

# Chapter 12

## Constituency in Mëbêngôkre independent clauses

Andrés Pablo Salanova

Université d'Ottawa

This paper presents a sketch in templatic form of the morphology and syntax of Mëbêngôkre, a Jê language from central Brazil, and evaluates various diagnostics for wordhood and constituency in the language. Diagnostics for constituent structure converge to identify a verb stem approximately coinciding with the word boundaries of my earlier practice, though its left edge is somewhat diffuse. Factors other than constituent structure, such as argumenthood and idiomaticity, are claimed to influence how elements in the template behave with respect to various diagnostics, and this effect is particularly clear with the elements that immediately precede the verb stem.

### 1 Introduction

This chapter describes independent verbal clauses in Mëbêngôkre, a Northern Jê language spoken in Central Brazil by approximately 13,000 people, divided among the Xikrin and the Kayapó nations. The methodology proposed in Tallman (2021) is applied to the planar structure of these clauses to establish whether the various diagnostics for constituency converge to a clearly discernible “word”.

The Jê languages of eastern South America are often described as being of the isolating type. This is not an adequate characterization. As will be evident in the following pages, the morphology of Mëbêngôkre, quite typical of what one finds in other Jê languages, is rather complex. One may nevertheless concede that the morphology of Mëbêngôkre contains relatively few productive concatenative affixes, and falls instead mainly into two classes: on the one hand, there is highly fusional morphology very close to the verb root, sometimes displaying limited

productivity; on the other hand, there are a number of “particles” exhibiting little morpho-phonological interaction with their hosts other than being phrased with them prosodically.

To my knowledge, the question of wordhood has not been explicitly broached in previous work on Jê languages. The practical considerations surrounding the creation of writing systems have forced certain decisions that may or may not be based on consistent application of phonological or morpho-syntactic criteria. The writing system of Mẽbêngôkre was devised by SIL missionaries in the 1970’s (Stout & Thomson 1974) and suffered a series of revisions before being stabilized at the time of the publication of the New Testament in 1996. The placement of word boundaries in this standard is not consistent from a phonological or morphological point of view, and seems quite counter-intuitive to most literate native speakers. In this chapter, I find that in independent verbal clauses the diagnostics mostly converge around a verb stem, with specific points in the planar structure where boundaries are less well-defined. Before I describe the issues, I offer an overview of the Mẽbêngôkre language and speakers and of the main facts of Mẽbêngôkre phonology and morphosyntax.

## **2 Mẽbêngôkre speakers and the author’s fieldwork**

Mẽbêngôkre belongs to the Northern branch of the Jê language family that in the period that preceded the European conquest occupied most of the Central Brazilian Plateau, as well as the interior regions of Southern Brazil. Jê is the main branch of the larger Macro-Jê family, which includes small families both to the east of Jê proper (Maxakalian, Kamakanan, Borum), to the west (Jabutian, Rikbaktsa, Chiquitanoan) and in pockets within Jê territory (Ofayé, Karajá); for classification, see Nikulin (2020). All Macro-Jê languages except for Chiquitano, its most distant member, are spoken entirely within the present-day borders of Brazil, in an area south and west of the Amazon and Madeira, and just shy of the Atlantic coast to the east. The Jê are known for living in large circular villages with a central plaza used for political debate and ritual, and for their efficient military organization, which kept Brazilian colonization at bay until the mid-20th century in many parts of the region.

The Mẽbêngôkre (‘those that are *ngôkre*’, i.e., ‘cavity of the water’, an opaque reference never satisfactorily explained in the literature) or Ngôkrejê, currently live mostly within the Xingu basin, with a few villages in the basin of western tributaries of the Araguaia. They most likely hail from farther east, though they were known to raid as far west as the Tapajós during the 20th century. All contemporary Mẽbêngôkre share this self-designation, yet for most of their recorded

history they did not see themselves as a unified polity, but rather lived as relatively small autonomous communities showing varying degrees of mutual antagonism; for a detailed history, see Verswijver (2018). Linguistically, their subdivisions are of little consequence except for the oldest of them, that between the Xikrin and the Kayapó, but even in this case the differences are relatively minor. The Kayapó account for at least four fifths of the total Mëbêngôkre population, and live in over two dozen villages in a mostly continuous stretch of land in southern Pará and northern Mato Grosso. The Xikrin live in around ten communities in two separate territories further north. The Kayapó's major subdivisions are Mëkräknöti, which includes Northern, Central and Southern Mëkräknöti – the latter also known as Mëtyktire – to the west of the Xingu river, and Gorotire or Djudjêtykti, which includes Kubëkräkênh and Kôkrajmôrô, to the east. This division is quite clear in sociological and political terms – the groups were contacted more than twenty years apart by different colonization fronts, and their post-contact history differs significantly – but irrelevant linguistically.<sup>1</sup> A further branch of the Kayapó, the Irã'ãmrãjre, living farther east in the savannas of the Araguaia, became extinct in the first half of the 20th century. Their language is documented in Sala (1920).

The Jê languages closest to Mëbêngôkre are Apinajé to the east and Kîsêdjê and Kajkwakhrattxi (also known as Suyá and Tapayúna, respectively) to the west and south. After contact, the Kajkwakhrattxi were relocated to Mëbêngôkre-Mëtyktire villages and maintain close ties with them. On the whole, however, peaceful contacts with neighboring groups only occurred exceptionally during the known history of the Mëbêngôkre. One case is the alliance between the Mëbêngôkre-Xikrin and the Xambioá (Karajá), later broken. Another case of an initially friendly contact that later soured is that between the Kayapó-Mëkräknöti and the Yudjá, also known as Juruna (see Verswijver 1982). Salanova & Nikulin (2020) discuss the linguistic effects of these contacts.

I conducted my first visit to the Mëbêngôkre-Xikrin in early 1996, and to the Mëbêngôkre-Kayapó later that same year, and have returned to the field almost yearly except for a 4-year period during my doctoral studies and in the peak years

---

<sup>1</sup>Nowadays, and perhaps throughout the recent history of the Mëbêngôkre, the identification of dialect differences is complicated by a number of recent migrations and influences. To cite only one example, many Gorotire Kayapó migrated to Xikrin communities starting in the late 1980's or early 1990's, and fulfill important roles due to their greater experience with non-indigenous society. The ascendancy of Kayapó in the Xikrin communities of Cateté makes identification of dialect differences difficult without a systematic survey, which I never carried out. Any differences between the varieties that are identified in my work are impressionistic, though confirmed by my main consultants, who are familiar with both varieties.

of the Covid-19 pandemic. Villages where I've recorded language data include the two main Xikrin villages in the Cateté Indigenous land, the Mëkrâknôti villages west of the Xingu (Baú, Kubêkàkre, Mëtyktire, Kapôt), the Gorotire village of Motukôre, and the small Gorotire settlement at Las Casas. I have also conducted linguistic research in the Brazilian city of Redenção, where an increasing number of Kayapó-Gorotire have been taking residence. I have recorded texts from both Kayapó and Xikrin speakers, going from the generation born just before contact (now over 80 years old) to people more or less contemporaneous with me (i.e., now in their 40's and 50's). The more careful linguistic elicitation – as well as the annotation and correction of translations and transcriptions – has been carried out since 2007 primarily with two individuals, both of them Kayapó-Gorotire by birth, but with strong Xikrin links.

In part because of the inability to return to the field during the period that the chapter was written, I've used very little elicitation to back the findings in this chapter. The methodology used can be described as corpus-assisted: my knowledge of the language, decanted over 25 years of research, is sufficient to produce hypotheses that can be easily tested by consulting the corpus of texts. Particularly useful for this chapter given its size was consultation of the New Testament translation mentioned above, very idiosyncratic in terms of lexicon but fairly reliable for questions such as order of particles. This corpus-assisted methodology yields easily verifiable results; however, with it certain questions can only be pushed so far: the template will rarely be densely filled in texts, and examples where more than one particle occurs in a certain zone are hard to come by, even if I know such configuration to be possible. The correct interpretation of the examples was verified during a trip to the field in the summer of 2022, when most of the chapter was already written.

### **3 Brief introduction to Mëbêngôkre syntax**

Like most other Jê languages, Mëbêngôkre is consistently head-final and predominantly head-marking. The one clear inflectional category, which cross-cuts the noun, verb and adposition classes, is person. Essentially a single series of person indices ("absolute", not given a case label in my glosses) exists for various functions (subject of non-finite or nominal intransitives, object of transitives, object of postposition, complement of nouns), though a small number of postpositions and a subset of transitive verbs govern an accusative series which is distinct from the absolute in the third person only (the verbs in question govern the accusative only when finite). Contrary to what happens in many other languages

of the region, Mëbêngôkre does not exhibit person hierarchy effects in its inflection – which is essentially tied to specific grammatical functions –, with a single exception: eccentric agreement with the subject in the case of a second person A and a third person P, in accusative-governing verbs. The following examples show this inflection in words of various categories.<sup>2</sup>

- (1) a. (finite verb)  
*ga i-pumũ, ba a-pumũ, ba-pumũ, gu omũ*  
 2NOM 1-see.FIN 1.NOM 2-see.FIN 1INCL-see.FIN 1INCL.NOM 3.see.FIN  
 ‘you see me, I see you, he/she/it sees us (incl.), we (incl.) see him/her/it’
- b. (nominal predicate)  
*i-kanê, a-kanê, ba-kanê, kanê*  
 1-sick 2-sick 1INCL-sick 3.sick  
 ‘I’m sick, you’re sick, we (incl.) are sick, he/she/it is sick’
- c. (adpositions)  
*i-mã, a-mã, ba-mã, ku-mã*  
 1-to 2-to 1INCL-to 3ACC-to  
 ‘to me, to you, to us (incl.) to him/her/its’
- d. (relational noun)  
*i-nã, a-nã, ba-nã, nã*  
 1-mother 2-mother 1INCL-mother 3.mother  
 ‘my mother, your mother, our (incl.) mother, his/her/its mother’

Arguably, the obligatoriness of person inflection on a lexical root reflects a deeper property, which cross-cuts the lexical categories, that I have called *relationality* in previous work (e.g., Salanova 2007 speaking about nouns; the notion is laid out systematically in Salanova & Nikulin in press). Relational roots have to be preceded either by person inflection (which is referential) or by a noun phrase, which I call the *complement*. In practice, this complement is often morphologically unexpressed, and the relationality of a root may only be seen in the rest of the paradigm and in the fact that anaphoric reference to some entity is always understood in the absence of an overt marker. The examples in (1) may be supplemented with the following, which show the complementarity between person indices and noun phrases:

<sup>2</sup>Example (1a) shows some of the morphophonology associated with third person inflection in some stems. For discussion, see Salanova (2011).

- (2) a. *prõ pumũ, omũ*  
 3.wife see.FIN 3.see.FIN  
 ‘saw his wife, saw him/her’  
 b. *prõ mã, ku-mã*  
 3.wife to 3ACC-to  
 ‘to his wife, to him/her’  
 c. *tep kanê, kanê*  
 fish poison 3.poison  
 ‘fish poison, its poison’

Non-relational roots, on the other hand, may never be directly preceded by inflection or by a nominal complement, and there is no implicit reference to a participant in that function. Devices such as adpositions or derivational mechanisms allow for the expression of, e.g., the possessor of an alienably possessed noun.

Number is independent of person except in the first person inclusive. Particles for paucal and plural exist which are placed immediately before the bound person indices or immediately after a nominative pronoun. In the case of first person inclusive, the number particles are bound with the nominative pronoun, making multiple exponence of person obligatory whenever a number index modifies a bound person index:

- (3) *ar i-nhõ kikre, gwaj ba-nhõ kikre*  
 PAUC 1-POSS house 1INCL.NOM.PAUC 1INCL-POSS house  
 ‘our (excl.) house, our (incl.) house’

The order of elements within the Mëbêngôkre clause is fairly rigid, though there are specific places, discussed in §5, where this rigidity can be overcome. The following diagram (from Salanova 2007) is broadly true for declarative sentences expressing categorical judgments (see discussion of example (12)), and forms the basis of the planar structure developed below. For convenience, we have added the numbers of positions in this planar structure to our earlier template.

- (4) | ... *left field* ... | ..... *middle field* ..... | *right field* |  
 Focus Tense Subject Aspect Objects Predicate Modifiers  
 1-2 3 4 5-8 9-12 13-16 17-22  
*djwỳ ne ba arỳm amim àr o=nhỹ*  
*farinha NFUT 1.NOM already self.for 3.roast.NFIN PROG*  
 ‘I’m already roasting *farinha* for myself.”

This template is insufficient in a number of situations, all of which are characterized by a difficulty in distinguishing simple from complex structures. One common case of this is when noun phrases are quantified or modified. If this happens in the focus position, the complex noun phrases behave like any other noun phrase, (5a); however, if they appear in object position, (5b), a same subject conjunction (ss) or the homonymous nonfuture marker (usually restricted to appear in second position after a single constituent) appears between it and the verb, and the verb takes a referential object index, as if its complement was recovered anaphorically from an earlier mention.<sup>3</sup>

- (5) a. *Tep amẽ=n ikjê=kêt ne ba ku-by.*  
 fish two=SS 3.partner=NEG NFUT 1.NOM 3ACC-grab.FIN  
 ‘I grabbed *three fish*.’
- b. *Ba tep amẽ=n ikjê=kêt ne ku-by.*  
 1.NOM fish two=SS 3.partner=NEG SS/NFUT 3ACC-grab.FIN  
 ‘I grabbed three fish.’

The latter construction is *sui generis*. It has what is plausibly a sentential conjunction in the middle of it, but it would still make little sense to consider it as consisting of two conjoined sentences, since the first person nominative subject would be hanging, without any semantic role tying it to the first conjunct.

Another case where the template seems to break involves sentential modifiers. The post-verbal position is generally reserved for elements that govern the non-finite or nominal form of the verb, something which is formally identical to subordinating it. There is a continuum between highly grammaticalized modifiers of this sort such as negation and cases where the construction is clearly biclausal. The following examples show a few points in the continuum. The subordinating element is in bold.

- (6) a. *Arým krĩ mǎ i-tẽm **kadjy**.*  
 already village to 1-go.NFIN for  
 ‘I’m about to go to the village.’
- b. *I-je ’ã akre **kadjy** ne me i-mǎ ku-ngã.*  
 1-ERG on 3.count for NFUT PL 1-to 3ACC-give.FIN  
 ‘They gave it to me so that I would count it.’

<sup>3</sup>Note that the number three is itself composed of two parts conjoined by a sentential conjunction: ‘[they are] two and [there is one with] no partner.’

- (7) a. *Mỳj dja ba i-je arēnh o?*  
what FUT 1.NOM 1-ERG 3.say.NFIN with  
'How should I tell it?'
- b. *Ba arỳm arēnh o nhỹ.*  
1.NOM already 3.say.NFIN with sit.FIN  
'I'm already telling it (sitting down).'
- c. *Kàj bê àmra o ku-m kabēn.*  
loud at 3.cry with 3ACC-to 3.speak  
'He spoke to him crying loudly.'

It might seem straightforward to say that only the (a) sentences in (6) and (7) involve grammaticalized modifiers, while all others should properly be called complex. However, the construction in (7b) has likely also been grammaticalized as a progressive, a construction where the postural verb is not interpreted literally. Other *prima facie* complex constructions seem grammaticalized to introduce semantic modifiers of the situation rather than new situations; in fact, there is no way of modifying duration in a clause, for example, other than what is seen in the following example, taken from the New Testament translation.

- (8) *Ar tyk ja pumũnh o kumex 'ã akati amãnhkrut ne*  
PAUC 3.dead this see.NFIN with much on day two and  
*ikjê=kêt ne 'ã mēdjja.*  
3.partner=NEG and on midday  
'They watched over the dead for three and a half days'. Literally: 'there were [passed] three days and [it was already] midday on [the extent of] them looking over the dead ones.' (Apocalypse 11:9)

My decision in this regard is pragmatic rather than principled. In previous work (Salanova 2008), I have insisted that all such cases are usefully considered to be complex. For the purposes of this chapter, however, it is worth recognizing the fact that a small class of governing post-verbal elements occur significantly more often than others in *prima facie* simple clauses, and their order is relatively fixed. These elements include a handful of postpositions which do double duty as aspectual modifiers, posture verbs that function in main clauses as progressive auxiliaries, negation, and a few manner modifiers recruited from the class of relational nouns. In such cases, even though a complex structure could be argued for, I've opted to consider the elements as part of a single independent clause template.



The same point could be made with certain coordinated structures, of which the following is an example.

- (9) *Ne kam ku-m kabën tën wadjà.*  
 ss then 3ACC-to 3.speak go.ss enter  
 ‘And he walked speaking to him as he entered.’ Literally: ‘talked to him and went and entered.’ (Acts 10:27)

In this chapter, I treat each of the strings between same-subject conjunctions (ss) as separate clauses, even if their semantic cohesion might suggest otherwise. The matter needs to be investigated further, to determine in particular whether the choice of verbs that are coordinated in this way is free or has crystallized into a small number of fixed senses.

## 4 Mëbêngôkre morphology, as traditionally understood

In the introduction I said that most morphology in Mëbêngôkre could be classified as either (1) elements very tightly bound with the verb root, often semantically idiosyncratic, non-concatenative or not fully productive, or (2) optional elements that have some degree of freedom in their ordering, and that seem bound to their hosts only as a consequence of being prosodically deficient. My implicit approach in previous work was to consider only the first type as morphology. The elements in the second class could be called “clitics” or “particles”, uninformative terms that I use here informally for elements that do not display any morphophonological interaction with their hosts, and display either some variability in prosodic phrasing (attaching alternatively to the right or to the left) or the possibility of appearing as free-standing elements.<sup>4</sup>

All morphology of the first type on Mëbêngôkre verbs is prefixal save for finiteness.<sup>5</sup> In addition to person inflection, there are two families of valency-reducing prefixes that are fairly productive: an anticausative and an antipassive. All but one subclass of transitive verbs, those that assign accusative case, as well as some relational nouns, have a prefixal relationalizer or transitivity prefix. These may be seen in the following examples.

<sup>4</sup>The form of clitics when they stand free may differ slightly from their form when they are phrased with other material. The unstressed demonstrative *ja* becomes stressed *jā* ‘this one’, for instance, while the focus-associated particle *bit* ‘only’ becomes *ajbit*.

<sup>5</sup>On nouns, there are diminutive and augmentative (in Kayapó it is a free-standing root), as well as “honorific” and vocative suffixes used on kinship terms, all of which could be considered to belong in this class. The diminutive is exemplified in (10c).

- (10) a. *i-bi-xa-djwỳ-rỳ*  
1-ANTIC-TR-put\_down.PL-NFIN  
'us coming down'
- b. *i-dju-ja-rě-nh*  
1-ANTIP-TR-tell-NFIN  
'me telling stories'
- c. *i-ka-my-re*  
1-REL-male-DIM  
'my little brother'

The cohesion of these affixes is clear enough for me to define a verbal word comprising them (the verb stem would be the verbal word without the person prefix; see the definition in §6.1). Note that none of these affixes is obligatory in general (they may be obligatory for a given root or class of roots), so that there are verb stems that consist only of a root: *rwỳ* 'go down (sg.)', *tě* 'go (sg.)', etc. This comes up again when discussing imperatives below.

There is some degree of semantic or morphological idiosyncrasy in all the affixes comprising the stem, and other than the finiteness suffix, their productivity is not high. The transitivity prefixes are borderline morphology, as they display a high level of lexical idiosyncrasy both in selection of hosts and in contribution to meaning, to the extent that in most cases it is impossible to separate them from the root. Despite this, they have been recognized as morphology in previous work; for discussion see Salanova (2014a), as well as de Oliveira (2005: 116–128) for the related language Apinajé. This is supported by a handful of sets of verbs that differ only in the prefix, as well as by a small number of cases where plural number is associated with the prefix.

Among the elements of the second type are several "clitics" in the informal sense proposed above. These fall into at least three different classes:

1. Noun-phrase internal modifiers, such as demonstratives, quantifiers, and particles like *bit* 'only', which may be free-standing or lean on the material to their left, with which they are invariably related morpho-syntactically (or, in the case of *bit*, via association with focus). These are slots 7 and 9 in the noun phrase template.
2. Postpositions, which, though normally related morpho-syntactically to the phrase on their left, may form an idiomatic unit with a verb to their right and phrase prosodically with it (in which case I consider them to be preverbs, slot 11 below). Postpositions may also appear to the right of the

non-finite form of a verb, either acting as a clausal modifier if the verb in question heads the matrix clause, or subordinating that verb to some other predicate (see discussion of examples (6)–(7)); this is slot 17 below.

3. Adverbial particles, which occupy specific positions in the clause (slots 3 and 6, and zone 5 below) and lean onto a host to their left. These normally do not form a morpho-syntactic constituent with their phrasal hosts, but rather are clause modifiers.

The most interesting of these is the second class: postpositions may phrase to their left with non-finite forms of verbs or with nouns, or to the right with verbs, but may be freestanding in certain circumstances. The following examples show these three possibilities.

- (11) a. *ba kàx=o krã-'yr o=dja*  
 1.NOM knife=with 3.head-cut.NFIN with=stand  
 'I'm cutting it with the knife.'
- b. *o ne ba krã-ta*  
 3.with NFUT 1.NOM 3.head-cut.FIN  
 'I cut it *with it*'

The behavior of adpositions is related to another gradient area of Měbêngôkre syntax, namely the distinction between arguments and adjuncts in the verb phrase, or, relatedly, the degree of idiomaticity of a sequence of adposition and verb. The interaction between these factors comes up again below, but a detailed discussion is beyond the scope of this chapter. For more information, see Salanova (2014b).

Aside from these morphological processes, compounding is rather important in Měbêngôkre, but since it is a phenomenon of the nominal domain it is discussed only briefly in this chapter, in §5.2. For the role of compounding in word formation, see Salanova & Nikulin (2020).

## 5 Měbêngôkre planar structure

The planar structure for independent clauses is provided in Table 1. A number of positions around the root are exclusive to verbs (e.g., valency-reducing prefixes, finiteness suffixes), but the template is essentially the same for matrix nominal predication, which is not discussed separately here.

Table 1: Verbal planar structure of Mëbêngòkre

Pos.	Type	Elements	Forms
(1)	slot	polar interrogative particle	<i>djām, djori</i>
(2)	slot	XP fronted for contrast	(open)
(3)	slot	tense/mood marker	<i>ne, dja</i>
(4)	slot	nominative subject NP or pronoun	(open)
(5)	zone	particles of varied semantics	<i>arým, on, 'ýr, tu, bit, kam, te, arek, etc.</i>
(6)	slot	subject paucal number	<i>ari</i>
(7)	slot	oblique subject NP or index	(open)
(8)	slot	oblique subject marker	<i>-te/-je, -mā, -bê</i>
(9)	zone	XP (adjuncts)	(open)
(10)	slot	direct object NP or index	(open)
(11)	slot	“preverb”	<i>o, mā, kam, 'ā, etc.</i>
(12)	slot	subject person	<i>i-, a-, ba-, 0-</i>
(13)	slot	valency-reducing prefixes	<i>aj-/bi-, a-/djā-/dju-</i>
(14)	slot	transitivity prefixes	<i>ka-, nhi-, py-, etc.</i>
(15)	<b>slot</b>	<b>VERB ROOT</b>	<b>(open)</b>
(16)	slot	non-finiteness suffix	<i>-r, -nh, -m, -k, -x</i>
(17)	slot	P (governed by auxiliary)	<i>o, mā, mo (&lt; mā + o)</i>
(18)	slot	subject person (on aux.)	<i>i-, a-, ba-, 0-</i>
(19)	slot	auxiliary root	<i>nhÿ, nō, dja, mō, etc.</i>
(20)	slot	non-finiteness suffix	<i>-r, -m, etc.</i>
(21)	zone	non-verbal governing modifier	<i>kêt, rā'ā, kadjy, mā, 'ýr, pro, kajgo, etc.</i>
(22)	slot	light manner predicates	<i>o, kute</i>
(23)	slot	nominal subordinate clause	(open)

Some freedom in ordering is allowed in three distinct areas of the clause: in the left field of the clause (zone 5), a number of particles with varied semantics (aspectual, conditional, frustrative, hearsay evidential, etc.) are ordered in a way that partly reflects their semantic scope; in the middle field (zone 9) various XP dependents of the predicate are ordered according to principles of selection (more object-like closer to the predicate, more adjunct-like farther from it); in the right field (zone 18) several modifiers that govern the predicate are ordered scopally. Slots 2 and 23 are for constituents that are information-structurally marked; while slot 2 is very often filled, slot 23 is used more rarely, and only for

heavy constituents such as non-finite subordinate clauses, in which case it is not entirely clear that the construction doesn't involve the paratactic juxtaposition of two clauses rather than a single complex clause.

Several additional complications are avoided by my choice to cut sentences wherever a conjunction appears, as discussed in §3. Finally, there are a few adverbial particles that have a variable or an as yet undetermined position in the clause, such as the durative *ari*. These are also excluded from the template.

One final general remark regarding the template that I propose is that it is modelled primarily on statements that convey categorical judgments, that is, those that have a theme-rheme structure. Clauses where this is not the case exhibit orders that deviate from the proposed template, though not radically: the subject, normally in 4, might appear as far into the clause as the left edge of zone 9. In the following sentence expressing athetic judgment, the subject appears after two particles from zone 5:<sup>6</sup>

- (12) *arým amrē Kajtire tē*  
       5      5      4      15  
       already hither Kajtire go.FIN  
       '(It is already the case that) Kajtire is coming.'

### 5.1 Interactions among positions in verbal clauses

Two interactions between positions in the template should be pointed out. Both of them have to do with the effects that governing elements in slots 17, 19, and 21 have on the finiteness of verbal heads, and the effect that verbal finiteness has in turn on the expression of arguments. The fact that so much changes in the clause according to whether the lexical predicate is governed or ungoverned by an auxiliary raises the question of whether the finite word, be it a lexical verb or an auxiliary, shouldn't be considered the head of the clause, with the template rearranged accordingly. Our decision regarding this is based on the mostly practical considerations raised on page viii.

The first of the interactions may be summarized as follows: if there is an element in slot 21, then the auxiliary (if present) will have a non-finiteness suffix in slot 20. If an auxiliary is present in slot 19 or a non-verbal governing modifier is present in 21, then it will be the verb root that will have a non-finiteness suffix in slot 16.

The second interaction affects the presence of person indices. Verbs or auxiliaries that are non-finite differ from finite verbs and auxiliaries in the expression

<sup>6</sup>For the thetic vs. categorical distinction, see Kuroda (1972).

of their arguments. A non-finite auxiliary will have a person index in slot 18, while a finite one won't.<sup>7</sup> Slot 12 is filled in most nominal predicates and non-finite intransitive verbs, and with a few finite intransitive verbs such as *kato* 'to exit' and *nhire* 'to let go'. In transitive verbs, position 10 is filled regardless of finiteness, but a third-person index in this position is in complementary distribution with a full noun phrase.

My separation of position 10 from position 12 in the template hinges on the fact that the former permits a noun phrase while the latter may only be filled by a person index, which can be co-referential with a noun phrase in slot 4. If a verb governs an oblique object, there will be an element present in 11, and a subject person index will occupy position 12 under specific circumstances. A typical example is the following:

- (13) *ba ku-m i-nhire*  
 4 10-11 12-15.16  
 1.NOM 3ACC-to 1-let\_go.FIN  
 'I let him/her go, I dropped him/her off.'

The first object of a verb, whether direct or oblique, may be differentiated from other objects by a number of diagnostics. Its interpretation is limited to certain thematic roles, for instance, and if oblique, it obligatorily strands its adposition when fronted.

The analytic choice to separate positions 10 and 12 in the template has practical value in that it simplifies the presentation of the structure of the clause. Still, scholars that are familiar with Northern Jê languages might find this separation arbitrary, since the subject person indices in 12 are in most cases formally identical to the indices used for objects, and the two positions are indistinguishable in the case of regular transitive verbs due to the absence of an object-governing adposition or "preverb" in 11. One fairly cogent objection is that a few verbs governing oblique objects have "expletive" direct objects in the accusative case (I thank Andrey Nikulin for bringing this up):

- (14) *ba pĩ=mã ku-ta*  
 4 9 10-15  
 1.NOM tree=to 3ACC-cut.FIN  
 'I'll fell the tree.'

---

<sup>7</sup>There are special cases: verbal auxiliaries appearing with stative main verbs, as well as the auxiliaries *pa* 'to complete', *oinore* 'to finish', and a few others, never take person inflection. This process is well understood albeit not fully described in my work, but discussing it here would take me too far afield.

Under the present approach, one would be forced to give the phrase *pĩ=mã* in (14) the status of an adjunct or second object at best, making it different from *ku-m* in (13), which is considered an oblique direct object. This would be quite counter-intuitive and arbitrary. Alternatively, we could consider *ku-* to be in slot 12. This is also a poor fit, as in all other cases the indices in this position are co-indexed with the subject.

Until we have a clear idea of the prevalence and proper analysis of verbs with expletive objects, however, I believe that the existence of a few verbs that behave as in (14) does not justify a change in the template. Regarding such cases, it is not clear to me whether the “expletive” direct object index has reference distinct from or identical to the oblique phrase, but analyses that would imply only a small adjustment to the template are possible for each of those situations. In the first case the construction would be an idiomatic expression with an implicit object, while in the second it could be described as involving differential object marking.

One further point raised by an anonymous reviewer has to do with the secondary exponence of finiteness in a handful of verbs that describe bodily functions, such as the following:

- (15) a. *arým ne itu*  
           2       3   15.16  
           already NFUT urinate.FIN  
           ‘S/he urinated.’
- b. *tu-ru kêt*  
           15-16       21  
           urinate-NFIN NEG  
           ‘S/he didn’t urinate.’

The appearance of initial *i-* in the finite forms of these verbs can plausibly be related to the allomorphy of valency-reducing prefixes that is discussed in diagnostic [20]. However, in this case the element that is prefixed is meaningless, and should not be considered a morpheme occupying a slot. If additional information later forces me to assign it to a slot, the likely candidate would be slot 13.

Mutatis mutandis, this applies to a more abstract palatalizing prefix identified in Nikulin & Salanova (2019), which is responsible for some synchronically irregular finiteness alternations such as *kate* ‘to shatter (FIN)’ vs. *ka’êk* ‘to shatter (NFIN)’: the consonant alternation is never the sole exponent of finiteness, and hence does not need to be considered a morpheme separate from the non-finiteness suffix.

Finiteness of the main predicate also affects the case of subjects, but this is a complex matter which I cannot address here (for discussion, see Salanova 2008, 2017a). Very broadly speaking, nominative subjects are found with finite verbs, while oblique subjects – a category that includes the ergative – are found with non-finite verbs. Different post-verbal modifiers complicate this picture by allowing nominative subjects to appear with non-finite main predicates. Further complications include the fact that a number of non-verbal predicates also require oblique subjects, that the ergative may optionally appear with active intransitive verbs if adjuncts intervene between it and the verb, and that a nominative pronoun can always be present in an independent clause, even if redundant.

## 5.2 The noun phrase template

The structure of noun phrases in Mëbêngôkre is examined in two previous papers, Salanova (2017b, 2020), and is not addressed in detail here. I provide the positions of the nominal planar structure in summary form in Table 2 to allow a simple comparison with the clausal template.

Table 2: Nominal planar structure of Mëbêngôkre

Pos.	Type	Elements	Forms
(1)	zone	Modifiers	<i>ap̃yñh</i> , PPs
(2)	slot	Complement NP	(open)
(3)	slot	Nominal relator	<i>ka-</i> , <i>nhi-</i> , <i>dju-</i>
(4)	<b>slot</b>	<b>Noun root</b>	<b>(open)</b>
(5)	zone	Governing modifier	<i>kaàk</i> , <i>kajgo</i> , <i>djw̃yñh</i> , <i>mex</i> , <i>punu</i> , <i>ti</i> , etc.
(6)	slot	Dimunitive and related	<i>-re</i> , <i>-jê</i> , <i>-wa</i> , <i>-ti</i>
(7)	zone	Determiners and related	<i>ja</i> , <i>wã</i> , <i>’õ</i> , <i>kw̃y</i> , etc.
(8)	slot	Adposition or case	(small class)
(9)	slot	Focus-sensitive particles	<i>bit</i>

The elements that make up the noun phrase are less differentiated than the various elements that compose the clause. It is not clear whether elements in slot 5 should in fact be considered distinct from the root in slot 4. The relationship between these two positions is formally no different than that between two roots in a “compound” (i.e., between positions 4 and 2): the word on the right is a relational word, and takes the one to its left as its complement. For further discussion, see the two papers cited above. Like in the case of post-verbal elements



in the clausal template, I have adopted a practical rather than a fully principled solution.

## 6 Constituency diagnostics

In this section, I describe all imaginable diagnostics for constituency applied to Měbêngôkre. By constituency diagnostic I refer to some generalization over the constructions of the language that identifies a subspan in the planar structure. In a first subsection I focus on diagnostics that are commonly applied cross-linguistically to identify words, such as non-interruptability, free occurrence, and so on, unfolded to capture various ways in which they can apply to the language. In the second subsection I discuss diagnostics that are typically used to identify larger constituents, such as pause and the domain of idiomatic interpretation. In the third subsection I discuss phonological and morpho-phonological processes with specific domains of application.

### 6.1 Morphosyntactic constituency

Among recurrent diagnostics for morpho-syntactic wordhood in descriptive studies are things such as interruptability and fixed order of elements, an identification that rides on a real or imagined contrast between syntactic and morphological principles of composition when it comes to their flexibility and productivity.

For this section, it is useful to define the *verb stem* as comprising slots 13-16. The stem functions as a unit for all the diagnostics in this section. In particular, given the fusional nature of the elements in slots 13, 14 and 15, I often do not show segmentation among them.

#### 6.1.1 Ciscategorial selection (13-16; 8-16)

Měbêngôkre independent clauses may be headed by nouns, non-finite verbs, or finite verbs. A number of elements farther from the head of the predicate can occur with predicates of all categories, so it is interesting to ask which slots around the head are conditioned to appear according to the category of root. In fact, this criterion clearly identifies the verb stem as I have just defined it, comprising slots 13-16. Valency-reducing prefixes from slot 13 and non-finiteness markers from slot 16 are never found on nouns. A handful of nouns appear to have transitivity prefixes from slot 14 serving as nominal relationalizers, e.g., *ka-ngô* ‘water or juice of...’, from *ngô* ‘water’ (see also (10c)). However, contrary to the transitivity prefixes, with appear in most transitive verbs, relationalizers appear

Table 3: Diagnostics applied to Mëbêngòkre

Diagnostic	Description
[1]	Ciscategorial selection of slots (13-16)
[2]	Ciscategorial selection of elements (8-16)
[3]	Minimal free occurrence in imperatives (10-15)
[4]	Minimal free occurrence in declaratives (4-15)
[5]	Maximal free occurrence (3-21)
[6]	Recursive interruption (11-22)
[7]	Non-recursive interruption (11-16)
[8]	Permutation rigidly reflects scope (12-22)
[9]	No permutation permitted (12-15)
[10]	Maximal span repeated in coordination (4-22)
[11]	Minimal span repeated in coordination (10-15)
[12]	Maximal span repeated in subordination (5-15)
[13]	Minimal span repeated in subordination (10-15)
[14]	Pause (11-16)
[15]	Domain for idiomatic interpretation (9-21)
[16]	Fortition at juncture (13-15)
[17]	Dissimilation of homorganic rimes (15-16)
[18]	Aphaeresis of palatal if initial in domain (12-15)
[19]	Suppletion for number (13-15)
[20]	Allomorphy of valency-reducing prefixes (13-15)
[21]	Vowel syncope if non-initial in domain (12-15)
[22]	Largest span on which only one stress occurs (11-16)
[23]	Domain in which stress is final (11-16)
[24]	Echo vowel after domain-final /r/ (11-20)

haphazardly in nouns and never form paradigms with nouns containing other prefixes. I therefore consider them to be a distinct morphological category from verbal transitivizers.

There is a broader way to define ciscategorial selection if one focuses not on the presence of a slot but on the set of elements that a slot may contain relative to the category of the head of the predicate. This is similar but different from the control of allomorphs discussed as diagnostic [20], as it involves elements that are meaningful in isolation, i.e., different adpositions appearing in position 8 according to the subclass of predicate that governs them and to whether this predicate is finite or nonfinite.

In the application of the diagnostics, these two ways of defining ciscategorical selection are distinguished.

### 6.1.2 Free occurrence (10-15; 4-15; 3-21)

Free occurrence refers to the ability of a certain sequence of elements to stand as a complete utterance. As is natural to expect, there are variables that affect the definition of the free occurrence span, and this requires that the diagnostic be fractured. A first-order fracture distinguishes between minimal and maximal free occurrence. The minimal free occurrence is the shortest independent utterance that spans the verb root. Maximal free occurrence is the single span that extends to cover all elements in the clause that may not appear as free utterances.

The first version of the diagnostic in particular may be further fractured in a number of ways. Given what was said above regarding obligatoriness of inflection, it is to be expected that differences arise between transitive and intransitive verbs, and between finite and non-finite forms of each. Finite intransitive verbs can in principle stand on their own in imperatives, as in (16a), though in practice the additional presence of an adverbial or particle from slots 5 or 9 is more idiomatic, as in (16b). Unless derived, these verbs consist of just the root.

- (16) a. *dja*,        *tě*,        *to*  
           15        15        15  
           stand.FIN go.FIN dance.FIN  
           ‘stand! go! dance!’
- b. *kàjmā dja*,        ‘*ỳrỳ* *tě*,        *tě=n*        *to*  
           9        15        9        15        15        15  
           upward stand.FIN up\_to go.FIN go.FIN=and dance.FIN  
           ‘stand up! go up to it! go dance!’

Transitive verbs, on the other hand, do not forfeit the requirement for person inflection even in the imperative.<sup>8</sup> The span involved would thus be 10-15. Like with transitive verbs, the presence of adverbials or particles is more idiomatic, but not an absolute requirement:

- (17) (*on*) *krě*,        (‘*ỳr*) *o=tě*,        *a-ma*  
           5    10.15    9    10.11=15    10-15  
           now 3.eat.FIN up\_to with=go.FIN 2>3-hear.FIN  
           ‘eat it! take it there! listen to it!’

<sup>8</sup>Because of a morphological idiosyncrasy of the third person, the object index is zero with some verbs, but is overt with others. See discussion on page v.

I take the transitive construction to be representative of the minimal free occurrence span with imperatives (diagnostic [3]). With declaratives (diagnostic [4]), a subject normally has to be present. When sentences are coordinated, third-person subjects are frequently omitted. In free-standing utterances omission of an overt subject outside of coordinated constructions is not normally idiomatic but does occur in the third person; that these few occurrences are instances of a morphological zero rather than of the absence of the position is suggested by the obligatoriness of the number particles in the subject-modifying position 6 if reference is plural, as well as by its anaphoricity, already mentioned in connection with third person indices.

Maximal free occurrence (diagnostic [5]) extends across a fairly large span of the sentence. Many post-verbal elements cannot be used as free forms and neither can tense markers (slot 3) or particles in slots 5-6. Regarding these left-peripheral elements, one could more insightfully say that there are two domains for bound elements in the middle field of the Měbêngôkre clause, one around positions 3-6, the other centered on position 15, and that a number of free-standing elements can appear elsewhere. For post-verbal elements, their bound status depends as much on prosodic and semantic properties as on the specific position they occupy. Elements in position 17 are always bound, while those in 19 are generally free, as they are identical to lexical verbs (in turn, positions 18 and 20 are bound to them); position 21 contains both free and bound elements: negation *kêt*, for instance, constitutes a complete utterance on its own, whereas prospective *mã* is always bound:

- (18) a. *ba kam ku-m arê-nh kêt*  
           4     5     9       10.15-16   21  
           1.NOM then 3ACC-to 3.say-NFIN NEG  
           ‘So I didn’t tell him/her about it/her/him.’
- b. *kêt*  
    21  
    NEG  
    ‘No; there isn’t any.’
- c. *ku-te ku-m arê-nh mã*  
    7-8     9       10.15-16   21  
    3ACC-ERG 3ACC-to 3.say-NFIN PROSP  
    ‘S/he is about to tell it to him/her.’
- d. # *mã*  
    21  
    PROSP

### 6.1.3 Non-interruptability (11-22; 11-16)

The non-interruptability diagnostic identifies the span overlapping the verb stem that cannot be interrupted by free forms. It is fractured into recursive [6] and non-recursive [7] interruption, i.e., spans that may be interrupted by a single free form and by multiple free forms, respectively. In Mëbêngôkre, these spans both begin in position 11 and extend to position 16 in the case of non-recursive interruption and to the end of the clause in the case of recursive interruption.

The element that may interrupt the span between the lexical verb and an auxiliary is either a manner modifier which syntactically becomes the main predicate, or the element *ari* ‘constantly’, which does not govern the preceding element but rather modifies the auxiliary that governs the verb. Examples of each type of interrupting element are as follows:

- (19) *Ta ne ami-jo mỳja ma-ri mex o=ba.*  
 2 3 9 10 15-16 – 17=19  
 3EMPH NFUT self-with thing know-NFIN well with=3.live  
 ‘He kept learning things properly for himself.’
- (20) *Nã bãm ami-wỳr kam ama-k ar o=i-ba.*  
 3 4.5 9 11 15-16 – 17=18-19  
 PRS 1.NOM.PRS self-up\_to 3.on wait-NFIN constantly with=1-live  
 ‘I keep waiting for him/her to come to me.’ (1 Corinthians 16:11)

One might expect that, at least in the case of manner modifiers, wherever simple interruption may occur, recursive interruption may as well. However, recursion of modifiers is not generally permitted in the language (this point may also be seen in the case of modification within a noun phrase, discussed briefly above). Multiple modification requires coordination, which by my definition establishes the boundary of a new planar structure. This leaves as the only recursive device interrupting a span the adjunction of modifiers in position 9:

- (21) *Me jã ne me arỳm kadjy ku-m arẽ-nh o=dja.*  
 2 3 4 5 9 9 12.15-16 17=19  
 PL this NFUT PL already for 3ACC-to 3.say-NFIN with=3.stand.FIN  
 ‘And these ones were talking to him for that purpose.’

### 6.1.4 Non-permutability (12-22; 12-15)

The order of elements in the Mëbêngôkre clause is overall fairly rigid, but permutation is possible in several zones, in addition to the possibility of movement to the clause-initial position 2. The former possibility is clear in the particles of

zone 5 and the phrases of zone 9, where ordering follows criteria of semantic scope or relatedness to the predicate. Fronting to clause-initial position is also a possibility for many particles found in slot 5, in addition to phrases in 4, 9 and 10:

- (22) a. *Arỳm ne ba ar a-mã i-kabēn jarē.*  
 2 (5) 3 4 9 9 10 15  
 already NFUT 1.NOM PAUC 2-to 1-speech say.FIN  
 ‘I’ve *already* told you my speech.’
- b. *I-kabēn ne ba arỳm ar a-mã arē.*  
 2 (10) 3 4 5 9 9 10.15  
 1-speech NFUT 1.NOM already PAUC 2-to 3.say.FIN  
 ‘I’ve already told you *my speech*.’

The left edge of the span identified by the non-permutability diagnostic is thus clearly after 10. The right edge is harder to identify.

Prima facie it might appear that the post-verbal modifiers can front to 2, as in (23b), as long as they are not finite auxiliaries. However, this is an epiphenomenon created by the homonymy between the post-verbal modifiers and adpositions, which may constitute a phrase with a morphologically null complement and are thus mobile. Elements like *kadjy* are always interpreted as adpositions (‘for the purpose of NP’) when fronted as in (23b), never as verbal modifiers (‘supposed to V’).

Claiming that what one sees in (24) does not involve the fronting of a manner modifier requires subtler argumentation, but in my view is equally justified: the two sentences are simply built differently, not related by movement, even if the meaning difference in this case is less obvious. The difference in construction can be seen in the fact that the sentence in question is finite (even if the verb does not have the morphology to show it). If *mex* were a governor of the verb, this would not be possible. It can also be detected in meaning, which in (24b) points to *mex* having been displaced from the object position.

- (23) a. *I-je ku-m ã-rã kadjy.*  
 7-8 9 10.15-16 21  
 1-ERG 3ACC-to 3.give-NFIN PROSP  
 ‘I’m supposed to give it to him.’
- b. *Kadjy ne i-je ku-m ã-rã.*  
 2 (21) 3 7-8 9 10.15-16  
 for NFUT 1-ERG 3ACC-to 3.give-NFIN  
 #‘I’m *supposed* to give it to him.’ (only: ‘What I gave him is for that purpose.’)

- (24) a. *I-je ipêx mex.*  
           7-8 10.15-16 21  
           1-ERG 3.make.NFIN good  
           ‘I made it well.’  
       b. *Mex ne ba ipêx.*  
           2 3 4 10.15  
           good NFUT 1.NOM 3.make  
           #‘I made it well.’ (only: ‘I made a good one.’)

Even if such fronting is not possible with the elements in slots 17-22, there are a few cases of permutation in situ which define a span that does not extend to the end of the clause. Auxiliaries may exceptionally appear after adpositional aspectual modifiers, as in example (25), from Romans 7:19, and in (26):

- (25) *Te i-mã i-jaxwe kînh kêt mã o=i-ba.*  
       (5) 7-8 10 15 21 21 17=18-19  
       in\_vain 1-to 1-evil like not PROSP with=1-live  
       ‘I do the evil that I don’t like.’  
       (26) *Me arek a-tykdjà kêt o=a-kri ngrire.*  
           4 5 12-15 21 17=18-19 –  
           PL still 2-fatigue NEG with=2-sit.PL small  
           ‘Stay and rest (catch your breath) a little bit.’

The interpretation of (25) is not fully clear to me. It is possible that the span beginning with *i-mã* (7-8) and ending with *mã* (the second 21) functions as an adjoined subordinate clause, and that *o=i-ba* doesn’t govern it but instead governs a morphologically null third person pronoun that co-refers with *i-jaxwe*. In the case of (26), the structure is straightforward, but the sentence has the disadvantage of having a nominal predicate (*tykdjà*), and of *tykdjà kêt* being an idiomatic expression of sorts. It may be seen, thus, that clear examples of permutation of post-verbal elements are rather hard to find. Still, the nature of these elements is such that permutation should be possible, and might be rare because of scope considerations: prospective scoping over progressive is conceivable, while the opposite is less so, for instance.

Given this, the qualitative distinction between the rigidly ordered morphemes of the verb stem and the freer though only marginally mobile elements that appear after the verb is captured by the two subcases into which the non-permutability diagnostic is fractured: [8] permutation is permitted but transparently reflects scope, and [9] permutation is not permitted at all.

### 6.1.5 Subspan repetition (4-22; 10-15; 5-15)

When clauses are coordinated or subordinated, part of the content of one of the clauses will typically be elided. The diagnostic of subspan repetition refers to the subspan of the clause that may appear repeated in coordination or in a non-finite dependent clause. This diagnostic is fractured according to each of these cases. In the case of coordinated structures, the diagnostic fractures further into [10] maximal subspan repeated in a coordinated structure involving the verb, and [11] minimal subspan repeated in a coordinated structure involving the verb. In the case of constructions involving subordination, the diagnostic is fractured between [12] the maximal span of elements that may occur in a subordinate clause, and [13] the subset of these that need to be present in any subordinate construction.

The application of these diagnostics is relatively straightforward. The following examples illustrate maximal and minimal examples in coordinated structures:

- (27) a. *Dja ba a-m arē ga arȳm i-kabēn ma.*  
           3 4 9 10.15 4 5 10 15  
           FUT 1.NOM 2-to 3.say.FIN 2.NOM already 1-speech hear.FIN  
           ‘I’ll say it and you’ll hear my words.’
- b. *Dja ba mǎ tē=n abym ar a-wȳr tē=n bōx.*  
           3 4 5 15 5 9 15 15  
           FUT 1.NOM away go.FIN=SS back PAUC 2-up\_to go.FIN=and arrive  
           ‘I’ll go away and return to you and arrive.’

As can be seen with the third conjunct of (27b), the conjunct can be as small as just the verb stem. On the other hand, if subjects in the conjoined clause are different, the conjunct will necessarily extend to the left all the way to the subject position, 4, as seen in (27a).<sup>9</sup> The second conjunct of (27b) represents an intermediate situation with identical subjects.

Subordinate clauses in general, since they can only be non-finite, have a more limited template which excludes the focus position and the position for nominative subjects. In example (28a), that template is maximally filled. Minimally, it must contain a verb stem, as in (28b).

<sup>9</sup>If the nominative subject is a speech-act participant, the conjunction is unexpressed. For third person subjects, the conjunction may be the same-subject conjunction *ne*, or the different-subject conjunction *nhym*. This raises the question of whether speech-act participant pronouns encompass the conjunction, as proposed by Nonato (2014) for a related language. For reasons of space, we cannot address here the consequences of this analytical step.



- (28) a. *Ga [ ku-te ajte akubyn me ba-wỳr ano-ro=ja ]*  
 4 7-8 5 5 9 10.15-16  
 2.NOM 3ACC-ERG again back PL 1INCL-up\_to 3.send-NFIN=this  
*pumũ.*  
 15  
 see.FIN  
 ‘You see that he has sent him back to us.’ (Luke 23:15)
- b. *Ne kam [ uma=je ] prõt.*  
 5 10.15 15  
 and then 3.fear=because 3.run  
 ‘And then s/he ran because of fear.’

## 6.2 Syntactic and semantic criteria (11-16; 9-21)

Pause is often used to diagnose morpho-syntactic domains. Here it is defined as the smallest span around the verb that can be delimited by pauses (diagnostic [14]); with that definition it defines the same span as non-recursive interruption (diagnostic [7]) and a couple of phonological diagnostics.

Another possible diagnostic for morpho-syntactic domains larger than the verb stem is based on the span of idiomatic interpretation (diagnostic [15]). The following are examples of idiomatic expressions in Měbêngôkre extending over various positions of the planar structure:

- (29) a. *ba arỳm i-tĩn prām*  
 4 5 10 15  
 1NOM already 1-life want  
 ‘I was afraid.’ (lit., ‘I wanted my life.’)
- b. *arỳm ne me bõ-m ku-mě*  
 5 3 4 9 10-15  
 already NFUT PL grass-to 3ACC-throw.FIN  
 ‘They expelled him.’ (lit., ‘They threw him to the grass.’)
- c. *ba pi’òk jarě-nh o=dja*  
 4 10 15-16 17=19  
 1NOM paper say-NFIN with=stand.FIN  
 ‘I’m lecturing.’ (lit., ‘I’m reading standing up.’)

In the case of (29a), one could say that positions 10 and 15 are parts of the idiom; in (29b), 9 and 15 are part of the idiom as well; in (29c), somewhat more tenuously,

19 may be argued to form an idiom with 15, since with a different auxiliary the interpretation is not of lecturing but rather of studying or reading for one's own sake.

A minimal counterpart for this diagnostic could also be defined, but yields less relevant results: only positions 14 and 15 combine to yield meaning in a systematically non-compositional way.

### 6.3 Phonological and morpho-phonological domains

The aim of this section is the identification of spans required by a number of phonological and morpho-phonological processes around the verbal base in Mẽbêngôkre. Morpho-phonology occurs in specific morpheme junctures, and is of limited relevance to define domains given the small number of morphemes that are affected. Still, a number of diagnostics may be defined on the basis of morpho-phonological processes that apply in certain spans but not elsewhere, and on the basis of allomorph selection. Among the former are [16] strengthening of palatals in certain environments (13-15), [17] dissimilation of high vowels next to homorganic codas (15-16), [18] dropping of certain consonants next to a person index and other allomorphic processes affecting vowel- or glottal stop-initial stems (12-15), [21] dropping of high back vowels in a stem conditioned by prefixation (12-15), and [19] suppletion for number (13-15). Among diagnostics based on control of allomorphs, I identified [20] allomorphy of valency-reducing prefixes based on finiteness (13-15). Not all of these diagnostics need to be discussed.

Diagnostic [16] is based on a fortition process that applies to certain coronal continuants in particular environments. The following distinct instances have been identified:

1. Fortition of /ɾ/ into /t/ or /n/ in the diminutive suffix *-re* when attached to a stem that ends in a noncontinuant coronal consonant. Examples include *kẽn-ne* 'small stone', *amàt-te* 'small piranha', *kwên-ne* 'small bird', *tỳx-te* 'pretty strong'.
2. Fortition of /j/ into /tʃ/ (orthographic *x*) in the honorific suffix *-jê* used in kinship terms, in contexts similar to the preceding process.
3. In Xikrin, /ɾ/ is fortitioned to /t/ or /n/ before a consonant within a certain domain which includes stems in compounds and some extra dependent categories: *par-kà* 'shoe' → [pat'kʌ], *bàr-prà* 'charcoal' → [bʌt'pɾʌ], *ar ga* 'you few' → [an'ga] (though in this case [an] is maintained in domain-final position, while [ar] is only found before vowels; see diagnostic [24]).

4. The morpho-phonological fortition of /j/ in verbs that receive the anticausative prefixes *aj-* and *bi-*. Examples of this in the lexicon are few, but the rule applies consistently: /bi-jabjer/ ‘to trickle’ → [bitʃaʔbjere], /bi-jaer/ ‘to play’ → [bitʃaʔere]. With the prefix *aj-*, employed with finite verb forms, there is some irregularity: /aj-jabij/ ‘to trickle’ → [atʃiʔbija], /aj-jae/ ‘to play’ → [aʔtʃe].

The applicability of this diagnostic around the verb is rather limited, since most environments for fortition occur in the nominal domain, but it does define a span that extends to the right from slot 20 to the left of the verb root, and, if one accepts the following data from a speech style called “angry speech”, where *-re* fails to fortition, excludes post-verbal modifiers in position 21:

- (30) a. *Ba on me’õ bũnh=re.*  
 1NOM now someone kill=DIM  
 ‘I’m going to kill (< *bĩ*) someone.’  
 b. *Ba on mỳja krõnh=re.*  
 1NOM now eat=DIM  
 ‘I’m going to eat (< *krẽ*) something.’

Diagnostics [18] and [21] refer to a family of stem changes, some of which are clearly morphologically triggered, while others likely rely on morpho-phonological domains. Certain stem-initial consonants on verbs and other lexemes get deleted when initial in a relational stem whose complement is not overt, while in almost exactly opposite circumstances a high back vowel on the initial syllable of the stem is dropped. The following data exemplify this.

- (31) a. *ngô jadjà*  
 water put\_in.FIN  
 ‘to fetch water’  
 b. *adjà*  
 3.put\_in.FIN  
 ‘to put it in’  
 (32) a. *kà kdjô*  
 skin peel.FIN  
 ‘to skin’ (Xikrin pronunciation)  
 b. *kudjô*  
 3.peel.FIN  
 ‘to skin it’

Elsewhere (Salanova 2011) I argued that the *prima facie* morpho-phonological process in (31) is not in fact domain-dependent but rather is the non-concatenative exponence of third person inflection. This is reflected in my glosses. The diagnostic is considered not to apply in such cases. The process in (32), on the other hand, does define a domain, differently in the Xikrin dialect (where it identifies span 10-15) than in the Kayapó dialect (where it identifies the same span as (31)).

The relevance of this diagnostic is likely greater for homologous morpho-phonological processes in closely-related languages (such as the realization Timbira prenasalized consonants, discussed in Salanova 2011), where proclitic elements that are not directly governed by the element that follows affect the application of the rule.

A strictly morphological diagnostic may be defined with reference to the exponence of number. Diagnostic [19] defines the maximal span around the verb root over which suppletion for number may apply. Such “suppletion for number” in Mëbêngôkre is not clearly a reflex of a productive morphological process. Though a number of verbs exist that oppose a singular/actional form and a plural form and a handful of these encode the opposition by means of non-suppletive morphology (mainly by substitution of the transitivity prefix), the distinction does not pervade the verbal lexicon of the language, and might be better characterized as relating pairs of lexically distinct verbs. If the distinction is considered morphological, then there is a clear maximum span for what may be suppleted.

Take the two verbs used for the plural and singular form of ‘descend’ or ‘be born’, respectively *bixadjwỳr* and *rwỳk*. The singular form suppletes for a plural form that includes transitivity prefixes and an anticausative prefix. The stem composed by *ja-* and *djwỳ-r* independently means ‘to lay down (plural)’.

(33)

13	14	15	16
<i>bi-</i>	<i>ja-</i>	<i>djwỳ</i>	<i>-r</i>
<i>rwỳ</i>			<i>-k</i>

In the case of all pairs of postural verbs, the plural form is a nominal predicate, a category which lacks a finiteness distinction. The following example is from the singular and plural verbs ‘to sit’:

(34)

15	16
<i>nhỹ</i>	<i>-r</i>
<i>krĩ</i>	

I consider this suppletion for number to be (residual) morphology based on the importance of the number distinction in other languages of the family, and identify a morphological span based on it. If suppletion is instead viewed as a matter of choice between two distinct lexical items, diagnostic [19] becomes a replacement test of sorts, where complex verbal bases are replaced by simple ones, again underscoring the validity of this intuitive span.

The morpho-phonological processes that I have discussed so far in this section could be described as lexical and structure-preserving. Further domains could be identified with reference to post-lexical or structure-filling processes, though I know of few such processes that apply over spans longer than a single syllable. Stress assignment is one, and I have defined diagnostics based on stress in the following way: diagnostic [22] identifies the largest span on which only one primary stress occurs (11-16), diagnostic [23] identifies the domain on which the position of stress is calculated (11-16; stress is final in this domain).

One last phonological process that is relevant for identification of domains is vowel epenthesis. Epenthesis is claimed to happen domain-finally after all coda segments in Stout & Thomson (1974), but in our own data this is only consistently the case after stem-final /r/ if final in a domain (11-20), [24]. If medial in the domain, /r/ will obligatorily resyllabify if followed by a vowel, without any epenthesis occurring; epenthesis still applies medially if /r/ is followed by a consonant in the Kayapó dialect of Měbêngôkre, though in the Xikrin dialect it strengthens to a dental stop with the same voicing and nasality features as the following consonant. A process of simplification of other consonant sequences (*mex jarē* ‘praise’, lit. ‘say good’ → [metʃa'rē]) likely applies in the same domain, but I lack precise data to confirm this.

## 7 Conclusion

Table 4 summarizes the results of applying the diagnostics described above to the planar structure. As can be seen in the table, there is a rather strong convergence of diagnostics that identify a span going from position 12 (or 10, or 11) to position 15. That position 16 is not included might be an artifact of the impossibility of applying many of the diagnostics to that position, filled by a lone consonant, or by my privative definition of non-finiteness, when in reality all verbal predicates should be classified as either finite or non-finite. This span is a good candidate for the verbal word in Měbêngôkre, and approximately coincides with the *verb stem* that I had implicitly defined in previous work and in my conventions for transcription. The indeterminacy of the left edge of this span is a matter that was

discussed briefly above: here, grouping seems to be less a matter of wordhood or constituency, but of semantic affinity or selection. An orthogonal set of diagnostics could be applied (and in fact were applied in the preparation of Salanova 2017c, even if not included in that publication) to test the affinity among peripheral elements and particular verbs.

Table 4: Application of the diagnostics to the verbal planar structure

		POSITIONS IN THE TEMPLATE																							
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
DIAGNOSTICS	1																								
	2																								
	3																								
	4																								
	5																								
	6																								
	7																								
	8																								
	9																								
	10																								
	11																								
	12																								
	13																								
	14																								
	15																								
	16																								
	17																								
	18																								
	19																								
	20																								
	21																								
	22																								
	23																								
	24																								

## Abbreviations

1	first person	NFIN	non-finite
2	second person	NFUT	non-future
3	third person	NOM	nominative
ACC	accusative	PAUC	paucal
ANTIC	anticausative	PL	plural
ANTIP	antipassive	POSS	possessive
DIM	diminutive	PROG	progressive
EMPH	emphasis	PROSP	prospective
ERG	ergative	PRS	present
FIN	finite	REL	relationalizing prefix
FUT	future	SS	same-subject
INCL	inclusive		conjunction
NEG	negative	TR	transitivity prefix

## References

- de Oliveira, Christiane Cunha. 2005. *The language of the Apinajé people of Central Brazil*. University of Oregon. (Doctoral dissertation).
- Kuroda, Sige-Yuki. 1972. The categorical and the thetic judgment. *Foundations of Language* 9. 153–185.
- Nikulin, Andrey. 2020. *Proto-Macro-Jê: Um estudo reconstrutivo*. Universidade de Brasília. (Doctoral dissertation).
- Nikulin, Andrey & Andrés Pablo Salanova. 2019. Northern Jê verb morphology and the reconstruction of finiteness alternations. *International Journal of American Linguistics* 85(4). 533–567.
- Nonato, Rafael. 2014. *Clause chaining, switch reference, and coordination*. Massachusetts Institute of Technology. (Doctoral dissertation).
- Sala, A. M. 1920. Ensaio de gramática Kaiapó. *Revista do Museu Paulista* XII. 395–429.
- Salanova, Andrés Pablo. 2007. *Nominalizations and aspect*. Massachusetts Institute of Technology. (Doctoral dissertation).
- Salanova, Andrés Pablo. 2008. Uma análise unificada das construções ergativas do Mëbêngokre. *Amérindia* 32. 109–134.
- Salanova, Andrés Pablo. 2011. A flexão de terceira pessoa nas línguas Jê. *LIAMES* 11. 75–114.

- Salanova, Andrés Pablo. 2014a. Reduplication and verbal number in Mëbengokre. In Hein van der Voort & Gale Goodwin Gómez (eds.), *Reduplication in Indigenous languages of South America*, 247–272. Leiden: Brill.
- Salanova, Andrés Pablo. 2014b. Sintaxis aplicativa, semántica causativa. In Francesc Queixalós, Stella Telles & Ana Carla Bruno (eds.), *Incremento de valencia en las lenguas amazónicas*, 154–190. Bogotá: Instituto Caro y Cuervo.
- Salanova, Andrés Pablo. 2017a. Ergativity in Jê languages. In Jessica Coon, Diane Massam & Lisa Demena Travis (eds.), *Oxford Handbook of Ergativity*, 1065–1088. Oxford: Oxford University Press.
- Salanova, Andrés Pablo. 2017b. O sintagma nominal como estrutura predicativa. In Francesc Queixalós & Dionei Moreira Gomes (eds.), *O sintagma nominal em línguas amazônicas*, 203–239. Brasília: Pontes.
- Salanova, Andrés Pablo. 2017c. The paradoxes of Mëbêngôkre’s analytic causative. In Heather Newell, Máire Noonan, Glynne Piggott & Lisa Travis (eds.), *The structure of words at the interfaces*, 160–184. Oxford: Oxford University Press.
- Salanova, Andrés Pablo. 2020. Counting and measuring in Mëbêngôkre and the count/mass distinction. *Linguistic Variation* 2(20). 300–11.
- Salanova, Andrés Pablo & Andrey Nikulin. 2020. A história que conta o léxico Mëbêngôkre. *Norteamentos (Sinop)* 13(33). 52–106.
- Salanova, Andrés Pablo & Andrey Nikulin. in press. A typology of relationalizing and absolutizing morphology in Lowland South American languages. Manuscript submitted to *IJAL*.
- Stout, Mickey & Ruth Thomson. 1974. Fonêmica Txukahamëi (Kayapó). *Série Linguística* 3. 153–176.
- Tallman, Adam J. R. 2021. Constituency and coincidence in Chácobo (Pano). *Studies in Language* 45(2). 321–383. DOI: 10.1075/sl.19025.tal.
- Verswijver, Gustaaf. 1982. The intertribal relations between the Juruna and the Kayapo Indians (1850-1920). *Jahrbuch des Museums für Völkerkunde zu Leipzig* 36. 305–315.
- Verswijver, Gustaaf. 2018. *The club-fighters of the Amazon. Warfare among the Kayapo Indians of Central Brazil*. 2nd edn. Almería: Turuti Books.