Chapter 6

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

Sandra Auderset, Carmen Hernández Martínez & Albert Ventayol-Boada

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 underingly 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 Duraznos 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 Duraznos, 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, ʣ], but <x, tx, ntx> = [ɕ, tɕ, ʑ]. 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 undergoer (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 *míí* 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>
<thead>
<tr>
<th>Gloss</th>
<th>Dependent</th>
<th>Independent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SG</td>
<td>í</td>
<td>yi'í</td>
</tr>
<tr>
<td>1PL.INCL</td>
<td>ò</td>
<td>mii-ó</td>
</tr>
<tr>
<td>1PL.EXCL</td>
<td>nti</td>
<td>nti'i</td>
</tr>
<tr>
<td>2SG.NHON</td>
<td>un</td>
<td>yó'ò</td>
</tr>
<tr>
<td>2SG.HON</td>
<td>ní</td>
<td>mii-ní</td>
</tr>
<tr>
<td>2PL</td>
<td>ntò</td>
<td>ntó'ó</td>
</tr>
</tbody>
</table>

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 *í* or *ni*, the incomplete 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

<table>
<thead>
<tr>
<th>INCMPL</th>
<th>CMPL</th>
<th>POT</th>
<th>Gloss</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>káchí</td>
<td>ikachi</td>
<td>kachi</td>
<td>‘say’</td>
<td>INCMPL high</td>
</tr>
<tr>
<td>xú’ni</td>
<td>ixu’ni</td>
<td>ku’ni</td>
<td>‘squeeze’</td>
<td>INCMPL high with stem alternation</td>
</tr>
<tr>
<td>ñin</td>
<td>ixòo</td>
<td>koo</td>
<td>‘live, stay’</td>
<td>INCMPL high with suppletion</td>
</tr>
<tr>
<td>núná</td>
<td>inùnà</td>
<td>nùnà</td>
<td>‘be open’</td>
<td>INCMPL with both morae high</td>
</tr>
<tr>
<td>kái</td>
<td>ikái</td>
<td>kái</td>
<td>‘burn’</td>
<td>INCMPL mid</td>
</tr>
<tr>
<td>kuá’àn</td>
<td>ixá’àn</td>
<td>kù’un</td>
<td>‘go’</td>
<td>segmental alternation only</td>
</tr>
</tbody>
</table>

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

\(^1\)In San Martin 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

<table>
<thead>
<tr>
<th>Pos.</th>
<th>Type</th>
<th>Elements</th>
<th>Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>slot connectors, question marker</td>
<td><em>an, ta, távà, chii, etc.</em></td>
<td></td>
</tr>
<tr>
<td>(2)</td>
<td>slot question words</td>
<td><em>nishi, ntxàa, etc.</em></td>
<td></td>
</tr>
<tr>
<td>(3)</td>
<td>slot focus (S/A/P, OBL, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4)</td>
<td>slot realis negation</td>
<td><em>kóó</em></td>
<td></td>
</tr>
<tr>
<td>(5)</td>
<td>zone adverbials</td>
<td><em>xàà, sa’a, và’a, vitxi, etc.</em></td>
<td></td>
</tr>
<tr>
<td>(6)</td>
<td>slot additive</td>
<td><em>va</em></td>
<td></td>
</tr>
<tr>
<td>(7)</td>
<td>slot intensifier</td>
<td><em>kuà’à, tontó</em></td>
<td></td>
</tr>
<tr>
<td>(8)</td>
<td>slot intensifier; again</td>
<td><em>ntxivà’a, yáá; tiki ∼ ti</em></td>
<td></td>
</tr>
<tr>
<td>(9)</td>
<td>slot intensifier; again</td>
<td><em>tiki ∼ ti; ntxivà’a, yáá</em></td>
<td></td>
</tr>
<tr>
<td>(10)</td>
<td>slot additive</td>
<td><em>va</em></td>
<td></td>
</tr>
<tr>
<td>(11)</td>
<td>slot modals</td>
<td><em>ni, ná</em></td>
<td></td>
</tr>
<tr>
<td>(12)</td>
<td>slot completive; potential negation</td>
<td><em>i; u/o ∼ i</em></td>
<td></td>
</tr>
<tr>
<td>(13)</td>
<td>slot causative ‘do’</td>
<td><em>sá</em></td>
<td></td>
</tr>
<tr>
<td>(14)</td>
<td>slot POT; CMPL class markers</td>
<td><em>ku; xi</em></td>
<td></td>
</tr>
<tr>
<td>(15)</td>
<td>slot iterative</td>
<td><em>nta ∼ nti</em></td>
<td></td>
</tr>
<tr>
<td>(16)</td>
<td>slot transitivizer ‘put’</td>
<td><em>chi</em></td>
<td></td>
</tr>
<tr>
<td>(17)</td>
<td>slot verb core</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(18)</td>
<td>slot additive</td>
<td><em>va</em></td>
<td></td>
</tr>
<tr>
<td>(19)</td>
<td>slot reciprocal</td>
<td><em>ta’an</em></td>
<td></td>
</tr>
<tr>
<td>(20)</td>
<td>slot temporal adv.</td>
<td><em>tuun, kìì</em></td>
<td></td>
</tr>
<tr>
<td>(21)</td>
<td>slot intensifier</td>
<td><em>tontó, kuà’à</em></td>
<td></td>
</tr>
<tr>
<td>(22)</td>
<td>slot intensifier; again</td>
<td><em>ntxivà’a, yáá; tiki ∼ ti</em></td>
<td></td>
</tr>
<tr>
<td>(23)</td>
<td>slot intensifier; again</td>
<td><em>tiki ∼ ti; ntxivà’a, yáá</em></td>
<td></td>
</tr>
<tr>
<td>(24)</td>
<td>slot additive</td>
<td><em>va</em></td>
<td></td>
</tr>
<tr>
<td>(25)</td>
<td>slot inside/being</td>
<td><em>ini</em></td>
<td></td>
</tr>
<tr>
<td>(26)</td>
<td>slot S/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(27)</td>
<td>slot P</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(28)</td>
<td>zone OBL, PP, LOC, adv.</td>
<td><em>nì, ví</em></td>
<td></td>
</tr>
<tr>
<td>(29)</td>
<td>slot discourse markers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Constituency in Tú’un Ntá’ví (Mixtec) of San Martín Duraznos

(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 mií-ntí ŋa kò’va  
  like already INCML.know-ADD TOP-1PL.EXCL CLF.THING amount  
  chikàà-ntí  
  POT.put(invisible)-1PL.EXCL  
  ‘so we already know what amount to put in’  
  SMD-0020-Huauzontle
b. sáàn 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 ŋa xàà ntóvà kíí sévóyá-va  
  and CLF.THING already INCML.sprout COP onion-ADD  
  tûyá’à-va xàà ntóvà ntxì’i-va  
  CLF.WOOD.chile-ADD already INCML.sprout POT.finish-ADD  
  tû-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’á* 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 săchuun kuà’à yáá ki’vi-ì
   tomorrow POT.do.work INTENS INTENS sister[f]-1sg
   ‘tomorrow my sister is going to work a lot’
   elicited
d. * itxààn săchuun yáá kuà’à ki’vi-ì
   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 săchuun tiki yáá ki’vi-ì
   tomorrow POT.do.work again INTENS sister[f]-1sg
   ‘tomorrow my sister is going to work a lot again’
   elicited
b. itxààn săchuun yáá tiki ki’vi-ì
   elicited
c. itxààn yáá tiki săchuun ki’vi-ì
   elicited
d. itxààn tiki yáá săchuun ki’vi-ì
   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 transitivizer.

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 transitivizer 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 i. 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 i. Comparison with other Mixtec varieties is not as instructive in this case, because the completive is either marked with tone alone (e.g., San Martin Peras Mixtec), or only displays a marker ni (e.g., Chalcatongo Mixtec, cf. Macaulay 1996: 74–75). We take i to be the basic, unmarked form, since it is more frequent and the one given in elicitation. We suspect that ni might mark deontic modality of past events, and so diachronically it might represent a combination of ná and i. 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 fit 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’; aspeckul 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.  

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 niku’un ini ‘remember’, kutátxi 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 ni ‘affirmative’. 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):

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

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 tún=kuʔu=ro
   2sg NEG=be.sick=2sg
   ‘As for you, you aren’t sick.’ (Macaulay 1996: 106)

b. pero niasu xí xʷá tándaʔá=∅ či tándaʔá=∅ xí pédrú
   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 tún=šãã̀ kṹñṹ=∅
   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 Oto- 
manguean languages (cf. Campbell 2024, Gutiérrez & Uchihara 2024, Nakamoto 
2024 [this volume]), SMD exhibits few tonal processes and few general phono-
logical rules.

We identify three phonological processes that apply to the verb complex, of 
which two concern segments and one concerns tone. These are: vowel overwrit-
ing, bimoraicity, and tone sandhi of dependent pronouns. The first two must be 
fractured into a minimal and maximal domain to render consistent spans, result-
ing in a total of five diagnostics. Throughout this section we also provide IPA tran-
scriptions 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.\textsuperscript{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’), CVʔV (e.g., tù’un ‘word, language’), or VCV 
(e.g., àsi ‘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 monomo-
raic 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

\textsuperscript{5}Long 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-tãː¹³-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-tàshi-ùn-ña?
     [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\textsubscript{i}ʔV\textsubscript{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:

\[(X)CV\textsubscript{i}ʔV\textsubscript{i} + V\textsubscript{j} \rightarrow (X)CV\textsubscript{i}ʔV\textsubscript{j}\]

<table>
<thead>
<tr>
<th>Pronoun</th>
<th>Gloss</th>
<th>Nasality</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>1SG</td>
<td>preserves nasality of base</td>
</tr>
<tr>
<td>èò</td>
<td>1PL.INCL</td>
<td>does not preserve nasality of base</td>
</tr>
<tr>
<td>un</td>
<td>2SG.NHON</td>
<td>always nasal</td>
</tr>
<tr>
<td>àn</td>
<td>3SG.F</td>
<td>always nasal</td>
</tr>
<tr>
<td>an</td>
<td>3SG.THING</td>
<td>always nasal</td>
</tr>
</tbody>
</table>

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 Huajuapan 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.

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

\(^7\)We 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 → yé’-i
   [ʒe⁵ʔe⁵] + [i¹] → [ʒe⁵ʔi¹]
   door + 1SG → ‘my door’

b. inù’ún + i → inù’-in
   [i’nùʔũ¹] + [i¹] → [i’nùʔi¹]
   CMPL.GO HOME + 1SG → ‘I went home’

c. ntxè’i + un → ntxè’-ùn
   [ⁿʣe⁵ʔi¹] + [ũ] → [ⁿʣe⁵ʔũ¹]
   clay + 2SG → ‘your (sg.) clay’

d. ko’nto + i → ko’nto-i
   [ko³ʔⁿdo³] + [i¹] → [ko³ʔⁿdo³i¹]
   bone + 1SG → ‘my bone’

e. xá’ntxá + i → xá’ntxá-i
   [ɕa⁵ʔⁿʥa⁵] + [i¹] → [ɕa⁵ʔⁿʥa⁵i¹]
   INCMPL.CUT + 1SG → ‘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 [ko¹ʔo¹] + un → kò’-ôn [ko¹ʔ-õ¹]
   plate + 2SG.NHON → ‘your plate’

b. ntó’o [ⁿdo⁵ʔo³] + un → ntó’-ôn [ⁿdo⁵ʔ-õ³]
   INCMPL.SUFFER + 2SG.NHON → ‘you are suffering’

Vowel overwriting is observed with vowel-initial dependent pronouns (cf. Table 4) in position 26 following a CVʔV base, and with elements in position 22, such as the intensifier ntxivà’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.  
\[ \text{ta } \text{sààn } \text{i-sùvá’-i} \quad \text{i-sísini-va-i} \]  
\[ [ta³ \text{sà}1 \text{i-su’βa²-i} \quad i¹-si’si’ni¹-βa²-i¹] \]  
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-Jardín  

b.  
\[ \text{lo’o } \text{ntxivá’-i} \quad \text{i-xi’i} \quad \text{nánà-i} \quad \text{tátà-i} \]  
\[ [\text{lo}³\text{o}³ \text{a’βa²-i} \quad i¹-ci’i¹ \text{na’a-na¹-i} \quad \text{ta’ta¹-i¹}] \]  
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.  
\[ \text{ta } \text{sa’a } \text{sa’a} \quad \text{i-nto’-án} \]  
\[ [\text{ta}³ \text{sa²}\text{a³} \quad \text{sa²}\text{a³} \quad i¹-do²-ã³] \]  
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 *i-sùvá’a-i*, 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

<table>
<thead>
<tr>
<th>Group</th>
<th>Generalization</th>
<th>Pronouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>no tone changes</td>
<td>1SG i, 2SG.HON ní</td>
</tr>
<tr>
<td>2</td>
<td>tone copying</td>
<td>2SG.NHON un</td>
</tr>
<tr>
<td>3</td>
<td>L alternating with H</td>
<td>1PL.INCL ò, 3SG.F ãn~ñà, 3.ANIM rì, 3.WOOD dùn</td>
</tr>
<tr>
<td>4</td>
<td>L alternating with M</td>
<td>1PL.EXCL nti, 3SG.M rà, 3PL ñà</td>
</tr>
</tbody>
</table>

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 sa³ʔa³ ka³ʧi³-ʔi³]  
1 5 17-26  
and like.that INCMPL.say-2SG.NHON  
‘And that’s how you say it.’  
SMD-0047-Cena
b. \textit{vitxi} \textit{i-kixà-ùn} \textit{yó’o}  
\text{[βi³tɕι³ i³-κi³ca¹-ũ¹ 3ο³o³]}  
\text{5 12-17-26 28}  
\text{now \ cmpl-arrive-2sg.nhon \ dem.prox}  
\text{‘Now you arrived here (…).’} \text{ SMD-0047-Cena}  

Tone sandhi can also be observed with elements in positions 21 to 25, illustrated by the tone realization of \textit{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 \textit{un} is the same regardless of the tone of the element following it.

\begin{align*}  
\text{(15) a. } & \text{ \textit{su} \ i-kuntàà \ ini-un \ gueritá} \quad \text{but \ cmpl-understand \ inside-2sg.nhon \ white.person} \\
& \text{\text{[su³ i³-κu³̣da:¹¹ i³ni³-ũ³ \ we’ri³ta³]}} \\
& \text{\text{1 12-17 25-26 27}} \\
& \text{‘But you understood \textit{guerita} (light-skinned girl).’} \text{ \quad \text{SMD-0047-Cena}}  
\end{align*}  

\begin{align*}  
\text{b. } & \text{ \textit{ta} \ sààn ni \ i-sàma \ ntxivà’-ùn} \quad \text{and \ then \ mod \ cmpl-change \ intens-2sg.nhon} \\
& \text{\text{[ta³ sà:¹¹ ni¹ i¹-sa¹ma³ \ “dzi¹βa¹ʔ-ũ¹]}} \\
& \text{\text{1 5 11 12-17 22-26}} \\
& \text{‘And so you’ve changed a lot.’} \quad \text{elicited}  
\end{align*}  

\begin{align*}  
\text{c. } & \text{ \textit{su} \ i-kuntàà \ ini-un \ shìtà} \quad \text{but \ cmpl-understand \ inside-2sg.nhon \ tortilla} \\
& \text{\text{[su³ i³-κu³̣da:¹¹ i³ni³-ũ³ \ [i¹ta³]}} \\
& \text{\text{1 12-17 25-26 27}} \\
& \text{‘But you understood \textit{shìtà} (tortilla).’} \quad \text{elicited}  
\end{align*}  

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

\begin{align*}  
\text{(16) a. } & \text{ \textit{ta} \ sààn jààn \ chikàà-ò-ña} \quad \text{and then \ dem.dist \ pot.put(invisible)-1pl.incl-3.thing} \\
& \text{\text{[ta³ sà:¹¹ hà:¹¹ tʃi³ka:¹¹-o³-ña³]}} \\
& \text{\text{1 5 3 17-26-27}} \\
& \text{‘and so we’ll put it in’} \text{ \quad \text{SMD-0020-Huauzontle}}  
\end{align*}
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>
<thead>
<tr>
<th>Diagnostic</th>
<th>Fracture</th>
<th>Left Edge</th>
<th>Right Edge</th>
<th>Size</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bimoraicity min</td>
<td>12</td>
<td>18</td>
<td>7</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>Bimoraicity max</td>
<td>1</td>
<td>27</td>
<td>27</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>Vowel overwriting min</td>
<td>17</td>
<td>26</td>
<td>10</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>Vowel overwriting max</td>
<td>4</td>
<td>28</td>
<td>24</td>
<td>3.2</td>
<td></td>
</tr>
<tr>
<td>Tone sandhi</td>
<td>-</td>
<td>17</td>
<td>27</td>
<td>11</td>
<td>3.3</td>
</tr>
</tbody>
</table>

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.\(^8\)

\[(17)\]
\[
a. \ t\ddot{a}an \\
17 \\
\text{INCMPL quake} \\
\text{‘There’s an earthquake (lit: [it] is quaking).’} \\
\text{elicited}
\]

\[
b. \ k\ddot{a}\ddot{a}n \\
17 \\
\text{POT speak} \\
\text{‘Speak!’} \\
\text{elicited}
\]

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 \(s\ddot{a}\). In the following, we will illustrate the issue and present the competing results.

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

If the causative marker \(s\ddot{a}\) is taken to be a shortened form of the verb \(s\ddot{a}'a\) – 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 \(s\ddot{a}\) is taken

---

\(^8\)Note 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.

\(^9\)Macaulay (1987) calls this “fast speech reduction”.

xxi
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-ntxítxà-ntò-an]  
         1 5  13-17-26-27
         and there CAUS-melt-2PL-3.THING
         ‘And then they dissolve it (…)’          SMD-0033-Espiritus
    b. kuíì-rì chii saa [i-kintxaa-va-i-ri]  
         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
         chikàà-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 $i$ or $ni$, 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>
<thead>
<tr>
<th>INCMPL</th>
<th>CMPL</th>
<th>POT</th>
<th>Gloss</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>tívi</td>
<td>itivi</td>
<td>tivi</td>
<td>‘wake up’</td>
<td>INCMPL high</td>
</tr>
<tr>
<td>xú’ní</td>
<td>ixu’ní</td>
<td>ku’ní</td>
<td>‘squeeze’</td>
<td>INCMPL high with stem alternation</td>
</tr>
<tr>
<td>iin</td>
<td>ixòo</td>
<td>koo</td>
<td>‘live, stay’</td>
<td>INCMPL high with suppletion</td>
</tr>
<tr>
<td>núná</td>
<td>inúnà</td>
<td>núnà</td>
<td>‘be open’</td>
<td>INCMPL with both morae high</td>
</tr>
<tr>
<td>kái</td>
<td>ikái</td>
<td>kái</td>
<td>‘burn’</td>
<td>INCMPL mid</td>
</tr>
<tr>
<td>kuà’an</td>
<td>ixà’an</td>
<td>kù’àn</td>
<td>‘go’</td>
<td>segmental alternation only</td>
</tr>
</tbody>
</table>

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 $nți/nta$, but not on the inflectional class markers $ku$ and $xi$, 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 inflection\textsuperscript{10}

<table>
<thead>
<tr>
<th>INCmpl</th>
<th>CMPL</th>
<th>POT</th>
<th>Meaning</th>
<th>Morphemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>chí-ntoo</td>
<td>i-chí-ntoo</td>
<td>chí-ntoo</td>
<td>‘put down, stack’</td>
<td>caus-be</td>
</tr>
<tr>
<td>16-17</td>
<td>12-16-17</td>
<td>16-17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ntá-koto</td>
<td>i-nta-koto</td>
<td>nta-koto</td>
<td>‘mend’</td>
<td>iter-take.care</td>
</tr>
<tr>
<td>15-17</td>
<td>12-15-17</td>
<td>15-17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ntá-chi-kàà</td>
<td>i-nta-chi-kàà</td>
<td>nta-chi-kàà</td>
<td>‘put again’</td>
<td>iter-caus-put(invis.)</td>
</tr>
<tr>
<td>15-16-17</td>
<td>12-15-16-17</td>
<td>15-16-17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sá-keta</td>
<td>i-sá-keta</td>
<td>sá-keta</td>
<td>‘finish sth.’</td>
<td>caus-put</td>
</tr>
<tr>
<td>13-17</td>
<td>12-13-17</td>
<td>13-17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>naní</td>
<td>i-xi-naní</td>
<td>ku-naní</td>
<td>‘be called, be named’</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>12-14-17</td>
<td>14-17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(19) a. u-ka’ntxa-ti-un shini ntá’-ùn
12-17-23-26 27
NEG.POT-cut-again-2SG.NHON head hand-2SG.NHON
‘Don’t cut your finger again!’ elicited

b. i-ka’ntxa-ti-un shini ntá’-ùn
12-17-23-26 27
NEG.POT-cut-again-2SG.NHON head hand-2SG.NHON
‘Don’t cut your finger again!’ elicited

c. i-ka’ntxa tiki-un shini ntá’-ùn
12-17 23-26 27
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 incompletive 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).
6 Constituency in Tù’un Ntà’ví (Mixtec) of San Martín Duraznos

(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-i 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. mìi-nì mâmà sánto’o-nì mìi-nì

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 kùní-i sánto’o-nì

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 incompletive
(one form – multiple meanings)

<table>
<thead>
<tr>
<th>INCMPL</th>
<th>CMPL</th>
<th>POT</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>sá’a</td>
<td>isá’a</td>
<td>sá’a</td>
<td>‘do, make’</td>
</tr>
<tr>
<td>nù’ùn</td>
<td>inù’ùn</td>
<td>nù’ùn</td>
<td>‘leave, go home’</td>
</tr>
<tr>
<td>xàà</td>
<td>ixàà</td>
<td>xàà</td>
<td>‘rot, decompose’</td>
</tr>
<tr>
<td>sá-nta-kàà</td>
<td>i-sá-nta-kàà</td>
<td>sá-nta-kàà</td>
<td>‘spread out’</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Diagnostic</th>
<th>Fracture</th>
<th>LeftEdge</th>
<th>RightEdge</th>
<th>Size</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free occurrence min</td>
<td>17</td>
<td>17</td>
<td>1</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>Free occurrence max</td>
<td>sá = sá’a</td>
<td>14</td>
<td>27</td>
<td>14</td>
<td>4.1</td>
</tr>
<tr>
<td>Free occurrence max</td>
<td>sá ≠ sá’a</td>
<td>11</td>
<td>27</td>
<td>17</td>
<td>4.1</td>
</tr>
<tr>
<td>Dev. biunique. min</td>
<td>one-to-many</td>
<td>17</td>
<td>17</td>
<td>1</td>
<td>4.2</td>
</tr>
<tr>
<td>Dev. biunique. max</td>
<td>one-to-many</td>
<td>13</td>
<td>17</td>
<td>5</td>
<td>4.2</td>
</tr>
<tr>
<td>Dev. biunique. min</td>
<td>many-to-one</td>
<td>15</td>
<td>17</td>
<td>3</td>
<td>4.2</td>
</tr>
<tr>
<td>Dev. biunique. max</td>
<td>many-to-one</td>
<td>12</td>
<td>23</td>
<td>12</td>
<td>4.2</td>
</tr>
</tbody>
</table>

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 Ruiz 2022).\footnote{It 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.

\[(21) \quad \begin{align*}
a. & \quad \text{taa } \text{ikán } xàà \quad \text{sa’-va } \quad \text{ntáa } \text{mii } \text{iti-nà} \\
& \quad \begin{array}{cccc}
1 & 3 & 5 & 17 & 26 \\
\text{and DEM.PROX} & \text{already} & \text{like.that-ADD} & \text{be} & \text{TOP cornfield-3PL.HUM} \\
\text{ikán} & \\
\text{DEM.PROX} & \\
\text{‘... And here, their cornfield is already like this here.’ SMD-0057-Tierra}
\end{array}
\end{align*}
\]

\[(21) \quad \begin{align*}
b. & \quad \text{taa } \text{ikán } \quad \text{sa’} \quad \text{xàà-va } \quad \text{ntáa } \text{mii } \text{iti-nà} \\
& \quad \begin{array}{cccc}
1 & 3 & 5 & 17 & 26 \\
\text{and DEM.PROX} & \text{already} & \text{like.that-ADD} & \text{be} & \text{TOP cornfield-3PL.HUM} \\
\text{ikán} & \\
\text{DEM.PROX} & \\
\text{‘...And here, their cornfield is already like this here.’ elicited}
\end{array}
\end{align*}
\]
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)  

<table>
<thead>
<tr>
<th>a.</th>
<th>chíntxeé  <em>ta’an</em> tóntó  ntxivà’a-na</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>INCMLHELP RECP INTENS INTENS-3PL.HUM</td>
</tr>
<tr>
<td></td>
<td>‘They really help each other a lot.’</td>
</tr>
<tr>
<td>b.</td>
<td><em>chíntxeé</em> tóntó <em>ta’an</em> ntxivà’a-na</td>
</tr>
<tr>
<td>c.</td>
<td><em>chíntxeé</em> tóntó ntxivà’a <em>ta’an-na</em></td>
</tr>
<tr>
<td>d.</td>
<td><em>ta’an</em> chíntxeé tóntó ntxivà’a-na</td>
</tr>
</tbody>
</table>

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.\(^\text{12}\) 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., *ishi* ‘hairy’ from *ishì* ‘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

\(^{12}\)We would like to thank Eric W. Campbell for pointing this out to us.
6 Constituency in Tù’un Ntá’vi (Mixtec) of San Martín Duraznos

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

<table>
<thead>
<tr>
<th>Form</th>
<th>Gloss</th>
<th>Base</th>
<th>Word class of base</th>
</tr>
</thead>
<tbody>
<tr>
<td>chiñú’un</td>
<td>worship sb.</td>
<td>ñu’un ‘fire’</td>
<td>noun</td>
</tr>
<tr>
<td>chiko’vá</td>
<td>measure sth.</td>
<td>kò’va ‘size, amount’</td>
<td>noun</td>
</tr>
<tr>
<td>chikanii</td>
<td>stop sth.</td>
<td>kanii ‘hit’</td>
<td>verb</td>
</tr>
</tbody>
</table>

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 Otomanguean 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 kii-àn [chii xià 1 11 17 20–26 1 5]
   so.that MOD POT.grow soon-3.THING because already
   kúmi-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. ntúta’an kuná-nto ve’e
17 17-26 27
INCMPL.should POT.open-2SG house
‘You (pl.) have to open your house.’
edited

The second type of subspan repetition that we report concerns syndetic linkage with conjunctions in position 1. We first briefly discuss ña ‘that’, because there are some additional considerations to take into account. The comparable marker xa\textsuperscript{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, ña appears to cover the same functions in SMD. Unlike Chalcatongo xa, however, in SMD there are several elements of the form ña 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 ña elements, which can be considered synchronically distinct. In this section, we are only concerned with ña as a marker of clause linkage, which we gloss as complementizer for lack of a better label.

Given that ña 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 ña’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 ña elements from each other. While we cannot provide a detailed analysis of the tonal realizations of these elements yet, we do observe that the ña-marking subordinate clauses always seem to have low tone – confirming that ña’a ‘thing’ is a probable source.

Further complications arise because ña is also used to modify nouns\textsuperscript{14}, 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 ña 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.

\textsuperscript{13}This form is not cognate with ña. For details on the distribution of the two forms in other Mixtec varieties see Hollenbach & Erickson (1995)

\textsuperscript{14}Whether 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 ña elements and their sources

<table>
<thead>
<tr>
<th>Element</th>
<th>Probable source</th>
</tr>
</thead>
<tbody>
<tr>
<td>3SG.F dependent pronoun, allomorph of ân</td>
<td>ña’á ‘woman’</td>
</tr>
<tr>
<td>clf.3SG ‘classifier’ for female beings</td>
<td>ña’á ‘woman’</td>
</tr>
<tr>
<td>3SG.THING dependent pronoun, allomorph of ân</td>
<td>ñà’a ‘thing’</td>
</tr>
<tr>
<td>clf.THING ‘classifier’ for things and abstract nouns</td>
<td>ñà’a ‘thing’</td>
</tr>
<tr>
<td>compl marker for subordinate clauses</td>
<td>ñà’a ‘thing’</td>
</tr>
</tbody>
</table>

(25) i-xinu oko kuìà [ñà kòó xa’á-va-i nií] 12-17 28 1 4 17-18-26 28  
CMPL-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 ña-linkage is different from that  
of asyndetic linkage, resulting in a test fracture. It runs from position 4 to 28,  
ilustrated in examples (26a) and (26b). Content question markers, focused  
constituents and discourse markers cannot appear in ña-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 [ñà 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 kaxà’an-v-ò 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 [ñà kòó kùni mìi-nà] 12-16.17 25-26 1 4 17 26  
CMPL.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 tāvà 'so that, in order to', chii ‘because’, soó/suu ‘but’, ŋàkán ‘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 ŋà-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 n’too tiemṕo vitxi [chi
17-26 17-26 28 1
POT.go-1PL.EXCL POT.cut-1PL.EXCL INCMPL.be time now because
tava-ŋa möso vi] 17-26 27 29
POT.take.out-3PL worker DM

'If we will go cut sugar cane around that time because they get the workers then.'

b. (...) [chi kaxí-v-ó ŋà yó’ó] [tí
1 17-18-26 27 1 5
because POT.eat-ADD-1PL.INCL CLF.THING here if good
và’a xáxí-ó-ŋà] [su
1726-27 1 1
INCMPL.eat-1PL.INCL-3SG.THING but if
[tí kuntasí-ti-ô]
17-23-26 5
POT.be.closed-again-1PL.INCL then
sáàn kuita-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.'
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. \[ta \text{kòò} \text{kuntaa} \text{ini-rà} \text{ní}] [ta \text{ukivi} \\
1 \hspace{1cm} 4 \hspace{1cm} 17 \hspace{1cm} 25-26 \hspace{1cm} 29 \hspace{1cm} 1 \hspace{1cm} 12.17 \]
and \text{NEG.REAL} understand inside-\text{3SG.M DM} and \text{NEG.POT.can} \\
\text{ka’an-rà} \hspace{1cm} \text{ní}] \\
17-26 \hspace{1cm} 29 \\
\text{POT.speak-3SG.M DM} \\
‘He doesn’t understand and he doesn’t want to speak.’ elicited

b. \[ntxáa \text{kù’ù-àn}] [ta \text{ñama} ntxikkò-àn \text{ñuu]}? \\
2 \hspace{1cm} 17-26 \hspace{1cm} 1 \hspace{1cm} 2 \hspace{1cm} 17-26 \hspace{1cm} 28 \\
where \text{POT.go-3SG.F} and when \text{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 ciscategorial selection and tonal inflection) are targeting “words” (rather than “phrases”).

Table 13: Morphosyntactic diagnostics and their results

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

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
Sandra Auderset, Carmen Hernández Martínez & Albert Ventayol-Boada

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

xxxvi
mentioned above. None of these spans are identical to any identified by a di-
agnostic, 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 Du-
raznos, 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 Inì 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 Foun-
dation Award (Grant number: 1660355) to the Linguistics Department of the Uni-
versity of California, Santa Barbara.

Abbreviations

1 first person  HON  honorific
2 second person HUM  human
3 third person  INCL  inclusive
ADD additive  INCMPL  incompletive
ANIM animals  INTENS  intensifier
CAUS causative  ITER  iterative
CLF classifier  M  masculine
CMPL completive  MOD  modality
COMPL completive  NEG  negative
COP copula  NHON  non-honorific
DEM demonstrative  PL  plural
DIST distal  POT  potential
DM discourse marker  PROX  proximal
EXCL exclusive  Q  question particle
F feminine  REAL  realis
FOC focus  RECP  reciprocal
References


Hollenbach, Barbara E. & Elena Erickson. 2013. Gramática del mixteco de Magdalena Peñasco (Sa’an Ñuu Savi). Tlalpan & Mexico: Instituto Lingüístico de Verano, A.C.


