like.that CMPL-happen-3SG.F
 'And like that like that it happened to her.' SMD-0047-Cena

In other varieties, this process applies to a wider range of bases, e.g., in Alcozauca Mixtec (Uchihara & Mendoza Ruíz 2022). In SMD, there is also a more general process of vowel overwriting, but it follows different rules. In connected speech, the first person plural inclusive marker *ò* often overwrites a final [a] or [u] of the preceding element. However, when asked to repeat the forms, speakers will undo this overwriting, e.g., *kaxá'an-v-ó* [eat-ADD-1PL.INCL] 'we will eat', which is repeated back as *kaxá'an-va-ó*. This never happens with the pronoun overwriting process described above. In fact, examples like **ì-sùvá'a-ì*, repeated from (13a) but with the final vowel restored, are deemed ungrammatical. Because the more general process is largely dependent on register and speech tempo, we do not discuss it further.

3.3 Tonal processes (17-27)

This diagnostic concerns the tonal changes triggered by the tone of adjacent elements and it excludes the tonal marking of inflection, which is discussed in §4.2. In SMD, tonal processes are quite rare, and in verbal predicate constructions they appear to be limited to dependent pronouns.

Dependent pronouns show interactions with their host with regard to their tonal realizations, i.e., they exhibit tone sandhi. These interactions fall into four groups and are summarized in Table 5. It is important to underscore that the tone sandhi processes identified are only observed with dependent pronouns and do not operate elsewhere in the language. Dependent pronouns in Group 1 do not exhibit tone sandhi and thus will not be discussed further. Group 2 consists of only one pronoun – ‘second person non-honorific’ *un* – which copies the tone of the preceding element. Groups 3 and 4 show alternations in similar contexts, but with different realizations. A detailed investigation and description of the sandhi patterns lies outside the scope of this chapter. Our observations so far indicate that the tone realizations are not only sensitive to the phonological characteristics of the preceding element, but also to its word class.

Table 5: Dependent pronouns and their tone realizations

Group	Generalization	Pronouns
1	no tone changes	1SG <i>ì</i> , 2SG.HON <i>ní</i>
2	tone copying	2SG.NHON <i>un</i>
3	L alternating with H	1PL.INCL <i>ò</i> , 3SG.F <i>àn~nà</i> , 3.ANIM <i>rì</i> , 3.WOOD <i>dùn</i>
4	L alternating with M	1PL.EXCL <i>ntì</i> , 3SG.M <i>rà</i> , 3PL <i>nà</i>

The tone sandhi diagnostic is applied so that it identifies the span overlapping the verb core, which contains the elements triggering tone sandhi on dependent pronouns. Given that dependent pronouns can never appear before the core – except in focused NPs, which are not discussed in this chapter – the left-most element they can interact with is the verb core. Examples (14a) and (14b) show that the verb core indeed triggers tone sandhi on the dependent pronoun *un* ‘second person singular non-honorific’.

- (14) a. *ta sa'a káchí-ún*
 [ta³ sa³?a³ ka⁵tʃi⁵-ũ⁵]
 1 5 17-26
 and like.that INCMPL.say-2SG.NHON
 ‘And that’s how you say it.’ SMD-0047-Cena

- b. *vitxi i-kixà-ùn* *yó'o*
 [βi³tɕi³ i¹-ki³ɕa¹-ũ¹] ʒo⁵ʔo³]
 5 12-17-26 28
 now CMPL-arrive-2SG.NHON DEM.PROX
 'Now you arrived here (...)' SMD-0047-Cena

Tone sandhi can also be observed with elements in positions 21 to 25, illustrated by the tone realization of *un* in examples (15a) and (15b). Elements in positions after the pronouns do not influence the tone realizations of pronouns. In examples (15a) and (15c) the tone realization of the dependent pronoun *un* is the same regardless of the tone of the element following it.

- (15) a. *su i-kuntàà* *ini-un* *gueritá*
 [su³ i¹-ku³ⁿda:¹¹] i³ni³-ũ³ we³ri⁵ta⁵]
 1 12-17 25-26 27
 but CMPL-understand inside-2SG.NHON white.person
 'But you understood *guerita* (light-skinned girl).' SMD-0047-Cena
- b. *ta sààn nì i-sàma* *ntxivà'-ùn*
 [ta³ sã:⁵¹ ni¹ i¹-sa¹ma³ n¹ɕi¹βa¹ʔ-ũ¹]
 1 5 11 12-17 22-26
 and then MOD CMPL-change INTENS-2SG.NHON
 'And so you've changed a lot.' elicited
- c. *su i-kuntàà* *ini-un* *shità*
 [su³ i¹-ku³ⁿda:¹¹] i³ni³-ũ³ ʃi¹ta¹]
 1 12-17 25-26 27
 but CMPL-understand inside-2SG.NHON tortilla
 'But you understood *shità* (tortilla).' elicited

Dependent pronouns used as P-arguments also exhibit tone sandhi, as illustrated in examples (16a) and (16b).

- (16) a. *ta sààn jààn* *chikàà-ò-ñā*
 [ta³ sã:⁵¹ hã:¹¹ tʃi³ka:¹¹-o¹-ɲa³]
 1 5 3 17-26-27
 and then DEM.DIST POT.put(invisible)-1PL.INCL-3.THING
 'and so we'll put it in' SMD-0020-Huauzontle

- b. *ta sáani chikàà-na-ñà*
 [ta³ sã:⁵⁵ni³ tʃi³ka:¹¹-na³-ɲa¹
 1 5 17-26-27
 and also POT.put(invisible)-3PL.HUM-3.THING
kuchúun-na-ñà jí'in-ña
 ku³tʃü:⁵³-na³-ɲa¹ hi³ʔi³-ɲa³]
 17-26-27 28
 POT.use-3PL.HUM-3.THING with-3.THING
 ‘And also they put it in and use it with that.’ SMD-0008-Hierbas

This diagnostic thus identifies a span from position 17 through 27.

3.4 Spans identified by phonological domains

Table 6 summarizes all the phonological diagnostics and their results. None of the spans converge, but two of them start at the verb core and two of them end at the P-argument slot. Given how much importance is ascribed to the bimoraic minimality constraint to identify prosodic/phonological words in Mixtec, we would have expected that it correlates much more with the other phonological domains. The absence of such convergences might indicate that bimoraicity does not play an important role for phonological constituency in SMD.

Table 6: Phonological diagnostics and their results

Diagnostic	Fracture	Left Edge	Right Edge	Size	Section
Bimoraicity	min	12	18	7	3.1
Bimoraicity	max	1	27	27	3.1
Vowel overwriting	min	17	26	10	3.2
Vowel overwriting	max	4	28	24	3.2
Tone sandhi	-	17	27	11	3.3

4 Indeterminate domains

In this section, we discuss the spans identified by diagnostics that could either be interpreted as phonological or morphosyntactic, depending on the theoretical background and morphemic analysis. They involve two diagnostics: free occurrence and deviations from biuniqueness.

4.1 Free occurrence (17; 14-27; 11-27)

Free occurrence is defined as the ability of an element to stand alone as a complete utterance. There are two interpretations of this diagnostic: we can look for the smallest (minimal) and largest (maximal) span that fulfills this criterion.

In the minimal interpretation, this diagnostic identifies the shortest span overlapping the verb core that can be single free forms. In SMD, imperatives and impersonal verbs can be used on their own as a single free form. They are marked for aspect-mood by tone but appear without any further segmental marking or person indexing (cf. examples (17a) of an impersonal verb and (17b) of an imperative). The diagnostic thus identifies just the verb core in position 17.⁸

- (17) a. *táan*
 17
 INCMPL.quake
 'There's an earthquake (lit: [it] is quaking).' elicited
- b. *kà'àn*
 17
 POT.speak
 'Speak!' elicited

In the maximal interpretation, this diagnostic identifies the longest span overlapping the verb core that can be a single free form. In SMD, the application of this test results in two different spans, depending on the interpretation of the causative formative *sá*. In the following, we will illustrate the issue and present the competing results.

The causative marker *sá* in position 13 is clearly related to the verb *sá'a* 'do, make'. There are two possible analyses here: i) the causative can be analyzed as a shortened form of *sá'a*, given that forms of the structure CV?V regularly contract to CV(V) in connected speech;⁹ or ii) the causative marker *sá* is a separate element that is only diachronically related to the verb *sá'a*. There is evidence for either interpretation and it is not clear *a priori* which interpretation is the correct one.

If the causative marker *sá* is taken to be a shortened form of the verb *sá'a* – a free form – and thus the same element, then the left edge of the construction is at position 14, i.e. right after the causative. If the causative marker *sá* is taken

⁸Note that a (non-imperative, non-impersonal) declarative verb cannot stand on its own as a complete utterance, but minimally appears with an S/A argument.

⁹Macaulay (1987) calls this "fast speech reduction".

to be a separate element from the verb *sá'a*, then the left edge of the span is at position 11. All positions after that and before the verb core have elements that are not free forms.

The right edge of the span is not affected by this issue and is in either case at position 27, i.e. it ends with dependent patient pronouns. Note that the full span can only be realized if no elements in positions 19 and 22 are present, since these are free forms. However, none of these forms is obligatory. In addition, due to an asymmetry in local versus non-local arguments, this only applies to third person patients, since first and second person patients have to be expressed by an independent (free) pronoun. Some examples of *long* free forms (indicated with square brackets) from naturalistic speech are provided in examples (18a) to (18c).

- (18) a. *ta ikán [sá-ntxixà-ntò-an]*
 1 5 13-17-26-27
 and there CAUS-melt-2PL-3.THING
 ‘And then they dissolve it (...)’ SMD-0033-Espiritus
- b. *kuî-rì chii saa [i-kintxaa-va-i-rì]*
 17-26 1 5 12-17-18-26-27
 green-3 because like.that CMPL-take.away-ADD-1SG-3
 ‘They are green because I just cut them.’ SMD-0062-Juana
- c. *sa'a-va koo-rà ta xàà [na]*
 5-24 17-26 1 5 11
 like.that-ADD POT.stay-3SG.M and already MOD
chikà-à-i-ra] ...
 17-26-27
 put(invisible)-1PL.INCL-3SG.M
 ‘It [the dried chili] will stay like this and then when I add it [to the pot] (I will add a bit more water to it)’ SMD-0020-Huauzontle

4.2 Deviations from biuniqueness (17; 13-17; 15-17; 12-23)

In this section, we discuss instances of deviations from biuniqueness, i.e. cases in which there a deviation from a one-to-one form-meaning correspondence. Such deviations have been associated with morphological structure or word-internal structure. In SMD, we find two types of deviations from biuniqueness: one form that codes multiple meanings (one-to-many), and multiple forms that express the same meaning (many-to-one). The latter is more commonly found in the verbal planar structure of SMD than the former. This diagnostic can be applied

in two ways, a minimal fracture, identifying the smallest span overlapping the verb core that exhibits deviations from biuniqueness, and a maximal fracture, which identifies the largest span that can show deviations.

As mentioned in §2, SMD verbs fall into inflectional classes. The aspect-mood exponents of these inflectional constitute many-to-one deviations. While the completive form is always marked with a preverbal element *ì* or *nì*, the incomplete and potential forms are often only marked by tone, the former showing a characteristic high tone on the first mora. The minimal interpretation of this diagnostic identifies the shortest span where tonal inflection can be observed. This consists of just the verb core in position 17 (cf. Table 7).

Table 7: Verbs (position 17) showing tonal inflection of different inflectional classes

INCMPL	CMPL	POT	Gloss	Class
<i>tívi</i>	<i>itìvi</i>	<i>tìvi</i>	‘wake up’	INCMPL high
<i>xú'ní</i>	<i>ixu'ní</i>	<i>ku'ní</i>	‘squeeze’	INCMPL high with stem alternation
<i>íin</i>	<i>ixòo</i>	<i>koo</i>	‘live, stay’	INCMPL high with suppletion
<i>núná</i>	<i>inúnà</i>	<i>núnà</i>	‘be open’	INCMPL with both morae high
<i>kàì</i>	<i>ikàì</i>	<i>kàì</i>	‘burn’	INCMPL mid
<i>kuà'àn</i>	<i>ixà'àn</i>	<i>kù'ùn</i>	‘go’	segmental alternation only

There is also a maximal interpretation of this diagnostic, which identifies the largest contiguous span overlapping the verb base that exhibits tonal inflection. In addition to the verb core, tonal inflection can also occur on the transitivizer marker *chi* and the iterative marker *nti/nta*, but not on the inflectional class markers *ku* and *xì*, nor on the causative marker *sá* (see Table 8 for examples). However, there are other positions that exhibit many-to-one relations. The maximal span of this diagnostic is, therefore, larger than that identified by tonal inflection.

The maximal domain identified by the many-to-one deviation ranges from position 12 to 23. The potential negation has three allomorphs: whether a verb takes *u* or *o* is lexically determined, but all verbs can alternatively take *i*, without any difference in meaning. After the verb base, this type of deviation from biuniqueness can be found in the adverbial *tiki~ti* in position 23. Examples (19a) to (19c) show such a span with three different forms, but with the same meaning.

¹⁰Examples are given with morpheme segmentation for convenience. Abbreviations: invis. = invisible; there are several ‘put’-verbs depending on whether the object is being placed inside of a container and thus becomes invisible, or remains visible after relocating it.

Table 8: Causative, iterative, and derived verbs showing tonal inflection¹⁰

INCMPL	CMPL	POT	Meaning	Morphemes
<i>chí-ntoo</i> 16-17	<i>ì-chi-ntoo</i> 12-16-17	<i>chi-ntoo</i> 16-17	'put down, stack'	CAUS-be
<i>ntá-koto</i> 15-17	<i>ì-nta-koto</i> 12-15-17	<i>nta-koto</i> 15-17	'mend'	ITER-take.care
<i>ntá-chi-kàà</i> 15-16-17	<i>ì-nta-chi-kàà</i> 12-15-16-17	<i>nta-chi-kàà</i> 15-16-17	'put again'	ITER-CAUS-put(invis.)
<i>sá-keta</i> 13-17	<i>ì-sá-keta</i> 12-13-17	<i>sá-keta</i> 13-17	'finish sth.'	CAUS-put
<i>nání</i> 17	<i>ì-xì-nání</i> 12-14-17	<i>ku-nání</i> 14-17	'be called, be named'	

- (19) a. *u-ka'ntxa-ti-un*
12-17-23-26
NEG.POT-cut-again-2SG.NHON
'Don't cut your finger again!' elicited
- b. *i-ka'ntxa-ti-un*
12-17-23-26
NEG.POT-cut-again-2SG.NHON head hand-2SG.NHON
'Don't cut your finger again!' elicited
- c. *i-ka'ntxa tiki-un*
12-17 23-26
NEG.POT-cut again-2SG.NHON head hand-2SG.NHON
'Don't cut your finger again!' elicited

The spans identified in this way do not coincide with those identified by the one-to-many deviations, which means that this diagnostic has to be fractured by type of deviation from biuniqueness and then further into a minimal and maximal domain for each.

The smallest span overlapping the verb core that exhibits one form with multiple meanings (one-to-many) is just the core in position 17. There is a small class of verbs that have the same form in the incomplete and potential, as illustrated in Table 9 and examples (20a) and (20b). The largest span that can exhibit this type of deviation from biuniqueness runs from the causative marker *sá* in position 13 to the verb core in position 17, cf. examples (20c) and (20d).

- (20) a. *ta nishi sá'a-ntó ntoo-ntò vitxi*
 1 2 17-26 17-26 28
 and how INCMPL.do-2PL POT.live-2PL today
 'And how do you (pl.) manage (lit.: do it) to live today?'
 SMD-0059-Padres
- b. *kòó xínì-ì nishi sá'a-ra káa*
 4 17-26 2 17-26
 NEG.REAL INCMPL.know-1SG how POT.do-3SG.M DEM
 'I don't know how he is going to do it.' SMD-0062-Juana
- c. *míí-ní màmà sánto'o-ní míí-ní*
 3 13.17-26 27
 TOP-2SG.HON mother CAUS.suffer-2SG.HON TOP-2SG.HON
 'You (pl.), mother, you're making yourself suffer.' SMD-0059-Padres
- d. *kòó kúnì-ì sánto'o-ní*
 4 17-26 13.17-26
 NEG.REAL INCMPL.want-1SG CAUS.suffer-2SG.HON
 'I don't want you (pol.) to suffer.' elicited

Table 9: Verbs with identical forms in the potential and incomplete (one form – multiple meanings)

INCMPL	CMPL	POT	Gloss
<i>sá'a</i>	<i>isá'a</i>	<i>sá'a</i>	'do, make'
<i>nù'ùn</i>	<i>inù'ùn</i>	<i>nù'ùn</i>	'leave, go home'
<i>xàà</i>	<i>ixàà</i>	<i>xàà</i>	'rot, decompose'
<i>sá-nta-kàà</i> 13-15-17	<i>ì-sá-nta-kàà</i> 12-13-15-17	<i>sá-nta-kàà</i> 13-15-17	'spread out'

4.3 Spans identified by indeterminate domains

In Table 10, we summarize the results of the diagnostics that could be interpreted as phonological or morphosyntactic. There is one convergence that concerns the verb core: the minimal free form and the minimal domain showing one-to-many correspondences both target this span. Otherwise, there are no convergences, but note that the right edge is in many cases at the verb core. For deviations of biuniqueness, this fits well with the idea that Mixtec languages are prefixing, i.e. the verbal "word" includes a few prefixes and the core, but everything after would be syntactical.

Table 10: Indeterminate diagnostics and their results

Diagnostic	Fracture	LeftEdge	RightEdge	Size	Section
Free occurrence	min	17	17	1	4.1
Free occurrence	max – <i>sá</i> = <i>sá'a</i>	14	27	14	4.1
Free occurrence	max – <i>sá</i> ≠ <i>sá'a</i>	11	27	17	4.1
Dev. biunique.	min – one-to-many	17	17	1	4.2
Dev. biunique.	max – one-to-many	13	17	5	4.2
Dev. biunique.	min – many-to-one	15	17	3	4.2
Dev. biunique.	max – many-to-one	12	23	12	4.2

5 Morphosyntactic domains

In this section, we discuss the spans identified by morphosyntactic diagnostics. We have identified five types of diagnostics.

5.1 Non-interruptability (14-20; 11-20; 3-25)

Non-interruptability identifies the span overlapping the core that cannot be interrupted by a free form (as defined in §4.1). In SMD, as in many other languages, this diagnostic identifies differing spans if the interrupting element is taken to be one single free form or a complex free form, such as a noun phrase.

The result of the non-interruptability diagnostic with a single free form depends on the interpretation of the causative element *sá* as either a form of the verb *sá'a* or as a separate formative (cf. the discussion in §4.1). If the causative is taken to be a form of the verb *sá'a* it constitutes a free form and the left edge of the span is right before it at position 14. If taken to be a separate element and thus a bound form, the leftmost boundary occurs at position 11. This is because the intensifiers/adverbials in 9 can stand on their own, for (when answering a question, and the following additive marker in position 10 can never appear without them. The reduction of bimoraic forms in connected speech is a well-known phenomenon in Mixtec languages (Pike 1945, Macaulay 1987, Uchihara & Mendoza Ruíz 2022).¹¹

¹¹It is often referred to as “fast speech reduction” but the opposition we find has more to do with connected speech (as it occurs in conversations and narratives) versus forms spoken in isolation or carefully (as is common in elicitation) and we see the difference in speech tempo as emerging from that.

The other elements in positions 14 through 15 are all bound. The rightward boundary of the span is in both interpretations at position 20, since the reciprocal marker *ta'an* cannot be used as a free form without a verb core.

The non-interruptability diagnostic with a complex free form with internal structure (e.g., a noun phrase) identifies a large span covering most of the verbal planar structure. The left edge is at position 3, because whole NPs can be focused. On the other side, the span ends at position 25, before the argument slots, which can be fit out by complex NPs.

5.2 Non-permutability (5-19)

The non-permutability diagnostic targets the span overlapping the core that contains elements that cannot be variably ordered. As with other diagnostics, it has more than one interpretation. It can be taken to include only elements that appear in one position exclusively or it can be taken to also include elements that can variably order and produce differences in scope. Since the latter is (so far) not attested in SMD, this diagnostic does not have to be fractured. Non-permutability thus identifies the span overlapping the core containing only positions whose elements cannot be variably ordered (while meaning remains the same).

The elements in slots 6 through 16, which appear before the verb, cannot variably order and are fixed in their position. The adverbials in position 5, however, can appear in either order with no difference in meaning. This is illustrated in the examples (21a) and (21b) with *sa'a* 'like this' and *xàà* 'already'. The adverbials in position 5 thus mark the leftward boundary of this span.

- (21) a. *taa ikán xàà sa'a-va ntáa míi iti-nà*
 1 3 5 17 26
 and DEM.PROX already like.that-ADD be TOP cornfield-3PL.HUM
 ikán

DEM.PROX

'... And here, their cornfield is already like this here.' SMD-0057-Tierra

- b. *taa ikán sa'a xàà-va ntáa mii iti-nà*
 1 3 5 17 26
 and DEM.PROX already like.that-ADD be TOP cornfield-3PL.HUM
 ikán

DEM.PROX

'...And here, their cornfield is already like this here.'

elicited

Of the elements after the verb base, most can also appear before it, i.e. they can variably order with it. This does not apply to the reciprocal *ta'an* in position 19, which constitutes the rightward boundary of this span. The reciprocal cannot variably order with other elements after the verb base either. Examples which illustrate this point are provided in (22) (partially repeated from §2).

- (22) a. *chíntxeé ta'an tóntó ntxivà'a-na*
 INCMPL.help RECP INTENS INTENS-3PL.HUM
 'They really help each other a lot.' elicited
- b. * *chíntxeé tóntó ta'an ntxivà'a-na* elicited
- c. * *chíntxeé tóntó ntxivà'a ta'an-na* elicited
- d. * *ta'an chíntxeé tóntó ntxivà'a-na* elicited

5.3 Ciscategorial selection (16-17; 17; 4-23)

An element which is Ciscategorial is one that exclusively combines with bases of a specific part of speech. In this chapter we are concerned with selectivity in relation to verbs. We ask what the span is that contains only ciscategorial elements or what the largest span is that contains ciscategorial elements on its left and right edges, the difference resulting in a minimal/maximal test fracture. The minimal interpretation of this diagnostic identifies the span overlapping the core in which all elements are ciscategorial with the core, i.e. they only combine with verbs. In SMD, this the minimal domain only identifies the verb core in position 17, because the elements in positions immediately before and after are both transcategorial. The additive, as explained in §2, also combines with nouns. The transitivizer *chi* in position 16 seems to also combine with nouns, cf. Table 11. However, one can observe that the tone patterns in the resulting verb form are not the same as in the base form with both noun bases: When *chi* combines with a verb base, the tones remain the same, but when it combines with noun bases, the tones of the bases all are raised one level. One possible analysis is that *chi* does not combine with noun bases in these cases, but with tonally derived verbs. This would make it ciscategorial, rather than transcategorial.¹² Such tonal derivations do occur in other parts of the grammar of SMD, for example in the derivation of adjectives from nouns with high tone (e.g., *ishí* 'hairy' from *ishi* 'hair'). However, the phenomenon is not sufficiently well studied to resolve the matter in this chapter. We thus fracture the minimal domain further, into a fracture in which we consider *chi* ciscategorial and one in which we consider it transcategorial. In the former

¹²We would like to thank Eric W. Campbell for pointing this out to us.

interpretation, the minimal span ends at position 16, since the iterative marker *nta/nti* combines with adjectives and verbs without a change in the tone pattern of the base. In the latter interpretation, the span consists of only the verb core in position 17.

Table 11: Examples of *chi* combining with different bases

Form	Gloss	Base	Word class of base
<i>chiñú'ún</i>	worship sb.	<i>ñu'un</i> 'fire'	noun
<i>chíko'vá</i>	measure sth.	<i>kò'va</i> 'size, amount'	noun
<i>chíkani</i>	stop sth.	<i>kani</i> 'hit'	verb

The maximal ciscategorial selection diagnostic identifies the largest span overlapping the core that can contain elements ciscategorial with verbs. The left edge of this span is at position 4, since many elements that can appear in the focus slot are transcategorial. The last ciscategorial element on the right edge is the adverbial *tiki* 'again' in position 23. All elements after that are transcategorial. The element *ini*, for example, can also be used with nouns as a preposition 'inside/in'. The dependent pronouns that appear in position 26 can also be used as possessors with nouns.

5.4 Subspan repetition (12-15, 12-26; 7-25, 4-28, 2-29, 1-29)

In this section we discuss subspan repetition, i.e. constructions in which the verb core and possibly other elements of the verbal planar structure are repeated. For each construction or construction type, we identify which elements can have scope over both conjuncts (or, more technically, repeated subspans) and which cannot. The minimal interpretation of this diagnostic identifies the smallest span overlapping the verb core that contains elements that cannot have wide scope. We have only found wide scope so far with dependent pronouns in position 26, temporal modifiers such as *vitxi* 'now, today' and *xina'á* 'long ago' in position 5, content questions in position 2 and at least some of the connectors in position 1. In the maximal interpretation, we consider the largest span of structure that can be conjoined, ignoring the possibility of wide-scope. The maximal spans identified by this diagnostic are different for each of the constructions we discuss. This test thus has to be fractured into 8 diagnostics (4 constructions with 1 minimal and 1 maximal domain each).

We start with a construction in which a verb is immediately followed by another verb without overt marking of the linkage. We refer to this construction as

asyndetic verb-verb linkage (AVVL). Macaulay (1996: 154–155) discusses this construction in the context of sentential complements. This fits well with our data: we have only observed this type of subspan repetition with the second verb being used as an argument of the first verb. While juxtaposition of clauses is often associated with parataxis, in languages like Mixtec (and most other Otomanguan languages) which lack non-finite verb forms, this association of juxtaposition with parataxis is less obvious. We have not systematically investigated prosody or morphosyntactic restrictions of the repeated subspan, but it is quite possible that such a study would reveal that they are ‘subordinated’ according to at least some criteria (cf. Palancar 2012 for a detailed study on Otomi).

The largest span that can be repeated in asyndetic linkage includes the verb up to the S/A-argument in position 26. The P-argument in position 27 cannot be repeated in AVVL and thus constitutes the right edge of this diagnostic. This is not surprising given that the second verb functions as the P-argument of the first, so this position is already occupied, cf. (23). The left edge is at position 13, because the potential negation can be repeated in the complement clause, as illustrated in (23b). Elements before the potential negation cannot be repeated. Thus the maximal span in AVVL runs from position 12 to 26.

- (23) a. *távà na kua’nu kû-àn [chii xàà*
 1 11 17 20-26 1 5
 so.that MOD POT.grow soon-3.THING because already
kùni-i [kaxi-i-ñà]]
 17-26 17-26-27
 INCMPL.want-1SG POT.eat-1SG-3.THING
 ‘So that it grows soon because I want to eat it already.’
 SMD-0009-Jardin
- b. *[ntúta’a-ntó [ukuná-nto ve’e]]*
 17-26 12.17-26 27
 INCMPL.should-2PL NEG.POT.open-2PL house
 ‘You (pl.) should not open your house.’ elicited

Of the elements included in the maximal AVVL span, only the S/A arguments in position 26 can have wide scope, as illustrated in examples (24a) and (24b). The minimal domain is thus only one position smaller than the maximal one.

- (24) a. *ntúta’a-ntó kuná-nto ve’e*
 17-26 17-26 27
 INCMPL.should-2PL POT.open-2PL house
 ‘You (pl.) have to open your house.’ SMD-0048-Mayordomia

b.	<i>ntúta'an</i>	<i>kuná-nto</i>	<i>ve'e</i>	
	17	17-26	27	
	INCMPL.should POT.open-2SG house			
	'You (pl.) have to open your house.'			elicited

The second type of subspan repetition that we report concerns syndetic linkage with conjunctions in position 1. We first briefly discuss *ñá* 'that', because there are some additional considerations to take into account. The comparable marker *xá*¹³ in Chalcatongo Mixtec is described as a subordinator optionally marking sentential complements in purpose, result and relative clauses (Macaulay 1996: 153–160). Based on a preliminary survey of our corpus, *ñá* appears to cover the same functions in SMD. Unlike Chalcatongo *xá*, however, in SMD there are several elements of the form *ñá* with different functions and probably different historical origins (see Ventayol-Boada 2021 for an analysis of the origins of third person pronouns and relativizers in SMD). In Table 12, we provide an overview of our current analysis, in which we identify two historical sources for five different *ñá* elements, which can be considered synchronically distinct. In this section, we are only concerned with *ñá* as a marker of clause linkage, which we gloss as complementizer for lack of a better label.

Given that *ñá* is highly generalized and as a linker and seems to have no semantic content, we think it's most reasonable to see it as a shortened form of *ñá'a* 'thing', which has a very general meaning itself. Note also that the two historical sources have different tone patterns (mid-low for 'woman' and low-mid for 'thing'), which might help separate the *ñá* elements from each other. While we cannot provide a detailed analysis of the tonal realizations of these elements yet, we do observe that the *ñá*-marking subordinate clauses always seems to have low tone – confirming that *ñá'a* 'thing' is a probable source.

Further complications arise because *ñá* is also used to modify nouns¹⁴, and it can at times be difficult to tell whether in a given context it introduces a subordinate clause or is modifying a noun. One such example is provided in (25), where the clause introduced by *ñá* could be interpreted as modifying the verb core or the NP 'twenty years' (e.g., 'It has been twenty years in which I didn't travel at all.'). We exclude such examples from the discussion here.

¹³This form is not cognate with *ñá*. For details on the distribution of the two forms in other Mixtec varieties see Hollenbach & Erickson (1995)

¹⁴Whether these constructions should be referred to as relative clauses or nominalizations is an open question outside the scope of this chapter.

Table 12: Current analysis of *ñā* elements and their sources

Element	Probable source
3SG.F dependent pronoun, allomorph of <i>àn</i>	<i>ñā'a</i> 'woman'
CLF.3SG 'classifier' for female beings	<i>ñā'a</i> 'woman'
3SG.THING dependent pronoun, allomorph of <i>àn</i>	<i>ñā'a</i> 'thing'
CLF.THING 'classifier' for things and abstract nouns	<i>ñā'a</i> 'thing'
COMPL marker for subordinate clauses	<i>ñā'a</i> 'thing'

- (25) *ì-xinu oko kùà [ñā kòò xa'a-va-ì níí]*
 12-17 28 1 4 17-18-26 28
 COMPL-run twenty year COMPL NEG.REAL POT.travel-ADD-1SG completely
 'It has been twenty years that I didn't travel at all.' SMD-0059-Padres

The maximal span that can be repeated in *ñā*-linkage is different from that of asyndetic linkage, resulting in a test fracture. It runs from position 4 to 28, illustrated in examples (26a) and (26b). Content question markers, focused constituents and discourse markers cannot appear in *ñā*-linkage. The minimal span excludes S/A-pronouns and temporal adverbials in position 5, since these have wide scope. The additive in position 24, however, can only appear there if preceded by an adverbial. The left edge of the minimal span is thus at position 7.

- (26) a. *ta xàà kivi [ñā chikà-ò kò'ò]*
 1 5 17 1 17-26 27
 and already INCMPL.be.able COMPL POT.put(invisible)-1PL.INCL plate
ta xàà kaxá'an-v-ó
 1 5 17-18-26
 and already POT.eat-ADD-1PL.INCL
 'And already we are able to set out the dishes and eat.'
 SMD-0005-ArrozAmarillo
- b. *ichikàà ini-nà [ñā kòó kùni míi-nà]*
 12-16.17 25-26 1 4 17 26
 COMPL.put(invisible) inside-3PL COMPL NEG.REAL INCMPL.want TOP-3PL
kà'àn-va-na]
 17-18-26
 POT.speak-ADD-3PL
 'They insist on not wanting to speak it.' SMD-0049-Medicinas

SMD also has other types of clause linkage markers in the same position, such as *távà* ‘so that, in order to’, *chii* ‘because’, *soo/suu* ‘but’, *ñàkán* ‘so, for that reason’, etc. A detailed study of each one of these markers lies outside the scope of this study and we thus treat them all together under the label of linkage with conjunctions.

The maximal span identified in this construction differs from that of asyndetic and *ñà*-linkage. It includes all positions except the first position (other connectors cannot co-occur with conjunctions) and the last position, which contains discourse markers. The span thus runs from position 2 to 28, illustrated in examples 27a and 27b. We thus need a further test fracture to account for this. Within this span, the leftmost element that can have wide scope are temporal adverbials in position 5. The additive following them in position 10, however, cannot appear without them, which means that the left edge of the minimal span is at position 7. The right edge is at position 29, since S/A-arguments cannot have wide scope in this construction.

- (27) a. *kù'un-nti* *ka'anxa-nti* *nfoo* *tiempó vitxi* [*chi*
 17-26 17-26 28 1
 POT.go-1PL.EXCL POT.cut-1PL.EXCL INCMPL.be time now because
tava-ña *móso ví]*
 17-26 27 29
 POT.take.out-3PL worker DM
 ‘We will go cut sugar cane around that time because they get the
 workers then.’
 SMD-0053-Carretera
- b. (...) [*chi* *kaxi-v-ó* *ñà* *yó'o*] [*tí*
 1 17-18-26 27 1 5
 because POT.eat-ADD-1PL.INCL CLF.THING here if good
và'a *xáxí-ò-ñà]* [*su*
 1726-27 1 1
 INCMPL.eat-1PL.INCL-3SG.THING but if
 [*tí* *kuntasí-ti-ó]*
 17-23-26 5
 POT.be.closed-again-1PL.INCL then
sààn *kuíta-va-n]*
 17-18-27
 POT.throw.away-ADD-3SG.THING
 ‘(We remove all the feathers from the chicken’s head) because we eat
 this [part of the chicken’s head] here, if we like to eat it, but if it puts
 us off, we will throw it away.’
 SMD-0046-Pollo

Clauses can be coordinated with the general connector *ta* ‘and’ and with the disjunctive marker *an* ‘or’. The maximal interpretation of this diagnostic identifies the whole planar structure, i.e. positions 1 to 29. This is a different span than identified by any of the other constructions, which means we have to fracture this diagnostic further. Two examples of large coordinated spans are provided in (28a) and (28b). As mentioned above, only few elements in SMD can have wide scope. The minimal diagnostic with coordination thus identifies the same span as linkage with conjunctions described above – that is the span from position 7 through 29.

- (28) a. *[ta kòò kuntaa ini-rà ní] [ta ukivi*
 1 4 17 25-26 29 1 12.17
 and NEG.REAL understand inside-3SG.M DM and NEG.POT.can
 ka'an-rà ní]
 17-26 29
 POT.speak-3SG.M DM
 ‘He doesn’t understand and he doesn’t want to speak.’ elicited
- b. *[ntxáa kù'ù-àn] [ta ñama ntxikokò-àn ñuu]?*
 2 17-26 1 2 17-26 28
 where POT.go-3SG.F and when POT.return-3SG.F village
 ‘Where is she going and when will she come back to the village?’
 elicited

We note that all the minimal spans apart from AVVL are identical. This is due to the already mentioned scarcity of forms that can have wide scope. We will consider all the minimal spans as one diagnostic. The reason for this is that they are not independent from each other, since for each subspan repetition construction which has a maximal domain that includes all wide-scope elements, the minimal domain will give the same result. In a sense, it does not tell us anything specific related to the construction. Further research and comparison with other languages is needed to investigate how cases like this one are best treated in the planar-fractal method.

5.5 Spans identified by morphosyntactic domains

We summarize all the morphosyntactic diagnostics and their results in Table 13. Four of the minimal domains converge, but this is because they only identify the verb core, which is rather uninformative. None of the larger spans converge. However, two of the maximal subspan repetition diagnostics differ by only one

position at the left edge. Furthermore, we can see that many of the spans end at the verb core. This is not surprising given that at least some of those diagnostics (like ciscategorical selection and tonal inflection) are targeting “words” (rather than “phrases”).

Table 13: Morphosyntactic diagnostics and their results

Diagnostic	Fracture	MinMax	Left Edge	Right Edge	Size
Non-interrupt.	simplex, <i>sá = sá'a</i>	min	14	20	7
Non-interrupt.	simplex, <i>sá ≠ sá'a</i>	min	11	20	10
Non-interrupt.	complex	max	3	25	23
Non-permut.		max	5	19	15
Ciscat. Selection	<i>chi=ciscat.</i>	min	16	17	2
Ciscat. Selection	<i>chi=transcat.</i>	min	17	17	1
Ciscat. Selection		max	4	23	20
Subspan Rep.	asyndetic	min	12	25	14
Subspan Rep.	asyndetic	max	12	26	15
Subspan Rep.	syndetic	min	7	25	19
Subspan Rep.	<i>ña-link.</i>	max	4	28	25
Subspan Rep.	conj.	max	2	29	28
Subspan Rep.	coordination	max	1	29	29

6 Summary and discussion

We summarize all the diagnostics and results in Figure 1, arranged by span size and colored by module. The span with the highest convergence level with 4 diagnostics is the verb core in position 17. However, no phonological diagnostic targets this span, only morphosyntactic and indeterminate ones. In our view it is not particularly informative for a minimal diagnostic to target the verb core, since this has to be included by definition.

The only other convergence is found with the span 15-17, identified by the maximal tonal inflection diagnostic and the minimal many-to-one deviations diagnostic.

The almost complete absence of convergences in SMD is remarkable but perhaps not completely unexpected, and it lends further support to the view argued in Pike (1945) that there is no sharp distinction between morphology and syntax (or between words and phrases) in Mixtec languages. We do identify convergences on edges: four diagnostics have their left edge at position 12, and four

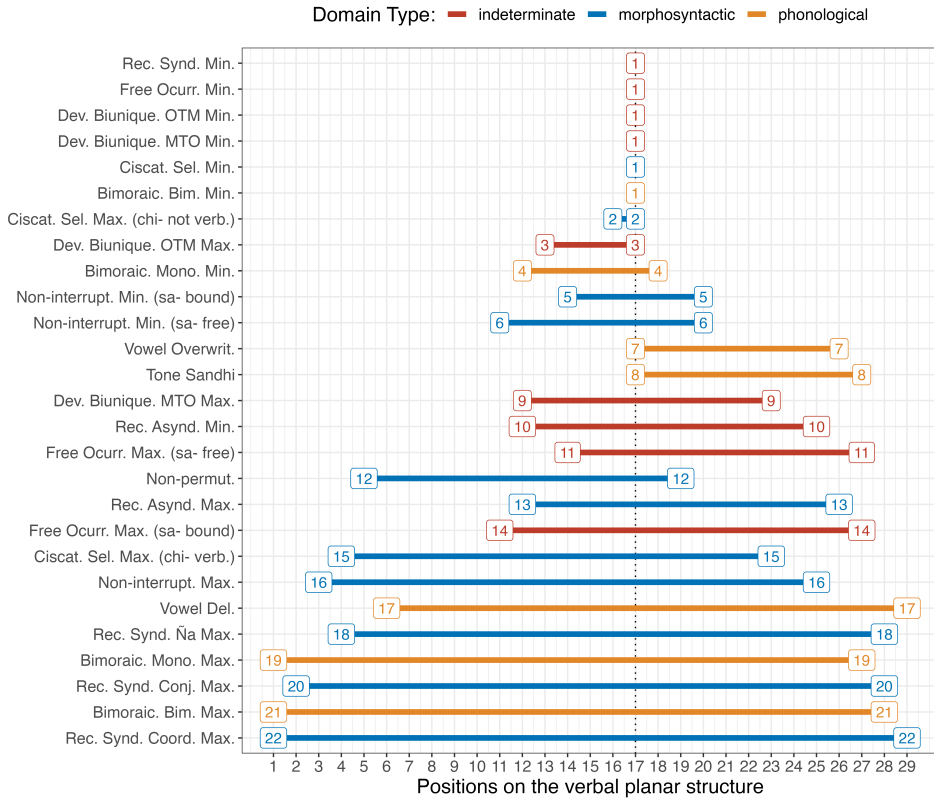


Figure 1: Constituency diagnostics and their results

have their right edge at position 27. This span could be argued to correspond roughly to what traditional analyses would call a “phonological word”, containing only the verb core with its “affixes” and “clitics”. In fact, it corresponds to the orthographic word including hyphens in the practical orthography of SMD as it is currently being used. However, it is not a well motivated level, since no single test, let alone multiple tests, targets this span.

Our results also help explain the different orthographic representations found in materials on Mixtec languages. Some, like Hollenbach & Erickson (2013), tend to write each morpheme separately, while others like Macaulay (1996) write many morphemes together as in one orthographic word, but separated by hyphens. In the practical orthography for SMD, our orthographic word excluding morphemes added with hyphens goes from position 12 to 17, while the orthographic word including hyphenated forms covers maximally from 12 to 27, as

mentioned above. None of these spans are identical to any identified by a diagnostic, but the shorter one roughly corresponds to the minimal bimoraicity constraint (although we write the additive in 18 with a hyphen), and the longer roughly corresponds to maximal free occurrence (even though the monomoraic modals in 11 are represented as separate “words”).

Acknowledgements

We would like to thank all our collaborators in the village of San Martín Duraznos, in particular: Catalina Martínez Ramírez, Pedro Pérez Mendoza, Reina Martínez Rendón, Juana Felipa Martínez, Cristóbal Hernández Martínez, Elvia Mendoza, Regina Rendón, Flor Melina Herón Reyes, Francisca Hernandez, and Eduardo Toledo. We also thank Iní G. Mendoza, Simon L. Peters, and Eric W. Campbell for insightful discussions during the composition of this chapter and Taylor Miller and Enrique Palancar for valuable comments on an earlier version. Our research is funded in part by the Endangered Languages Documentation Fund (Grant number: SG0566) to Sandra Auderset and a National Science Foundation Award (Grant number: 1660355) to the Linguistics Department of the University of California, Santa Barbara.

Abbreviations

ADD	additive	ITER	iterative
ANIM	animals	M	masculine
CLF	classifier	MOD	modality suffix
CMPL	completive	NHON	non-honorific
COP	copula	PROX	proximal
DEM	demonstrative	Q	question particle
DM	discourse marker	REAL	realis
EXCL	exclusive	RECP	reciprocal
F	feminine	THING	things, abstract concepts
HON	honorific		
HUM	human	TOP	topic
INCMP	incompletive	WOOD	wooden things
INTENS	intensifier		

References

- Auderset, Sandra & Carmen Hernández Martínez. 2019. *Hacia un análisis de la flexión verbal en el tù'un Na ñuu sá matxí ntxè'è (mixteco de San Martín Duraznos)*. Talk presented at the 9th Conference on Indigenous Languages of Latin America.
- Auderset, Sandra & Carmen Hernández Martínez. 2022. *Documenting Tù'un Na Ñuu Sá Matxí Ntxè'è, a Mixtec language of Oaxaca, Mexico*. Endangered Languages Archive. <http://hdl.handle.net/2196/f3856126-2e8b-4e98-8709-3b6f0e09f639>.
- Auderset, Sandra, Adam J. R. Tallman & Carmen Hernández Martínez. 2021. *Revisiting clitic phenomena in Mixtec*. Talk presented at the 54th Annual Meeting of the Societas Linguistica Europaea. National & Kapodistrian University of Athens (online).
- Campbell, Eric W. 2017. Otomanguean historical linguistics: Past, present, and prospects for the future. *Language and Linguistics Compass* 11(4). 1–22.
- Campbell, Eric W. 2024. Constituency in Zenzontepec Chatino. In Adam J.R. Tallman, Sandra Auderset & Hiroto Uchihara (eds.), *Constituency and convergence in the Americas*, 367–418. Berlin: Language Science Press. DOI: 10.5281/zenodo.13208554.
- Daly, John. 1973. Tone sandhi and rule ordering in Peñoles Mixtec. *Work Papers of the Summer Institute of Linguistics, University of North Dakota Session* 17(1). 4.
- Feist, Timothy & Enrique L. Palancar. 2016. Tracing the emergence of inflectional tone in Cuicatec. In Enrique L. Palancar & Jean-Léo Léonard (eds.), *Tone and inflection: New facts and new perspectives*, chap. 11, 267–294. Berlin: Walter de Gruyter.
- Gutiérrez, Ambrocio & Hiroto Uchihara. 2024. Words as emergent constituents in Teotitlán del Valle Zapotec. In Adam J.R. Tallman, Sandra Auderset & Hiroto Uchihara (eds.), *Constituency and convergence in the Americas*, 305–365. Berlin: Language Science Press. DOI: 10.5281/zenodo.13208552.
- Hollenbach, Barbara E. 2003. The historical source of an irregular Mixtec tone-sandhi pattern. In Mary Ruth Wise, Thomas N. Headland & Ruth M. Brend (eds.), *Language and life: Essays in memory of Kenneth L. Pike*, 533–550. Dallas: SIL International & University of Texas at Arlington.
- Hollenbach, Barbara E. & Elena Erickson. 1995. Cuatro morfemas funcionales en las lenguas mixtecanas. In Ramón Arzápalo Marín & Yolanda Lastra (eds.), *Vitalidad e influencia de las lenguas indígenas en latinoamérica: II Coloquio Mauricio Swadesh*, 284–293.

- Hollenbach, Barbara E. & Elena Erickson. 2013. *Gramática del mixteco de Magdalena Peñasco (Sa'an Nuu Savi)*. Talpan & Mexico: Instituto Lingüístico de Verano, A.C.
- Hunter, Georgia G & Eunice V. Pike. 1969. The phonology and tone sandhi of Molinos Mixtec. *Linguistics* 7(47). 24–40.
- Josserand, J. Kathryn. 1983. *Mixtec dialect history*. New Orleans: Tulane University. (Doctoral dissertation).
- Kaufman, Terrence. 1988. Otomanguan tense/aspect/mood, voice, and nominalization markers. unpublished monograph.
- Longacre, Robert E. 1957. *Proto-Mixtecan*. Bloomington: Indiana University.
- Macaulay, Monica A. 1987. Cliticization and the morphosyntax of Mixtec. *International Journal of American Linguistics* 53(2). 119–135.
- Macaulay, Monica A. 1993. Argument status and constituent structure in Chalcatongo Mixtec. In *Proceedings of the nineteenth annual meeting of the Berkeley linguistics society*, vol. 19, 73–85.
- Macaulay, Monica A. 1996. *A grammar of Chalcatongo Mixtec*. Berkeley: University of California Press.
- Nakamoto, Shun. 2024. Constituency in Ayautla Mazatec. In Adam J.R. Tallman, Sandra Auderset & Hiroto Uchihara (eds.), *Constituency and convergence in the Americas*, 231–264. Berlin: Language Science Press. DOI: 10.5281/zenodo.13208548.
- Padgett, Erin Padgett. 2017. *Tools for assessing relatedness in understudied language varieties: A survey of Mixtec varieties in Western Oaxaca, Mexico*. Grand Forks: University of North Dakota. (MA thesis).
- Palancar, Enrique L. 2012. Clausal juxtaposition and subordination: New light from Otomi. In Volker Gast & Holger Diessel (eds.), *Clause linkage in cross-linguistic perspective: Data-driven approaches to cross-clausal syntax*, 37–76. Berlin: Mouton de Gruyter.
- Palancar, Enrique L. & Heriberto Avelino. 2019. Inflectional complexity and verb classes in Chichimec. *Amerindia* 41. 323–360.
- Penner, Kevin. 2019. *Prosodic structure in Ixtayutla Mixtec: Evidence for the foot*. Edmonton: University of Alberta. (Doctoral dissertation).
- Pike, Eunice V. & John H. Cowan. 1967. Huajuapán Mixtec phonology and morphophonemics. *Anthropological Linguistics* 9(5). 1–15.
- Pike, Kenneth L. 1944. Analysis of a mixteco text. *International Journal of American Linguistics* 10(4). 113–138.
- Pike, Kenneth L. 1945. A problem in morphology-syntax division. *Acta Linguistica* 5(1). 125–138.

- Pike, Kenneth L. 1948. Tonemic perturbations in Mixteco, with special emphasis on tonomechanical subclasses. In Kenneth L. Pike (ed.), *Tone languages*, 77–94. Ann Arbor: University of Michigan Press.
- Tallman, Adam J. R. 2020. Beyond grammatical and phonological words. *Language and Linguistics Compass* 14(2). 1–14. DOI: 10.1111/lnc3.12364.
- Tallman, Adam J. R. 2021. Constituency and coincidence in Chácobo (Pano). *Studies in Language* 45(2). 321–383. DOI: 10.1075/sl.19025.tal.
- Uchihara, Hiroto & Juana Mendoza Ruíz. 2022. Minimality, maximality and perfect prosodic word in Alcozauca Mixtec. *Natural Language & Linguistic Theory* 40. 599–649.
- Ventayol-Boada, Albert. 2021. From classifiers to subordination: Nominal origins of relativizers and subordinators in Tù'un na Ñuu Sá Mátxíí Ntxè'è (Mixtec). *Amerindia* 43. 177–210.
- Zwicky, Arnold M. 1977. *On clitics*. Bloomington: Indiana University Linguistics Club.