Abstract

Tuparí, an indigenous Brazilian language of the Tupian family, has innovated a highly productive finite embedded clause construction that retains the morphosyntactic hallmarks of matrix clauses – without any neutralization in tense or evidentiality. I offer a synchronic analysis of these finite embedded clauses and propose a specific grammaticalization pathway that can account for their diachronic emergence: the clausal nominalizer hè developed out of a homophonous third person pronoun, allowing for paratactic constructions to be reanalyzed as involving true subordination. Both functions of hè (as a pronoun and as a clausal nominalizer) remain in use today, giving rise to occasional ambiguity. An additional aim of this paper is to evaluate the Tuparí facts in light of the literature on the Final-over-Final Condition (FOFC), a proposed syntactic universal. I will show that the language’s embedded clauses are unexpected on the most restrictive formulation of FOFC (Holmberg 2000) but can be accommodated without issue once FOFC is restricted to apply within Extended Projections. Situating Tuparí in the broader FOFC typology allows for a more fine-grained understanding of the distribution of the categorial features [+nominal] and [+verbal] in the language’s syntax.
1 Introduction

This paper offers a theoretically and typologically informed examination of finite embedded clauses in Tuparí, an indigenous language from the Brazilian Amazon. These finite embedded clauses retain the morphosyntactic hallmarks of matrix clauses, without any neutralization in tense, aspect or evidentiality. They are a historical innovation unique to Tuparí within the Tuparían branch of the Tupían family: no comparable constructions are attested in any of the sister languages, which rely on non-finite subordination strategies only. Hence in this paper I aim not only to provide a synchronic analysis of Tuparí embedded clauses but also to propose a specific grammaticalization pathway that can account for their diachronic emergence. I argue that the clausal nominalizer hè developed out of a homophonous third person pronoun, allowing for paratactic constructions to be reanalyzed as involving true subordination. Both functions of hè remain in use in contemporary Tuparí speech, giving rise to occasional ambiguity. Finally, this paper examines the Tuparí facts in light of the typological findings concerning the Final-over-Final Condition (FOFC), which was put forth by Holmberg (2000). Situating Tuparí in the broader FOFC typology allows for a more fine-grained understanding of the distribution of the categorial features [+nominal] and [+verbal]. I will argue that the C projection in Tuparí is unlike the explicitly [+verbal] categories of tense, evidentiality and aspect in that it lacks a categorial feature altogether.

Tuparí belongs to the Tuparían branch of the Tupían language family, which is among the most internally diverse and geographically dispersed families in all of South America (Rodrigues & Cabral 2012; Eriksen & Galucio 2014; van der Voort 2015). At the beginning of the European invasion, Tupían languages were spoken all along the Amazon River proper, on the Atlantic coast, and in much inland territory as well. Tuparí today has 350 fluent speakers, all of whom live in the Brazilian state of Rondônia. This state, which borders Bolivia, is considered the likely homeland of the Tupían-speaking peoples: it is here that the greatest number of the family’s branches are found – and in dense geographic concentration, as well (Vander Velden 2010). The Tuparí language remains vital in several villages on the Rio Branco Indigenous Territory, which is home to majority of the ethnic Tuparí; this is where I have carried out my own fieldwork. A number of ethnic Tuparí also live on the Rio Guaporé Indigenous Territory, located several hundred kilometers to the west; however, intergenerational transmission of the language is reported to have ceased there. Maps are provided in Figures 1 and 2.

Figure 1 The Brazilian state of Rondônia, located in the country’s Northern Region (from Google Maps).
Prior to my own fieldwork, which began in 2013 at the invitation of two indigenous teachers, the available descriptive materials on the language discussed lexicon, phonology and morphology (Caspar & Rodrigues 1957; Seki 2001; Alves 2004). The syntactic generalizations reported in this paper have not been described in print before, though some data involving finite embedded clauses were provided in our prior publications (Singerman 2019: 424–428).

All examples are given in a four-line format, with the first line showing the standard orthography used by the Tuparí-speaking schoolteachers on the Rio Branco and followed in Tupari et al. (2016). For certain examples I also provide the discourse context. Relevant metadata are provided for each example: for conversational excerpts, the date on which the example was uttered; for texts, the name of the author/narrator. In keeping with a broader commitment to rely on naturalistic material, I use non-elicited utterances whenever possible. By doing so I hope to demonstrate the degree to which finite embedded clauses have taken hold in the language. Relying on natural data also makes it possible to highlight the contexts where the morpheme hè could be parsed either as the innovative clausal nominalizer or as the third person pronoun from which that nominalizer has grammaticalized.

The paper opens in §2, which provides an overview of the structure of matrix clauses in Tuparí. §3 then discusses the major morphosyntactic properties of finite embedded clauses and the uses to which those clauses are put. §4 presents several diagnostics that demonstrate the extent to which these finite embedded clauses maintain the syntactic properties and inflectional categories of matrix clauses. With this synchronic picture in place, §5 contrasts the clausal nominalizer hè against the homophonous third person pronoun and argues that formerly paratactic constructions were reanalyzed as involving true subordination. Finally, §6 situates the Tuparí facts within the typological landscape of syntactic headedness.

2 The structure of Tupari matrix clauses

This section summarizes the defining properties of matrix clauses in Tuparí, with attention paid both to finiteness categories (clause type, tense, evidentiality) and to the distribution of head-
In all four examples the direct object, ewaet ‘your hammock’, immediately precedes the transitive verb āpe ‘hang up’. Object-verb order is absolute in the Tupari VP; indeed, this head-final syntactic property is shared by all the languages belonging to the Tupari an branch of the Tupían family (Galucio 2001; Braga 2005; Aragon 2014; Nogueira 2019). Examples (b) and (d) additionally contain an auxiliary from the aux go series, so named because of a diachronic connection with the lexical verb ‘go’; like all of the language’s auxiliaries, these must follow rather than precede the lexical VP. Hence the Tupari AuxPs are head-final, too, just like the object-verb VPs beneath them.

The four utterances in (1) feature the second position particle =mākērō ‘DON’T.KNOW’, which converts wh-questions into expressions of ignorance or doubt on the speaker’s part. It is but one of a larger set of clause-typing particles, given in Table 1, all of which encliticize onto the clause-initial syntactic constituent. These particles are sensitive to whether the clause-initial constituent is [+wh] or [−wh]; only nāpe ‘EMPHATIC’ is indifferent. These particles also have considerable effects on the interpretation and availability of the non-witnessed evidential suffix -pnẽ/psira. Given that the clause-typing particles instantiate the highest functional level in the Tupari clause and are sensitive to the [±wh] status of the clause-initial XP, we analyze them here as the realization of a high, head-initial C projection.
In between the head-final VP and AuxPs, on the one hand, and the head-initial CP, on the other, we find the TP – a projection that is realized in heterogeneous ways. Returning to the four-way paradigm in (1), tense is marked by the near past suffix -t/n in (a) and (b) but by the distant past particle in (c) and (d). Despite the fact that near past -t is a suffix on the predicate while distant past =õpot is an enclitic located in second position, speakers treat these morphemes as paradigmatically contrastive members of a single grammatical class. An additional similarity between these two morphemes (one a second position particle, the other a predicate-final suffix) is that both will combine with members of the AUX₇ series of auxiliaries to create intermediate temporal interpretations. In (1b), the near past suffix -t combines with tero’e ‘AUX₇,SG’. The result is a temporal interpretation intermediate between (1a) and (1c): the hammock hanging event took place at least a few months before UT and possibly as far back as a full year or two. And in (1d), the distant past particle =õpot combines with the auxiliary tet’e (an allomorph of tero’e) to achieve a temporal interpretation more remote than (1c): the hammock hanging event took place many, many years before UT.

The distant past particle =õpot and the near past suffix -t/n are part of a larger set of tense morphemes. The durative -pbi’a (used for past habitual readings and also, on occasion, for present habitual ones) patterns like near past -t in that both are suffixes that attach at the right edge of the predicate. In a similar way, =ko∼ke ‘polite.fut’ (which indicates a degree of deference to the addressee and is often employed to make requests or issue polite commands) and =kut ‘ANCIENT.PST’ (which is used for events that took place prior to the speaker’s birth) pattern like =õpot ‘DISTANT.PST’ in that they, too, sit in second position. Despite this surface heterogeneity, all of these morphemes are in paradigmatic contrast with one another. For this reason I have analyzed them as instantiations of the same syntactic head, T. Textual evidence shows that these morphemes pattern alike in running discourse, as well (Singerman 2018b: 298–303). Note that there are also several tense auxiliaries used for the future and present tenses. Like all other auxiliaries in this language they pattern as head-final; that is, they follow rather than precede their complements. Detailed discussion of all of the language’s tense morphology is provided in Singerman (2018b: chapters four and five) and in Singerman (2020).

Examples (1c) and (1d) contain both a clause-typing particle (=mākerõ ‘DON’T.KNOW’) and a second position tense particle (=õpot ‘DISTANT.PST’). As those utterances demonstrate, when a single clause contains both a clause-typing particle and a tense particle, the two must occur in exactly that order and without any intervening material. (2) provides additional examples of utterances that contain both kinds of particles. Observe the strict order and linear adjacency between the particles.

(2) Clause-typing particles and tense particles are adjacent in second position

a. Epoanerõ’omka e’a nākop ko ’on
   [ e-poanẽ-ro-’om-ka-a e-’a ] = nākop = ko = ’on
   [ 2SG-get.better-NMZ-NEG-VBZ-TH 2SG-if.SG ] = MAYBE = POLITE.FUT = 1SG
   emākakap.
e-māk-ap
2SG-send-ADV.FOC

‘If you-SG don’t get better, maybe I will send you [from the village to the city].

conversation: 2018-08-06

| Particle | Gloss     | Function                          | Clause-initial XP |
|----------|-----------|-----------------------------------|-------------------|
| nē       | YES/NO    | builds polar questions            | [-wh] only        |
| mākerõ   | CONFIRMATIVE | builds tag/biased polar questions | [-wh] only        |
| nākop    | MAYBE     | turns propositions into statements of doubt | [-wh] only |
| ’et      | SADLY.NOT | expresses regret that some event failed to take place | [-wh] only |
| pa’a / ta’a | ASSERTIVE | creates extra-assertive declaratives; allomorphy indexes speaker gender | [-wh] only |
| mākerõ   | DON’T.KNOW | turns content questions into statements of ignorance | [+wh] only |
| nāpe     | EMPHATIC  | builds questions with extra emotive content | [+wh] or [-wh] |
Singerman (2020) proposes that the tense particles arrive in second position via an application of T-to-C Head Movement (Travis 1984 and much subsequent work). The Tuparí CP is an invariably head-initial projection, with a single XP occupying its specifier position; as a result, the application of T-to-C Head Movement brings the tense particles to second position, as well. Note that no Head Movement applies with the tense suffixes -t ‘near.pst’ and -pbi’a ‘durative’, which attach to the right edge of the predicate; in separate work we analyze these suffixes’ surface position as the result of an application of the post-syntactic operation of Lowering (see Embick & Noyer 2001; Harizanov & Gribanova 2019 and other research within the Distributed Morphology framework). The crucial takeaway is the following: the second position particles appear to instantiate a head-initial TP, whereas the predicate-final suffixes would appear to instantiate a head-final one. If we are correct in analyzing both of these sets of morphemes as undergoing dislocation processes to arrive at their surface positions, then neither set reveals the underlying headedness of the TP; rather, that headedness remains indeterminate. Following Singerman (2020), (3) and the other trees in this paper indicate the indeterminate headedness of the TP using dotted lines.2

(3) The structure of matrix clauses in Tuparí; following Singerman (2020), the indeterminate underlying headedness of the Tense Phrase is indicated by dotted lines

The Tuparí Evidential Phrase sits in between the highest auxiliary projection and the TP, as in the tree in (3). The EvidP is realized overtly by the suffix pnẽ/psira, which agrees with the subject in number. While the TP has, at least on the surface, both head-initial and head-final realizations, the EvidP is consistently head-final: it is realized as a suffix on the highest head in

---

2 There is strong language-internal evidence that subjects in Tuparí are generated in a low position, in keeping with the VP-Internal Subject Hypothesis (Koopman & Sportiche 1991; McCloskey 1997), and arrive in the clause’s left periphery via movement. To keep things simple, the trees in this paper show the subject as generated in Spec,V rather than Spec,v.
the predicate complex. (4) shows a minimal pair for evidentiality. In both utterances the clause-initial XP is the NP subject. The lexical VP, which describes a kinship relation, is followed by two auxiliaries: tero’a and te’eka. Tense is marked with the second position particle =õpot. The absence of -pnẽ ‘EV.SG’ in (a) yields an obligatory [+ WITNESSED] interpretation.

(4) Minimal pair for the witnessed/non-witnessed evidential contrast
a. Isìt õpot o’apaynā tero’a
   i-si-t =õpot o-‘apay-nē-a tero’e-a
3-mother-NUC =DISTANT.PST 1SG-paternal.aunt-VBZ-TH AUX—a—SG-TH
   te’eka.
   te-eka-a
3C-AUXwāir—SG-TH
‘His mother was my paternal aunt (WITNESSED).’ [the speaker knew her aunt]
conversation: 2018-08-02

b. Pamẽkgen õpot möket malokare ototonā
   Pamẽk-en =õpot möket maloka-re o-toto-nē-a
   Pamẽk-NUC =DISTANT.PST long.ago maloka-OBL 1SG-grandfather-VBZ-TH
   tero’a te’ekapnā.
   tero’e-a te-’eka-pnẽ-a
   AUX—a—SG-TH 3C-AUXwāir—SG-EV.SG-TH
‘Pamẽk was my grandfather/male ancestor long ago in the maloca (NON-WITNESSED).’ [the speaker did not know Pamẽk]
conversation: 2017-08-04

When there is no auxiliary present, then the evidential suffix will attach to the lexical verb itself:

(5) Evidential suffix attaches to the lexical verb when there’s no auxiliary
Easat mākẽrõ te’ausi patnan?
e-asa-t =mākẽrõ te-a’usi pat-nē-a-n
2sg-old.older.brother-NUC =CONFIRMATIVE 3C-wife marry-EV.SG-TH-NEAR.PST
‘Your-SG older brother got married (NON-WITNESSED), right?’
conversation: 2017-08-04

In this example the near past suffix -t/n is separated from the evidential suffix by the ‘theme vowel’ -a, a morphological linker element whose distribution is not relevant for the present discussion. Note that no evidential morphology is present in the four-member paradigm in (1) because the clause-typing particle =mākẽrõ ‘DON’T.KNOW’ neutralizes the language’s witnessed/non-witnessed contrast. (The clause-typing particle =nākop ‘MAYBE’ does so as well.)

We have now seen that matrix clauses in Tuparí are characterized by three layers of syntactic headedness. Head-initiality reigns at the top of the clause, in the CP layer, whereas head-finality obtains between the VP and the EvidP. In between these two layers is the TP, which exhibits a mixed set of properties. Before turning to the structure of finite embedded clauses, in §3, I wish to comment on evidence for null functional morphology in Tuparí. Although the glosses in this paper generally omit null morphemes, there is evidence for at least two different null tense heads. One of these provides an immediately-before-UT interpretation and is located in second position – just like the particles =õ pot ‘DISTANT.PST’, =ko∼ke ‘POLITE.FUT’ and =kut ‘ANCIENT.PST’. Positing this null particle helps to explain the distribution of the nominative enclitics. Consider the distribution of =en ‘2SG’ in the four-utterance paradigm in (1): ‘en occurs at the far right edge of the predicate, following =‘NEAR.PST’, in (1a) and (1b) but sits in second position, following =õ pot ‘DISTANT.PST’, in (1c) and (1d). That is, it is linearly parasitic on the tense morphology. Now take a look at (6a) and (6b). In both =on ‘1SG’ surfaces in second position. Since the nominative enclitics linearly follow the tense morphology, a null tense morpheme must sit in second position in (b) – just as overt =ko ‘POLITE.FUT’ does in (a).

3 There is no pronominal prefix attached to the auxiliary in tero’a in (4a) and (4b) because of a morphological haplology constraint that bars te-‘3C’ prior to the singular allomorph of the AUX—a—series (and prior to the singular allomorph of the lexical verb ‘go’). If the auxiliary root here were pā’e rather than singular tero’e, the third person pronominal prefix te- would be overt. The same haplology constraint is responsible for the lack of te- prior to tero’e in examples (15c), (17b) and (45).
Evidence for a null tense particle located in second position

a. Nĩka etet’ero’are ko ‘on watoa
  [ ∅-nĩk-a e-tet’e-ro’are ] = ko = ‘on w-ato-a
  [ 3-write-TH 2SG-AUX,SG-while ] = POLITE.FUT = 1SG 1SG-bathe-TH
  owáram.
o wan-am
1SG-go.nearby-ADV.FOC
‘Let me go a short distance to bathe while you-SG are writing it down.’
WhatsApp: 2018-01-22

b. E’era eyěrō’are ‘on
  [ e-‘et-a e-yē-ro’are ] = ∅ = ‘on
  [ 2SG-sleep-TH 2SG-AUX,SG-while ] = IMMEDIATE.PAST = 1SG
  waorosap.
w-aoros-ap
1SG-arrive.SG-ADV.FOC
‘I arrived [just now] while you-SG were sleeping, lying down.’
conversation: 2016-11-15

That the nominative enclitics most commonly occur in second position in superficially tenseless clauses indicates that such clauses contain a null second position tense particle, per the gloss in (6b).

3 Finite embedded clauses

Embedded clauses in Tuparí come in two flavors. First, there are older non-finite nominalizations with cognates found throughout the Tupian family. Second, there is an innovative finite embedded clause construction that retains the morphosyntactic properties of matrix clauses. §3.1 provides examples of a few of the older, non-finite embedding strategies. In §3.2 our attention turns to the innovative finite embedded clauses and describes the morphology that the nominalizer he can exhibit. §3.3 then discusses the uses of finite embedded clauses, namely, as internally headed relatives and as headless relatives.

3.1 Conservative retentions: non-finite nominalization strategies

Tuparí retains several non-finite subordination strategies which are shared by the other members of the Tuparian branch of the Tupian family (see Galucio 2011a,b, 2014; Aragon 2014; Nogueira 2019; Galucio & Nogueira 2018). Crucially, these strategies are non-finite: they never contain tense, aspectual, evidential, or clause-typing morphology.

The pair of utterances in (7) shows the actor nominalizer -at/an. The nominalization tepuop’orap kot’oaret ‘one who wants to learn’ serves as the subject in (b), behaviorally identical to akurapap’at ‘spider’ in (a). Note the presence of the nuclear case, required on non-pronominal subjects.

4 The only exception to this generalization is that the nominalizer -ap/am, referred to as a ‘circumstantial nominalizer’ in the literature on the Tuparian languages, can attach at a syntactic height above that occupied by aspectual auxiliaries (Singerman 2019: 436–440).

5 The distant future auxiliaries require their VP complement to be nominalized with -ro/to, as shown in (7):

6 A reviewer asks about the distribution of the nuclear case suffix -et/en/t/n. This suffix is required on all non-pronominal subjects, such as isìt ‘his mother’ in (4a), the proper name Pamẽkgen in (4b), akurapap’at ‘(the) spider’ in (7a) and the deverbal nominalization tepuop’orap kot’oaret ‘the one who wants to learn’ in (7b). In this sense the suffix looks like a marked nominative; however, it can also occur on non-pronominal direct objects that are not discourse new. Importantly, it never occurs on focused NPs, as shown by various examples in §5. For this reason I have previously described the nuclear case as a kind of ‘anti-focus’ marker, sensitive to both grammatical relations and information structure (see Singerman 2018b: 64–71, which raises challenges for the alternate analyses advanced by Alves 2004 and Cabral et al. 2017). Note that utterances like the two in (7) – with a nuclear-marked NP subject serving as the clause-initial syntactic constituent – are never interpreted with focus on the subject. §5 provides more examples demonstrating the nuclear case’s sensitivity to information structure.
The actor nominalizer suffix

a. Akurap'at temeren koro pete’a.
   akurap’a-t te-men-en ko-ro pete’a
   spider-NUM 3C-husband-NUM eat-NMZ DISTANT.FUT + 3SG
   ‘The spider will eat its own husband.’
   conversation: 2017-08-17

b. Tepuop’orap kot’oaret sorowaro pete’a.
   te-puop’orap kot’oy-at-et s-orowa-ro pete’a
   3c-learning want-NMZ,actor-NUM 3-search.for-NMZ DISTANT.FUT + 3SG
   ‘The one who wants to learn [lit.: the learning wanter] will search for it.’
   conversation: 2018-08-30

The pair of utterances in (8) makes the same point for the object nominalizer prefix iy/y-, which converts a transitive verbal root (here, pek ‘buy, ask for’) into a nominal that can then be possessed (here, by the NP eop pẽan ‘your older paternal uncle’). Hence eop pẽan iypekgere ‘from your paternal uncle’s bought thing / purchase’ in (8b) behaves just like the non-derived oblique NP pot’are ‘from the boar’ in (8a).

The object nominalizer prefix

a. Eop koro’om nẽ ’en pot’a-re?
   e-arop ko-ro’-om =nẽ =’en pot’a-re
   2SG-food eat-NMZ-NEG =YES/NO =2SG boar-OBL
   ‘Have you-5G not eaten your food, from the boar meat?’
   conversation: 2016-02-08

b. Eop pẽan iypekgere nẽ earop koro’om eman
   e-op pẽan iy-pek-ere =nẽ e-arop ko-ro’-om eman
   2SG-father elder NMZ,object-buy-OBL =YES/NO 2SG-food eat-NMZ-NEG still
   ’en?
   =’en
   =2SG
   ‘Have you-5G still not eaten your food, from your paternal uncle’s bought thing / purchase?’
   elicitation: 2017-08-02

The suffix -at/an ‘NMZ,actor’ and the prefix iy/y- ‘NMZ,object’ attach to small pieces of syntactic structure – ones that lack the Evid, T, and C projections shown in (3). The same applies to the other nominalizing affixes discussed in our prior publications. What is more, these affixes all have cognates in the other members of the Tuparí an branch of the Tupían family (Galucio & Nogueira 2018) and in more distant Tupían branches, as well (see da Cruz & Praça 2019 on Tupi-Guaraní). So they qualify as conservative retentions, ones that reconstruct all the way to Proto-Tupían.

3.2 Major structural properties of finite embedded clauses

The distinguishing formal feature of finite embedded clauses in Tuparí is that they always bear the morpheme hè—he at their right edge. Evidence will be presented that hè is a nominalizer, just like the affixes discussed in the previous subsection. But hè differs from those affixes in that it scopes over whole clauses – clauses which can contain tense and evidential morphology as well as a subset of the clause-typing particles. This subsection discusses how hè may take the full range of nominal morphology and can perform the same syntactic roles that non-derived nominals do.

Example (9a) shows a fully-formed declarative clause that contains overt tense and evidential morphology. In (9b), the addition of hè converts the matrix declarative clause into an NP that gets interpreted as an internally headed relative. (9c) provides the full sentential context.

Matrix declarative converted into an embedded clause by the nominalizer hè

a. Tarupat teotsirat.
   tarupa-t te-ot-sira-a-t
   non.indigene-NUC 3C-go.PAUC-EV,PL-TH-NEAR.PST
   ‘The non-indigenous people left (NON-WITNESSED).’
b. tarupat teotsirat hé
[ tarupa-t te-ot-sira-a-t ] hé
[ non.indigene-NUC 3C-go.PAUC-EV.PL-TH-NEAR.PST ] HÉ-NUC
‘the non-indigenous people who left (NON-WITNESSED)’

The nominalizer hé in (c) bears the nuclear case, as required of all non-pronominal subjects (see footnote 6). The entire embedded clause serves as the clause-initial XP in this utterance, as demonstrated by the fact that it immediately precedes the second position particle =né ‘YES/NO’. Hence (9c) is parallel to (10), where the subject is Kopere ema’en ‘the language of the Djeoromitxí’:

(10) NP subject in clause-initial position, prior to the second position clause-typing particle
Kopere ema’en nê nam erop’a?
[ Kopere ema’ẽ-n ] =nê nam erop’a
[ NP Djeoromitxí language-NUC ] =YES/NO difficult
‘Is the language of the Djeoromitxí people difficult?’
conversation: 2017-08-16

Since =né and the other clause-typing particles always encliticize onto the clause-initial XP, example (9c) shows that a finite embedded clause built with hé functions as a single syntactic constituent.

The nominalizer hé can take the full range of case morphology. Several intransitive verbs take optional arguments that are marked with the instrumental suffix -p/m. One such predicate is apsikat ‘think, think about, remember’. In (11a), the instrumental suffix attaches to ha ‘this place, here’; in (11b), to both the pronoun en ‘2sg’ and the clausal nominalizer hé.

(11) The verbal predicate apsikat ‘think about’ takes an instrumental-marked argument
a. Èkgo eaora etera e’a nê
[ e-ek-o e-aot-a e-tet-a e-’a ] =nê
[ 2SG-house-INS 2SG-leave.SG-TH 2SG-go.SG-TH 2SG-when.SG ] =YES/NO
ke ‘en ham eapsikatsam?
=ke ‘en ha-m e-apsikat-sẽ-am
=POLITE.FUT =2SGN this.place-INS 2SG-think-RSLT_{yesterday}.SG-ADV.FOC
‘When you-SG arrive at your home, are you going to think about this place?’
conversation: 2016-02-17

b. Wapsikara ‘on éró, ma’ã ‘en herōwap hem.
w-apsikat-a =‘on en-o [ ∅-ma’ã-a =‘en herōwap ] hé-m
1SG-think-TH =1SG 2SG-INS [ 3-speak-TH =2SG yesterday ] HÉ-INS
’I thought about you-SG, about what you-SG said yesterday (WITNESSED).’
conversation: 2017-08-14

The distribution of the instrumental suffix here shows that embedded clauses built with hé belong to the same morphosyntactic category as nominal roots like ha ‘this place’ and pronouns like en ‘2SG’.

In more conservative Tupari speech, right-dislocated direct objects bear two case suffixes: nuclear -et/ẽ/en and locative -pe, in that order (see also 2b, in §2). Right-dislocated objects must be resumed in situ by a pronominal proclitic, incorporated object, or full NP. This is shown by (12a), where the right-dislocated object is ouoka iaetpe ‘my water container’; the third person proclitic s- attaches to the verb at ‘get’. When the right-dislocated object is an embedded
The embedded clause in (b) is fully finite: it contains the non-witnessed plural evidential suffix and the second position tense particle =kut ‘ANCIENT.PST’.

Possessors and the complements of postpositions are morphologically bare, as shown by (13). The speaker of this two-sentence utterance uses the postposition tere ‘on, by means of’ first with a simple nominal and then with an entire finite embedded clause. As expected, the nominalizer hè in (b) is bare – just as moto ‘motorcycle’ is in (a).

Nominal predicates resemble possessors and the complements of postpositions in that they, too, lack any overt morphological marking. Hence when a finite embedded clause serves as a predicate, hè is bare. Compare the simplex nominal predicates in (14a) and (14b) against the embedded clause in (14c):

(14) **Example of a finite embedded clause serving as a nominal predicate**

a. Tupari ‘on.

   Tupari =’on

   Tuparí =1SG

   ‘I am Tuparí.’

   everyday speech

b. Amẽko eri’at ‘on.

   amẽko eri’at =’on

   dog owner =1SG

   ‘I am the owner of a dog.’

elicitation: 2015-10-10
I now turn to two final morphological parallels between finite embedded clauses and other nominals. Negative/privative -‘om is a strictly nominal suffix: it attaches only to nominal bases, never to verbal ones (Singerman 2018a). This means that a verbal root must first be nominalized if it is to be negated. (See 23b, 29a and 31 for examples where -‘om attaches to puop ‘smart, knowledgeable; know’, which is formally a nominal, and 2a, 8, 28 and 41a for examples where -‘om attaches to a nominalized verb.) It is telling that hè can be negated with -‘om, just as any other nominal can.

(15) The negative/privative suffix -‘om attaches to nominal bases only

a. Tupari’ommẽ.
   Tuparí-‘om = e
   Tuparí-NEG = 3
   ‘He/she isn’t Tuparí.’
   everyday speech

b. Hare kiaripotkarat’ommẽ.
   hare ki-ari-potkat-at-‘om = e
   here IMPRSG.GO.hungry-NMZ,ACEG-NEG = 3
   ‘Here no one goes hungry.’ / ‘Here there is no one who goes hungry.’
   conversation: 2018-08-15

c. Mõket tero’apbi’ae he’ommẽ.
   [ mõket tero’e-a-pbi’a = e ] he-‘om = e
   [ long.ago exist.SG-DURATIVE = 3 ] HÈ-NEG = 3
   ‘The one that long ago used to exist (witnessed) is no more / no longer exists.’
   WhatsApp: 2021-01-14

In (15a) the negative/privative -‘om attaches to the monomorphemic nominal root Tupari; in (15b), to the deverbal actor nominalization kiaripotkarat ‘one who goes hungry’. And in (15c), in which the speaker describes an old house of hers that has since been demolished, -‘om attaches to hè. That embedded clauses built with hè can be directly negated by -‘om further testifies to their status as nominalizations.

An additional parallel between finite embedded clauses and other nominals concerns number. Third person NPs in Tuparí usually lack number marking; one must look to verbal agreement to determine the intended interpretation. For instance, the paucal verbal root at ‘GO.PAUC’ and the plural evidential suffix in (9a) force a paucal interpretation of the NP subject tarupat, which could in principle be interpreted either as ‘the non-indigenous person’ or as ‘the non-indigenous people.’ There is however an optional plural-like suffix, -‘eat ‘many’. Like negative/privative -‘om, it is restricted to nominal bases.

(16) The plural-like suffix -‘eat attaches to nominal bases only

a. Èop tero’a né ‘en ote’earere?
   e-eop tero’e-a =nẽ =‘en ote-‘eat-ere
   2SG-get.used.to AUXSG-SG-TH = YES/NO = 2SG 1PL.EXCL-MANY-OBL
   ‘Have you-SG grown used to all of us-excl?’
   conversation: 2016-12-15

b. Wappe kut Tupari’earet te’era sakapsira.
   wap-pe =kut Tuparí-‘eat-et te’et-a s-aka-psira
   hammock-LOC =ANCIENT.PST Tuparí-MANY-NUC 3C-sleep-TH 3-AUXV-newer,PL-EV.PL
   ‘The Tupari used to sleep in hammocks (nonwitnessed).’
   conversation: 2017-08-03
Plural-like -'eat can attach to a pronoun (ote '1PL.EXCL’ in 16a), to a monomorphemic nominal (the ethnynym Tupari in 16b), and – as shown by (c) – to the clausal nominalizer hè. The embedded clause in that example is fully finite, with overt tense, evidential and aspectual morphology; it serves as a right-dislocated subject.

The examples in this subsection have demonstrated that finite embedded clauses behave just as non-derived nominals do. They may bear the full range of case suffixes and may serve in any and all syntactic roles, ranging from sentential subjects to the complement of postpositions.

They can also be negated by -'om and pluralized with -'eat, both of which are strictly nominal suffixes.

### 3.3 Uses of finite embedded clauses

Finite embedded clauses in Tuparí are frequently employed as INTERNALLY HEADED RELATIVE CLAUSES, a construction attested in various languages of the Americas (Platero 1974; Gorbet 1976; Cole 1987; Williamson 1987; Basilico 1996; Salanova 2011; Boyle 2016; Gordon & Munro 2017; Hanink 2021), as well as Japanese and Korean (Hiraiwa 2017; Ohara 2018) and several languages of South Asia (see Subbārāo 2012: chapter six and references therein).

(17) shows internally headed relatives where the subject of an intransitive verb or an auxiliary serves as the head. Observe from (b) that there is no requirement that the internal head (in bold) be the initial XP within the embedded clause.

(17) Internally headed relatives where the head is the subject of an intransitive verb

a. het'oet kuret etere teyã hè
   [ het'op-et kut-et e-tere te-yê-a ] hè
   [ that-NUC child-NUC 2SG-on.top 3C-exist.sg-th ] hè
   'the child that is sitting/lying on top of you-SG'
   WhatsApp: 2017-07-24

b. here õpore wirik eri'aret tero'ap hè
   [ here =õpot =e wirik eri'at-et tero'e-ap ] hè
   [ then =DISTANT.PST =3 field owner-NUC exist.sg-ADV.FOC ] hè
   'the owner of the field that existed / the owner that the field had (WITNESSED)'
   text narrated by Iracema Taydyup Tupari

(18) shows internally headed relative clauses where the head is the subject of a transitive verb:

(18) Internally headed relatives where the head is the subject of a transitive verb

a. eamigot mensagem mäka te'ekat hè
   [ e-amigo-t mensagem mäk-a te-'eka-a-t ] hè
   [ 2SG-friend-NUC message send-TH 3C-AUX sg-th-near.pst ] hè
   'your-SG friend who used to regularly send messages (WITNESSED)'
   conversation: 2018-07-28
Example (b) demonstrates that the internal subject can be a pronoun, here the third person nominative enclitic =e (which must follow the tense particle =õpot ‘DISTANT.PST’; see §2).

A very common kind of internally headed relative in my corpus is one where the internal head is the direct object of a transitive verb.\(^8\) (19) provides four examples, with a mix of pronominal and non-pronominal objects.\(^9\)

\[(19) \text{ Internally headed relatives where the head is the object of a transitive verb}\]

\[\text{a. kuret atsā } \text{en ei rowaere } \text{hè}\]
\[[\text{kut-et} \text{ at-sē-a } =\text{en e-i rowap-ere } ] \text{hè}\]
\[\text{‘the child that you-.sg are holding, sitting down, in your photograph’}\]

\[\text{b. sítèynan } \text{hè}\]
\[[\text{s-itèy-nẽ-a-n } =\text{en } ] \text{hè}\]
\[\text{‘the thing that you-.sg brought (NON-WITNESSED)’}\]

\[\text{c. Vania yam } =\text{en y-om-nẽ-am } \text{hè}\]
\[[\text{Vania yam } =\text{en } ] \text{hè}\]
\[\text{‘the thing that you-.sg just gave to Vania (NON-WITNESSED)’}\]

\[\text{d. omemsiremsiren ipeknan } \text{hè}\]
\[[\text{o-memsiremsin-en i-pek-nẽ-a-n } ] \text{hè}\]
\[\text{‘the thing that my grandchild bought (NON-WITNESSED)’}\]

Examples (19b) through (19d) show the three overt allomorphs of the third person pronominal proclitic: s- before short oral vowels, y- (realized as [ɲ]) before nasal vowels, and i- elsewhere. Uniquely among the set of pronominal proclitics, the third person also has an optional null allomorph; it is available prior to consonant-initial verbs only (Singerman 2018b: chapter two). So the head of an internally headed relative may be null when it is the object of a consonant-initial verb, as in (20):

\[(20) \text{ Internally headed relatives where the head is null prior to a consonant-initial transitive verb}\]

\[\text{a. Josué aropnã poaroa o’e } \text{hè}\]
\[[\text{Josué aropnã } \text{∅-poaro-a } o-’e ] \text{hè}\]
\[\text{‘the thing that I put away for Josué’}\]

\[\text{b. wa’usipapet màksira } \text{hè}\]
\[[\text{w-a’usipap-et } \text{∅-māks-sira-a ] } \text{hè}\]
\[\text{‘the thing that my mother-in-law sent (NON-WITNESSED)’}\]

---

\(^8\) There are no ditransitive verbs or double object constructions in Tuparí.

\(^9\) The positional information in example (19a) (i.e., that the subject is horizontal) is encoded in the resultative suffix. This suffix makes a horizontal-vertical contrast with singular subjects; with plural subjects, that contrast is neutralized. See also examples (11a), (21a) and (53a). The Resultative Phrase itself sits above the lexical VP but beneath the lower of the two auxiliary projections; see (3) and, for more detailed discussion, Singerman (2019: §6).
The verb in (20b) bears the plural evidential suffix because in-laws are treated, as a matter of respect, as non-singulars.

There is a crucial difference between examples like (20a) and (20b) – where the head of the relative clause is the null object of a consonant-initial verb – and embedded clauses that are used as HEADLESS RELATIVES. The interpretation of the latter kind of relative clause implies a null adverbial element:

(21)  Finite embedded clauses used as headless relative clauses: location, time

a. totot tepsiksā teyā
   [ toto-t te-epsik-sē-a te-yē-a ]
   [ grandpa-NUC 3C-sit.down-RSLT,SG-TH 3C-AUX,SG-TH
   te-‘eka hē
te-‘eka-a ] hē
   3C-AUX,SG-TH ] HĒ
   'the place where grandpa had been sitting down earlier in the day'
   conversation: 2018-08-24

b. medikot owētõan here
   [ mediko-t o-wetom-a-n ] hē-re
   [ doctor-NUC 1SG-let.somebody.know-TH-NEAR,PST ] HĒ-OBL
   'at the time when the doctor let me know (WITNESSED)'
   conversation: 2018-08-15

The implicit adverbial can also be one of manner; this is indicated outside of the finite embedded clause itself by nēkat ‘resemblance, resembling’ or a derivation of it:

(22)  Finite embedded clauses used as headless relative clauses: manner

a. teapap’a haet porae hē nēkatke
   [ te-apap’a hap-et pore-a =e ] hē nēkatke
   [ te-head hair-NUC cut-TH =3 ] HĒ resembling
   'resembling the way that he cut his hair (WITNESSED)'
   conversation: 2017-08-29/30

b. tarupa aropnã òpot irik’enã te-puop’otnē hē
   [ tarupa aropnã =òpot irik’enē-a te-puop’ot-nē ] hē
   [ non.indigene for =DISTANT,PST work-TH 3C-learn-EV.SG ] HĒ
   nēkaremankia nēkaremankia
   perfectly.resembling
   'perfectly resembling the way that he had learned to work for non-indigenous people (NON-WITNESSED)'
   text narrated by Pedro Kup’eoyt Tupari

The finite embedded clause in (22a) is a verbal clause without overt finiteness morphology. Per the generalizations presented in §2, this yields an obligatory [+WITNESSED] evidential interpretation and an immediately-prior-to-UT temporal interpretation. Example (22b), meanwhile, contains both the second position tense particle =òpot ‘DISTANT,PST’ and the non-witnessed singular evidential suffix.

Similarly to the above examples of headless relatives, finite embedded clauses can be used with factive predicates such as puop ‘know, be knowledgeable about’. In such instances, too, there is no internal head. The nominalizer hē takes oblique morphology because of the selectional requirements of the matrix predicate puop; see also (29), in §4:

(23)  Headless relative serving as the oblique argument of puop ‘know, be knowledgeable about’

a. Puop =‘on [ i-et-et Abo ] he-re
   know =1SG [ 3-name-NUC Abo ] HĒ-OBL
   ‘I know that his name is Abo.’
   WhatsApp: 2018-04-17
3.4 Summary

This section has shown that Tuparí makes use both of non-finite nominalizations that stretch far back in the history of the Tupían family and of fully finite embedded clauses. The latter are marked with hê at their right edge and are employed as internally headed relatives, as headless relatives and as the complement of factives. In the next section we investigate the internal structure of these embedded clauses in greater detail. Evidence will be presented that they contain the very same functional projections found in matrix clauses: while only a subset of the clause-typing particles may be embedded, embedded clauses maintain all of the language’s tense, evidential and aspectual distinctions.

4 How much structure is there inside of finite embedded clauses?

In many languages embedded clauses exhibit different structural properties than matrix clauses. For instance, they may expose a more limited range of finiteness contrasts when compared to matrix clauses (Nikolaeva 2007; 2013). The availability of may also differ between matrix and embedded clauses, as famously shown by asymmetric Verb Second in Germanic.

This section asks to what degree Tuparí embedded clauses resemble / differ from matrix clauses. It turns out that these embedded clauses exhibit a host of root syntactic phenomena: except for reduced use of the clause-typing particles, they are indistinguishable from matrix clauses in many respects. We establish this point by examining the maintenance of second position effects (§4.1) and the retention of tense, evidentiality and clause type contrasts (§4.2). These generalizations lead us to the analysis in (24): a right-headed nominal projection, headed by hê, takes a complement that itself contains the same functional projections that characterize matrix clauses.

(24) The structure of finite embedded clauses in Tuparí

```
NP
   CP
     XP
       C
       C'
       TP
         T
         EvidP
           T
           AUXHABITUAL
           Evid
           AUXGO / AUXPOSITIONAL
           AUXHABITUAL
           RsltP
             AUXGO / AUXPOSITIONAL
             Rslt
             VP
               NP
                 V'
                   NP
                     V
```
4.1 Second position effects

Scholarship on Verb Second and related phenomena has established a number of ways that matrix and embedded clauses can differ with regards to second position effects (see Holmberg 2015 for theoretical overview and Wolfe & Woods 2020 for a range of case studies). Multiple Tupían languages exhibit second position phenomena; the most famous may be Karitiâna, which tends to be V2 in matrix clauses but is verb-final in embedded environments (Storto 2011; 2014; Rocha da Silva 2016; Storto et al. 2018).

Unlike Karitiâna or Germanic, there are NO second position asymmetries in Tupari: if a particle located in second position in matrix clauses is available within finite embedded clauses, it will be located in second position there too. These placement facts remain constant even when word order permutations take place. The two examples in (25), drawn from the same text, demonstrate.

In (a) the initial constituent within the embedded clause is the evidential-marked VP *tepuop'otnã*; in (b), it is the postpositional phrase *tarupa aropnã* ‘for the non-indigenes’ and the VP occurs in the post-second position field. In both examples distant past =*õpot* sits in second position.

(25) **Distant past particle =*õpot* remains in second position inside of embedded clauses**

| a. | *tepuop'otnã* =*õpot* =e *hẽ nêkaremankia* |
|---|---|
|   | [[ *te-puop'ot-nẽ-a* ] =*õpot* =e ] *hẽ nêkaremankia* |
|   | [[ *3c-learn-EV.SG-TH* ] =*DISTANT.PST* =3 ] *hẽ perfectly.resembling the way that he learned (NON-WITNESSED)* |
| b. | *tarupa aropnã* =*õpot* =irik'ẽnã *tepuop'otnẽ* *hẽ* |
|   | [[ *tarupa aropnã* ] =*õpot* =irik'ẽnẽ-a *te-puop'ot-nẽ* ] *hẽ* |
|   | [[ *3c-learn.EV.SG* ] =*DISTANT.PST* *work-TH 3c-learn.EV.SG* ] *hẽ nêkaremankia* |
|   | *perfectly.resembling* |
|   | *perfectly resembling the way that he learned to work for non-indigenous people (NON-WITNESSED)* |
|   | text narrated by Pedro Kup'eoyt Tupari |

§2 noted that the placement of the nominative enclitics provides evidence for a null tense particle that sits in second position. The same enclitic placement facts that we saw in that section for matrix clauses apply within embedded clauses, as well. Compare the following two embedded clauses against the superficially tenseless matrix clause in (6b). Given the placement of the nominative enclitic =*‘en ‘2sg’* in (26) (repeated from 19c), there must be a null tense particle in second position here:

(26) **Evidence for null tense particles in second position in finite embedded clauses**

| Vania yam =∅ =*‘en* =y-om-nẽ-am ] *hẽ* |
|---|---|
| [ Vania to =IMMEDIATE.PAST =2SG 3-give.EV.SG-ADV.FOC ] *hẽ* |
| ‘the thing that you-SG just gave to Vania (NON-WITNESSED)* |
| conversation: 2016-01-01 |

To reiterate, Tupari possesses no second position asymmetries. Hence the complement of *hẽ* can contain one or more head-initial syntactic projections at the highest level, as in (24).

4.2 Finiteness categories

The Tupari witnessed/non-witnessed evidential contrast is not optional; rather, it must be marked in all past tense declarative matrix clauses and in a subset of non-declaratives, as well. The same contrast obtains inside of past tense embedded clauses, too. This is shown by (27), where the internal head is the pronominal proclitic *s-.* The embedded clause in (a) contains no evidential suffix and is thus interpreted as witnessed: that is, the speaker saw the addressee bring the object in question. The embedded clause in (b), on the other hand, is marked as [–WITNESSED] and is interpreted as such.

---

10 Embedded clauses in Karitiâna lack the tense and mood morphology that characterizes positive polarity matrix clauses; that is, the language has no finite embedding of the sort described here for Tupari.
(27) The non-witnessed evidential contrast is maintained in finite embedded clauses

a. Sítèsa òpot 'en hè nãkop.
   [ s-ìtès-a = òpot = 'en ] hè = nãkop = ∅
   [ 3-bring.SG-TH = DISTANT.PST = 2SG ] HÈ = MAYBE = 3
   'It might be the thing that you-SG brought (WITNESSED).'
   conversation: 2016-11-19

b. Sítèynã òpot 'en hè nãkop.
   [ s-ìtèy-nè-a = òpot = 'en ] hè = nãkop = ∅
   [ 3-bring.SG-EV.SG-TH = DISTANT.PST = 2SG ] HÈ = MAYBE = 3
   'It might be the thing that you-SG brought (NON-WITNESSED).'
   elicitation: 2018-07-29

Speakers provide clear, consistent judgments on the felicity of examples like these. The maintenance of the obligatory witnessed/non-witnessed contrast inside of finite embedded clauses demonstrates that these clauses must contain an Evidential Phrase, just as matrix clauses do.

The full range of tense and aspect contrasts is maintained within embedded clauses, as well. §3 provides various examples of embedded clauses that contain the ancient past particle =kut, the distant past particle =òpot, or the near past suffix -t/n. embedded clauses also maintain the periphrastic tenses shown in (1b) and (1d), with a predicate-final suffix or second position particle combining with the AUXxo auxiliary series. Example (28) shows this periphrasis in both the matrix and embedded clause. (The matrix clause’s initial constituent is the VP, with the clausal nominalization serving as the direct object of top ‘see; this VP is then followed by the second position clause-typing particle =nè ‘YES/NO’.)

(28) Periphrastic tense construction in both the matrix clause and the embedded clause

Amẽkot sa otero’at ‘on hèt
[ VP [ amẽko-t si-a o-tero’e-a-t = ‘on ] hè-t
[ VP [ jaguar-NUC shoot-TH 1SG-AUXxo.SG-TH-NEAR.PST = 1SG ] HÈ-NUC
topto’omk’ nè etero’at ‘en?
top-to-’om-ka-a ] = nè e-tero’e-a-t = ‘en
see-NMZ-NEG-VBZ-TH ] = YES/NO 2SG-AUXxo.SG-TH-NEAR.PST = 2SG
'Did you-SG not see the jaguar that I shot?'
conversation: 2018-07-26

Embedded clauses also maintain matrix clauses’ three-way contrast between the polite future (a second position particle with modal overtones, often used for requests or commands), the near future (which combines an auxiliary with the suffix -pwa/mwa/p’a/m’a), and the distant future (also an auxiliary, used for events that will not happen any earlier than tomorrow). We already saw an instance of the distant future inside an embedded clause (example 23b, in §3.3). (29) shows the polite and near futures:

(29) Maintenance of future tense contrast in finite embedded clauses

a. Puop’om nè ‘en katkaere ke lap
   puop’om = nè = ‘en [ katkaere = ke ip-ap ]
   know-NEG = YES/NO = 2SG [ when = POLITE.FUT come.SG-ADV.FOC ]
   hèrè.
   HÈ-OBL
   'Do you-SG not know when he will come here?'
   conversation: 2016-01-20

b. Kiaraere nā otet’e oneporet èsapwa
   kiaraere nā o-tet’e oneporet [ e-sa-pwa
   happiness-OBL PROG 1SG-AUXxo.SG 1SG.also [ 2SG-come.SG-TH-FUT
   ‘eronā hèrè.
   ‘e-ronā ] hè-re
   AUX.SG-aGain ] HÈ-OBL
   'I too am happy that you-SG are going to come here again.'
   WhatsApp: 2016-10-05
In short, all tense and aspect contrasts are retained within matrix clauses. This fact provides strong evidence that embedded clauses contain the same Tense projection that matrix environments do.

We now turn to clause type. Most of the clause-typing particles are unavailable in embedded clauses; this is unsurprising, given their key role in marking speech acts (Sadock & Zwicky 1985; König & Siemund 2007). But = nākop ‘MAYBE’, which serves to convert propositions into statements of doubt, can be used within embedded clauses without issue. (30) shows = nākop inside an embedded clause which itself serves as the matrix clause’s initial XP, immediately prior to the second position tense particle = ke ‘POLITE.FUT’.

(30) The clause-typing particle = nākop ‘MAYBE’ can occur inside finite embedded clauses
Poatpoatkut’at nākop teyā hē ke ‘en
[ poatpoatkut’a-t = nākop te-yē-a ] hē = ke = ‘en
[ good.looking-NUC = MAYBE 3c-exist,sg-th ] hē = POLITE.FUT = 2sg
ey’etēy!
c-y-etē
ey 2SG-OBJ.FOC-bring.sg
‘Please bring the good-looking one that there may be / whatever good-looking one there is!’
WhatsApp: 2018-07-30

In addition, some speakers allow for the embedding of = mãkẽrõ ‘DON’T.KNOW’, which converts wh-questions into statements of ignorance:\footnote{Embedding = mãkẽrõ ‘DON’T.KNOW’ is not accepted by all of my consultants and may in fact be restricted to the speech of younger Tuparí. At least one consultant – an excellent middle-aged speaker – approved (30), which contains an embedded token of = nākop ‘MAYBE’, but rejected utterances such as (31). It is possible that this rejection is because embedded = mãkẽrõ ‘DON’T.KNOW’ is redundant when the matrix predicate is itself puop’om ‘not know’.}

(31) The particle = mãkẽrõ ‘DON’T.KNOW’ is acceptable in finite embedded clauses for some speakers
Puop’om ’on katkaere mãkẽrõ ko ’on aodeiam
puo-p’om = ’on [ katkaere = mãkẽrõ = ko = ’on aodeia-m
know-NEG = 1SG [ when = DON’T.KNOW = POLITE.FUT = 1SG village-INS
oterap here.
o-tet-ap ] hē-re
1SG-go,SG-ADV,FOC ] hē-OBL
‘I don’t know when I may go back to the village.’
WhatsApp: 2017-07-29

It is likely that the other clause-typing particles are unavailable in embedded clauses because they are used to mark speech acts such as polar or tag questions. So their resistance to embedding would be a pragmatic rather than narrowly syntactic fact. It is of course possible that in certain contexts clause-typing particles other than = nākop ‘MAYBE’ and = mãkẽrõ ‘DON’T.KNOW’ could be embedded. But the existence of such contexts would only strengthen the point made in this section, namely, that finite embedded clauses built with hē contain the same C projection known from matrix clauses.

4.3 Summary

Embedded clauses in Tuparí retain the second position effects that characterize matrix clauses and may contain all of the matrix clauses’ functional projections: EvidP, TP, CP. Only a subset of clause-typing particles may occur in finite embedded clauses, but this is likely not a syntactic fact; rather, it is a pragmatic consequence of these particles’ use as markers of root-only speech acts. The next section proposes a diachronic origin for Tuparí finite embedded clauses that accounts for their many structural resemblances to matrix clauses.

5 The diachronic origin of finite embedding in Tuparí

As finite embedded clauses are unique to Tuparí within the Tuparí an branch of the Tupían family, we are forced to ask where they came from diachronically. Based on structural ambiguities that
persist through the present, I will argue that a third person pronoun was deaccented and came to be reanalyzed as a clausal nominalizer. It was this process of deaccentuation that made possible the reinterpretation of erstwhile parataxis as true subordination.

5.1 The strong third person pronoun hè versus the clausal nominalizer hè

Tuparí pronouns come in two types: weak nominative enclitics, which are unstressed and parasitic on certain pieces of tense morphology, and strong pronominal roots, which can bear the full range of nominal morphology and serve specific informational roles (Singerman 2020). As shown by Tables 2 & 3, these two sets of pronouns are partially homophonous for speech act participants but differ markedly in the third person. The third person nominative enclitic can surface as = e but is null following a subset of clause-typing particles as well as /e/-final words (see 27). The third person strong pronoun, meanwhile, is hè, homophonous with the language’s clausal nominalizer. (Other demonstratives may also be used for anaphoric reference.) But unlike the clausal nominalizer hè – which is unstressed and is not usually preceded by any prosodic break – pronominal hè can and does take stress. I gloss it as PRON and translate it as ‘that one’, ‘those ones’, ‘that thing’ or ‘those things’, depending on number/animacy.

| SINGULAR | DUAL | PLURAL |
|----------|------|--------|
| 1INCL | ‘on’ | ‘okit’ | ‘okitwat’ |
| 1EXCL | ‘ote’ |        |        |
| 2      | ‘en’  | wat    |        |
| 3      | e (but ∅ in certain contexts) |        |        |

We have already looked at several declarative utterances where the NP subject (marked with the nuclear case) serves as the clause-initial XP. Such utterances are pragmatically neutral, often occurring in out-of-the-blue contexts; they are never interpreted with focus on the subject. To achieve a subject focus reading, one must switch from the neutral construction in (32a) to the marked one in (32b):

(32) A neutral declarative clause versus a clause with the subject focused
   a. Òwet Tupari.
      o-op-et Tupari
      1sg-father-NUC Tuparí
      ‘My father is Tuparí.’
      everyday speech
   b. Òpbe Tuparit.
      o-op = e Tupari-t
      1sg-father = 3 Tuparí-NUC
      ‘It is my father who is Tuparí.’
      conversation: 2016-11-21

In (a) the nuclear case appears on the clause-initial NP subject (òwet ‘my father’) and the nominal predicate (Tupari) is morphologically bare. But in (b) the clause-initial NP òp ‘my father’ is bare; the third person nominative enclitic = e occurs in second position; and the nuclear case surfaces on Tuparit. The discourse context for the latter example is revealing: the speaker was contrasting the ethnic affiliation of her mother, who was Aruá, with that of her
father, who was Tuparí. Subject focus constructions like (32b) always rely on contrast in the discourse, as further illustrated by the three utterances in (33).

(33) Three examples of the subject focus construction, with the third person enclitic & nuclear case highlighted

a. CONTEXT: I ask whether a pet parrot is male or female. My friend replies that it could be either; it is hard for humans to tell.
   
   Aoro emanë puowet.
   aoro eman =e puop-et
   parrot only =3 know-NUC
   'It’s only the parrot that knows for sure.'
   conversation: 2016-01-10

b. CONTEXT: I show a painful red bump on my arm to a friend. He recognizes the bump as a bee sting and concludes that a bee stung me without my having noticed at the time.
   
   Kapbe nā ètom’ensipnan.
   kap =e nā e-etom’en-si-pnē-a-n
   'It was the parrot that knows for sure.'
   conversation: 2017-08-29

c. CONTEXT: A speaker explains that she cannot say whether a particular animal tastes good because she has never tried it. Her brother was the only one who had eaten it, years back.
   
   Õa õpore ikat.
   o-oa =õpot =e i-ko-a-t
   1SG-brother =DISTANT.PST =3 3-eat-TH-NUC
   'It was my brother who ate it (WITNESSED).'
   conversation: 2017-08-14

In my corpus one of the most frequent usages of the third person pronoun hē is as a clause-initial focus. (34) provides two examples akin to the subject focus constructions in (32b) and (33).

(34) Pronominal hē can serve as a clause-initial focused argument

a. CONTEXT: I ask friends if they like the Ouroeste coffee brand; they say yes, enthusiastically.
   
   Hē nā otekafe kot’oaet.
   hē =∅ =nā ote-kafe kot’oap-et
   PRON = 3 =FOCUS 1PL.EXCL-coffee favorite-NUC
   'That thing [= the Ouroeste brand] is our-EXCL favorite coffee.'
   conversation: 2016-12-15

b. CONTEXT: Some friends are discussing the late husband of a respected matriarch. They comment that he had been Tuparí through and through.
   
   Hē õpore nā Tupari ta’atenemnam.
   hē =õpot =e =nā Tupari ta’ate-nē-mnē-a-n
   PRON =DISTANT.PST =3 =FOCUS Tuparí true-VBZ-EV.SG-TH-NUC
   'That one [= the matriarch’s late husband] was a true Tuparí (NON-WITNESSED).'
   conversation: 2017-08-10

It is also possible for the pronoun hē to be a possessor or the complement of a postposition:

(35) Pronominal hē can serve as the focused complement of a postposition

CONTEXT: A friend from the village of Nazaré is studying in a nearby village but says he will return home the next day. Since one needs a boat to reach Nazaré, I ask how he plans to get there. He says he is waiting for his parents and grandparents to visit him; he will then hitch a ride home with them.

a. Hē yope nā otero peo’ap eret.
   hē yope =nā o-tet-ro peo’ap eret
   PRON along.with =FOCUS 1SG-go.SG-NMZ DISTANT.FUT + 1SG tomorrow
   'I'll be going tomorrow along with those ones [= the speakers’ parents and grandparents].'
   WhatsApp: 2020-06-30
Pronominal hè can also serve as a predicate. In (36), hè ‘PRON’ is the predicate inside of the thought report that serves as the direct object of ke ‘say, think’; the subject of the thought report is ‘en ‘2sg’.

(36)  **Pronominal hè can serve as a nominal predicate**  
**CONTEXT:** A friend is surprised to learn that I am from the United States; she thought that I was from Germany.

a.  
Hè nākop ‘en kàpbi’a ‘on wan’om.  
[ hè =nākop =‘en ] ke-a-pbi’a = ‘on wan’om  
‘However I was saying/thinking that you might be that thing [ =from Germany].’

conversation: 2016-12-11

In all of the above examples, the pronoun hè serves as the clause-initial focused constituent or as a subpart of the focused constituent. But this pronoun can also serve in the opposite informational role, that is, as the backgrounded material in an argument focus construction. This backgrounded material must always bear the nuclear case, as we saw in (32b), (33), (34a) and (34b):12

(37)  **Pronominal hè can be backgrounded in subject focus constructions**  

a.  
**CONTEXT:** I ask a friend whether he is his mother’s eldest child. He says no.

Ner’om. Danieoe nā hèt.  
ner’om Danieo =e =nā hè-t  
‘No, it is Daniel who is that thing [=the eldest child of the speaker’s mother].’

conversation: 2017-08-12

b.  
**CONTEXT:** Over dinner I ask a friend whether she is an Evangelical Christian. She says no, at which point the hostess chimes in.

ôrë hèt.  
on =e hè-t  
1SG =3 PRON-NUC  
‘It is I who am that thing [=an Evangelical].’

conversation: 2016-11-20

The above examples demonstrate the interpretive flexibility of pronominal hè. Its referent may be a discourse-salient human, as in (34b) and (35). The referent does not, however, need to be animate: in (34a) hè refers back to a particular brand of coffee. Pronominal hè can also be interpreted as referring to an abstract quality or state, as in (36) (where hè is interpreted as ‘from Germany’) and (37b) (where it is interpreted as ‘Evangelical Christian’). Note that the interpretation of hè with regards to number is flexible, too: in (35) there is no overt marking on hè but it is still interpreted as plural, referring back to the speaker’s parents and grandparents.13

We have now seen that the strong third person pronoun hè and the clausal nominalizer hè have very different distributions despite their segmental homophony. The pronoun can serve both as a focus and as backgrounded material against which some other constituent is highlighted. It bears sentential stress when focused. The clausal nominalizer, however, always comes right after an entire finite clause, does not bear stress, and tends not to be preceded by a pause.

---

12 The backgrounded material in object focus constructions sometimes lacks the nuclear case for reasons that appear to be morphological in nature, as in (1) (and in 38c and 39b).

13 See also the discussion surrounding the plural-like suffix -e at, at the end of §3.2.
5.2 From parataxis to subordination

In keeping with the grammaticalization pathways for embedded clauses discussed by Givón (2012; 2015; 2016), among others, I propose that the elimination of the pause prior to pronominal hè led to a reanalysis of the pronoun as a nominalizer, one enclitic onto the immediately preceding clause. This grammaticalization pathway accounts for the segmental homophony between the pronoun hè and the clausal nominalizer hè, and it helps us to make sense of those cases where hè can be parsed either way.

To begin, consider example (38). This is an object focus construction like (30) (in §4.2) and (1) (in footnote 12); the matrix verb bears the object focus prefix y-. The focused nominal in (38c) is the embedded clause kurem ’ote nẽ hè ‘the thing that we-EXCL made just now’. This embedded clause functions as an internally headed relative, with the null third person object of the verb nẽ ‘do, make’ serving as the internal head. The lack of overt tense morphology in the embedded clause ensures a just-before-UT interpretation.

(38)

a. Finite declarative clause
Kurem ’ote nẽ.
kurem = ’ote ∅-nẽ
just.now = 1PL.EXCL 3-make
’We-EXCL made it just now.’

b. Clausal nominalization of (a)
kurem ’ote nẽ hè
[ kurem = ’ote ∅-nẽ ] hè
[ just.now = 1PL.EXCL 3-make ] hè
‘the thing that we-EXCL made just now’

c. Clausal nominalization from (b) serves as the focused direct object in the matrix clause
Kurem ’ote nẽ hè nã oymk.
[ kurem = ’ote ∅-nẽ ] hè = ∅ = nã o-y-mak
[ just.now = 1PL.EXCL 3-make ] hè = 3 = FOCUS 1SG-OBJ.FOC-send
’What I sent you was the thing that we-EXCL made just now.’

The utterance in (c) was produced without any pause whatsoever prior to the clausal nominalizer hè. If there had been a pause there, the parse would have been very different: Kurem ’ote nẽ. || Hè nã oymk ‘We-EXCL made it just now. It was that thing [ = what we-EXCL made just now] that I sent to you’. It is the deaccentuation of hè and the elimination of the preceding prosodic break that ensures that (38c) is parsed hypotactically (‘what I sent you was the thing that we-EXCL made just now’) rather than paratactically (‘we-EXCL made it just now; it was that thing that I sent to you’).

Now consider the two-clause utterance in (39). This example differs from (38c) in that it was uttered with a prosodic break before hè.

(39) Two-clause utterance with prosodic break before the focused pronoun hè

CONTEXT: My friend tells me he has had lunch, so I ask Kat’are eyko? ‘What did you eat?’ This is how he replies.

a. Marlit pot’a opet màkñà herõwap.
Marli-t pot’a ope-t màk-nê-a herõwap
Marli-NUC boar thigh-NUC send-EV.SG-TH yesterday
’Marli sent boar’s thigh yesterday (NON-WITNESSED).’

b. || hè oyko.

hè = ∅ o-y-ko
PRON = 3 1SG-OBJ.FOC-eat
’It was that thing [ = the boar thigh that Marli sent yesterday] that I ate.’

The first line is a subject-initial, informationally neutral declarative of the sort we saw in (4b), (7) and (32a). The second line is an object focus construction where the clause-initial nominal
is the strong pronoun hè, just like (1); as expected, the verb bears the object focus prefix y-.
With a clear pause before hè, (39) must be parsed as paratactic: ‘Marli sent boar’s thigh yesterday (NON-WITNESSED).’ It was that thing that I ate.’ Were that pause eliminated – with accompanying deaccentuation of hè – the parse would be hypotactic: ‘What I ate was the boar’s thigh that Marli sent yesterday (NON-WITNESSED).’

Research into grammaticalization often seeks to identify areas where multiple parses or interpretations are available synchronically; the availability of multiple parses helps to identify diachronic changes that were recently completed or even ones that are still in progress (Narrog & Heine 2011, 2017). We have seen that multiple parses are indeed available for hè: speakers produce utterances like (39b), where hè is preceded by a pause and is interpreted as a pronoun, and ones like (38c), where hè is not preceded by a pause and serves as a clausal nominalizer. That non-elicited, everyday talk includes both uses of hè provides evidence for the diachronic pathway proposed here.

Just as importantly, there are contexts where multiple synchronic parses are unavailable – even adjusting for prosody. When a finite embedded clause is both preceded and followed by matrix clause material, then there is no way to insert a pause before hè to achieve parataxis; the only possible parse is one of subordination. Compare (40a) and (40b). In both utterances the clause-initial constituent is the adverbial here ‘so, and, then’; the tense particle = ke ‘POLITE.FUT’ sits in second position, followed by the nominative enclitic = ’en ‘2SG; and the verb itès∼etès ‘bring.sg’ bears the adverbial focus suffix -ap. The sole substantive distinction between the two utterances is the internal complexity of the direct object: the third person proclitic s- in (a), an entire embedded clause in (b).

(40) Two utterances that differ regarding the internal complexity of the direct object (in bold)

(a) Here ke ’en sitèsap ham
here =ke =’en s-itès-ap ham
so =POLITE.FUT =2SG 3-bring.sg-ADV.FOC here
oteptoaptenã.
ote-potop-ap-tenã
1PL.EXCL-see-NMZ-PURPOSIVE
‘So please bring him here, for us-EXCL to see.’
WhatsApp: 2017-06-22

(b) Here ke ’en pekat ’on hèt
here =ke =’en [ ₕ-pek-a-t =’on ] hè-t
then =POLITE.FUT =2SG [ 3-ask.for-TH-NEAR.PST =1SG ] HÈ-NUC
etèsap.
etès-ap
bring.sg-ADV.FOC
‘Then please bring the thing that I asked for [some days/weeks back].’
conversation: 2015-12-28

Example (40b) does not enjoy the structural ambiguity that we saw with (38c) and (39). That is, it would not be possible to insert a pause before hèt and in so doing to create a well-formed paratactic structure. This is because the finite embedded clause in (40b) is linearly positioned in the middle of the matrix clause: it is sandwiched between the clause-initial XP here ‘then’ (followed by the tense particle = ke and the nominative enclitic = ’en) and the verb etèsap (which bears the adverbial focus suffix -ap because the clause-initial XP is the adverbial here ‘so, and, then’). This is an example of unambiguous syntactic embedding, with one fully finite clause nested inside another.

Utterances like (40b), with a finite embedded clause embedded in the middle of a finite matrix clause, are not unusual. For a particularly impressive example take (41), a subject focus construction just like the ones examined in §5.1. As in (40b), the finite embedded clause in this example is pekat ‘on hèt ‘the thing that I asked for’; it serves as the direct object of the verb etès ‘bring.sg’. But the entire object-plus-transitive-verb complex is backgrounded here, as shown by its taking the nuclear case. The clause-initial strong pronoun en ‘2SG’ is focused.
Finite embedded clause inside of a subject focus construction (with the embedded clause in bold)

CONTEXT: A friend asks me to bring him a needed item from the city, but I forget to do so. When I tell him I didn’t bring what he asked for, he jokes that he will fight me. Playing along, I ask him why. This is how he replies.

a. Érê nā pekat 'on hèt
   en =e =nā [∅-peka-t] =‘on ] hè-t
2sg =3 =FOCUS [3-ask.for-TH-NEAR.PST =1SG ] HÈ-NUC
etèy-to’omkapnaren.
etèy-to’om-ka-pnè-a-n-en
bring.SG-NMZ-NEG-VBZ-EV.SG-TH-NEAR.PST-NUC
‘(Because) it was you who did not bring (NON-WITNESSED) the thing that I asked for.’

WhatsApp: 2017-07-30

The near past suffix occurs twice in this utterance: once on the embedded verb pekat and once on the matrix verb etèy-to’omkapnaren. In addition, both the embedded clause and the matrix clause are specified for evidentiality: the embedded clause bears no overt evidential suffix because the speaker (trivially) witnessed his own act of requesting an item from town; but the matrix clause is marked as non-witnessed since the speaker did not see my failure to bring the requested item. (In Singerman 2019: appendix we discuss how the evidential suffix interacts with negation and first person subjects.) In sum, utterances such as (40b) and (41) (see also 28, in §4.2) demonstrate that the Tuparí language allows for the grammatical categories of evidentiality and tense to be embedded recursively.

The grammaticalization pathway proposed here predicts the clausal complements of the nominalizer hè to exhibit no radical differences from normal matrix clauses. That is, if the nominalizer hè developed out of the destressing of the strong third person pronoun following independent finite clauses, then the clauses nominalized by hè should retain features characteristic of matrix finite clauses in general. As we already saw in §4, this prediction is born out: all tense, aspect and evidential contrasts are retained within embedded clauses, and a subset of the clause-typing particles are available within them as well. What is more, word order flexibility of the sort known from matrix clauses occurs in embedded ones. Although Tuparí VPs and AuxPs are head-final, the language is not strictly verb-final on the surface; post-predicate and right-peripheral constituents are common. (See 40a, above, where the verb is followed by ham ‘here’ and otepotoaptenã ‘in order for us-excl to see’.) Embedded clauses, too, allow for a wide variety of syntactic constituents to occur at their right edge. Hence the nominalizer hè can linearly follow an oblique-marked nominal (example 19a); a lexical verb inflected for tense and/or evidentiality (9c, 17c, 12b, 20b, 21b); a nominative enclitic that is positionally parasitic on tense morphology (13b, 18b, 19b, 22a, 40b); a verb that bears the adverbial focus suffix -ap (17b, 19c, 29a); and so on. That finite embedded clauses built with hè exhibit the same constituent ordering patterns found in matrix clauses follows straightforwardly if they descend from independent clauses rather than, say, from non-finite constructions that managed to acquire finiteness properties over time. (The latter process, referred to by Givón 2016 as REFINITIZATION, does seem to have occurred elsewhere in the Tupian family, however. See Rose 2013, 2016 on evidence for this process in the Tupi-Guaraní branch.)

5.3 Functional advantages of the innovative finite embedded clauses

I hypothesize that the innovative finite embedded clauses have encroached on the functional territory of the older, non-finite nominalizations thanks to their ability to exploit the language’s elaborate systems of tense, aspect and evidentiality. As the older nominalizations are incompatible with the finiteness morphology that characterizes matrix clauses, they are useful only when the TAME question is irrelevant or not at-issue.

CONTEXT: A Tuparí friend notices a colorful lanyard attached to his friend’s backpack.
Eynè nè?
e-y-nè =nè =∅
2sg-NMZ object make =YES/NO =3
‘Was this thing made by you?’ / ‘Is this thing your making?’

conversation: 2016-11-10
The speaker of (42) was curious as to who had made the lanyard on his addressee's backpack; when the event of lanyard-making took place was not of concern. All of the non-finite nominalizations shown in §3.1 receive similarly generic or unspecified interpretations as far as tense, aspect and evidentiality are concerned. But the nominalizations built using hè are different: they permit speakers to employ the TAME contrasts which they are already accustomed to using in matrix contexts. This makes possible the expression of nuanced distinctions within individual utterances, as in (43) (repeated from §3.2).

(43)  Tense contrasts between matrix and embedded clauses

| Omoto    | pêka | oteroa’t | 'on hè tere nà |
|----------|------|----------|---------------|
| o-moto   | pêk-a | o-tero’e-a-t | 'on hè tere =nà |
| [ 1SG-motorcycle buy-TH 1SG-AUXSG-TH-NEAR.PST =1SG ] HÈ on =FOCUS |

òsa  o’e

1SG-come.SG-TH 1SG-AUX.SG

'It was on the motorcycle of mine that I bought some time ago that I came here.'

WhatsApp: 2018-02-04

The matrix clause in this utterance employs a near past construction, built using the auxiliary 'e ‘AUX.SG', that indicates that the speaker's act of coming back to the village took place on the same day as – but several hours prior to – the Utterance Time. The embedded clause, meanwhile, uses the same auxiliary-plus-suffix periphrasis seen in (1b) and (28): this combination of near past -t with tere ‘AUXSG ensures that the speaker's act of motorcycle-purchasing is interpreted as having taken place at least several months prior to, but no more than a year or two before, UT. Hence using a finite embedded clause allows the speaker of (43) to articulate the different amounts of time that have elapsed since buying the motorcycle, on the one hand, and returning to the village, on the other.

Nuanced evidential and aspectual distinctions are possible as well. The entire finite embedded clause in (44) serves as the possessor of (h)et 'name'. The nominalization functions as an internally headed relative; its internal head is the NP object aoro non 'other parrot'.

(44)  Tense, aspect, and evidentiality contrasts between the matrix and embedded clauses

| Aoro     | vendekapnarè | hè | heret |
|----------|---------------|----|------|
| [ aoro nò-n vendeka-pnè-a-n ] =e ] HÈ het-et |
| [ parrot other-NUC sell-EV.SG-TH-NEAR.PST =3 ] HÈ name-NUC |

Cassionambi’a.

Cassio-nè-a-mbi’a

Cassio-VBZ-TH-DURATIVE

'The name of the other parrot that she sold (NON-WITNESSED) was Cassio (WITNESSED).'

conversation: 2018-08-30

The embedded verb, vendeka 'sell', bears the singular evidential -pnè and near past -n. The matrix predicate, meanwhile, bears durative -pbi’a/mbi’a, a unique member of the language's inventory of inflectional morphology in that its meaning conflates tense, aspect and evidentiality. While almost all of the other past tenses are morphosyntactically compatible with the evidential suffix -pnè/psira – and therefore semantically compatible with non-witnessed interpretations – the durative is not. Hence the matrix clause in (44) must be interpreted such that the speaker had met the parrot in question and heard it addressed as Cassio. In terms of aspect, durative -pbi’a/mbi’a is used only (a) for actions that were repeated over and over again and (b) for permanent or at least very long-lasting states. (See 14c and 15c.) So the matrix and embedded clauses in example (44) encode different values for all three clausal categories: tense, aspect, and evidentiality. Being named Cassio was a permanent state of the pet parrot, to which the speaker personally bore witness; whereas the selling of Cassio to a new owner was a one-time event that the speaker did not see take place.

While all of these tense, aspect and evidential contrasts are easily expressed through the innovative clausal nominalizations built with hè, they are neutralized in the non-finite subordination strategies inherited from Proto-Tupían. Speakers are aware of this
distinction. Consider example (45), in which the actor nominalizer suffix first discussed in §3.1 appears:

(45) context: A friend informs me that his family’s pet macaw is alive and well.
Kaykay’at tero’aem, ewekaret.
kaykay’a-t tero’e-a-em e-wek-at-et
macaw-NUC exist.SG-TH-still 2SG-bite-NMZ_ACTOR-NUC
‘The macaw is still there, the one that bit you / the you-biter.’
WhatsApp: 2020-09-26

The speaker of this utterance confirmed that one could replace the non-finite nominalization with a full finite embedded clause, as in (46):

(46) Kaykay’at tero’aem, eweka òpore hét.
kaykay’a-t tero’e-a-em [ e-wek-a = òpot = e ] hè-t
macaw-NUC exist.SG-TH-still [ 2SG-bite-TH = DISTANT.PST = 3 ] HÈ-NUC
‘The macaw is still there, the one that bit you (WITNESSED).’

As it is non-finite, the highlighted actor nominalization in (45) lacks any tense or evidential specification; but the embedded clause in (46) successfully conveys the temporal remoteness of the biting event and the fact that the speaker saw it take place. So it is perhaps surprising that the speaker described the latter, fully finite option with the Portuguese mais certo ‘more correct.’

With regards to the language’s evidential contrast, I should note that certain clause-typing particles invert the deictic orientation of the evidential suffix -pnẽ/psira from speaker to addressee, a phenomenon known as INTERROGATIVE FLIP (Bhadra 2020). Other clause-typing particles, meanwhile, neutralize the witnessed/non-witnessed contrast altogether. Singerman (2019: 411–428) discusses the interaction between evidentiality and clause type and shows that the evidential contrast inside of finite embedded clauses projects over the matrix clause-typing particle in the manner of a presupposition. That is, an embedded clause will maintain a speaker-oriented evidential contrast even when the matrix clause’s evidential contrast has flipped from speaker to addressee or has been neutralized. This finding buttresses the central claim of this subsection: finite embedded clauses allow speakers of Tuparí to employ the language’s full range of TAME distinctions in subordinate contexts. This functional advantage has likely contributed to the finite embedded clauses’ growing role in everyday discourse – and to their appropriation of functions that were previously carried out by non-finite constructions alone.

5.4 Discussion

This section has argued that Tuparí developed finite embedding through a grammaticalization process that is well-attested for other languages: the elimination of a prosodic break between two separate clauses allowed for paratactic structures to be reinterpreted as involving true subordination. In particular, the third person pronoun hè came to be reinterpreted as an enclitic clausal nominalizer. Abstracting away from prosodic cues, two different interpretations are sometimes available; see for instance (38c) and (39). But in other cases the only possible interpretation is that of subordination: see (40b) and (41), where a prosodic break could not be inserted prior to hè to achieve a paratactic parse. Utterances such as (40b) and (41) demonstrate the extent to which finite embedding has been grammaticalized in Tuparí.

It appears that a similar pathway has allowed for other Tupían languages to develop finite subordination strategies, as well. Now, no such strategies are described for the remaining members of the Tupían family’s Tuparí an branch; these employ exclusively non-finite constructions of the sort illustrated in §3.1 (see Galucio 2011a, b, 2014 on Sakurabiá). One must look to more distant relatives to find clausal nominalizations similar to the ones described in this paper. Moore (1989, 2012) shows that Gavião, of the family’s Mondé branch, builds nominalized embedded clauses using mène ‘abstract nominalization’ and mât ‘concrete nominalization’. Like hè, these nominalizers follow rather than proceed the clauses which they embed. An additional parallel with Tuparí is that the nominalized embedded clauses in Gavião retain multiple tense/aspect distinctions from matrix clauses. Moore further observes that mène and mât are homophonous with independently attested pronouns that are used as demonstratives, similar
to how English *that* can serve both as a complementizer and as a demonstrative. This pronoun-nominalizer homophony indicates that clausal nominalizations in Gavião likely arose through the same general pathway proposed here for Tuparí. I must stress, however, that there is no evidence to suggest that contact between speakers of Tuparí and Gavião led both languages to develop finite embedded clauses. It is true that a sister dialect of Gavião, Aruá, was historically spoken in the Rio Branco region, and there have been cases of intermarriage between Aruá and Tuparí speakers. However, the Aruá have been small in number and have tended to maintain closer ties with the Makurap than with the Tuparí (Denny Moore, p.c.). So the most reasonable conclusion is that the existence of finite embedded clauses in both Gavião and Tuparí is due to the two languages' having undergone parallel, but independent, diachronic changes. The lack of any phonological similarity between Gavião *méne / mát* and Tuparí *hè* is a reflection of those changes' independence, as is the fact that the two languages' nominalizers enjoy different scopal properties: *méne / mát* can nominalize VPs in addition to whole clauses, while *hè* only ever nominalizes clauses. (An additional difference between the two languages' finite embedded clauses is that auxiliaries in Gavião embedded clauses take a special suffix, -néè, which never occurs on matrix auxiliaries. Nothing comparable occurs in Tuparí, whose embedded clauses preserve matrix finiteness morphology without alteration.)

If the pathway argued for in this section correctly explains the rise of finite embedded clauses in Tuparí, then we do not need to invoke language contact – with Tupían or non-Tupían languages – to explain this development from parataxis to subordination. This is an important point to stress given that the Amazon Basin constitutes a region of considerable multilingual interaction, with structural convergence argued to have taken place in many contact zones and between speakers of many different languages (Beier et al. 2002; Aikhenvald 2002; Stenzel 2005; Michael 2014; Epps & Michael 2017; Epps 2020). The languages of the Rio Branco region belong to a broader contact zone that straddles the Brazilian-Bolivian border, the Guaporé-Mamoré linguistic area (Crevels & van der Voort 2008). Indigenous multilingualism was the norm on the Rio Branco at least through the mid-twentieth century, with the Tuparían language Makurap serving as a pan-ethnic lingua franca (Caspar 1956; 1957; 1975). That multilingual system has since given way to asymmetrical Tuparí-Portuguese bilingualism, with the latter commanding a greater degree of prestige. Yet despite intensive bilingualism with Portuguese on the Rio Branco and despite the presence of a fair number of Portuguese borrowings in Tuparí, there is no evidence that Tuparí finite embedded clauses are the result of Portuguese influence. Portuguese embedded clauses are introduced by an initial complementizer or relative pronoun; their phrase structural organization is head-initial rather than head-final; when used as relative clauses, they are externally rather than internally headed; and so on. Since Tuparí embedded clauses bear no material or structural resemblance to the Portuguese ones, their diachronic emergence cannot be explained via contact. The parataxis-to-hypotaxis change described in this section is best analyzed as an *autochthonous development* within Tuparí grammar.

6 Tuparí finite embedded clauses and syntactic typology

Having analyzed the synchronic structure of Tuparí finite embedded clauses and having provided an account of their diachronic emergence, we now turn to a question of broader interest: where does this language fit into the typology of syntactic headedness? The analysis of finite embedded clauses presented in §4 posits a right-headed NP, headed by *hè*, whose complement contains the same functional projections that characterize finite matrix clauses, including a head-initial CP. This analysis is typologically surprising given recent scholarship on the crosslinguistic distribution of head-initial and head-final phrase structure. In particular, Holmberg (2000) proposed that a head-final projection cannot immediately dominate a head-initial one:

\[(47) \text{ If a phrase } \alpha \text{ is head-initial, then the phrase } \beta \text{ immediately dominating } \alpha \text{ is head-initial. If } \alpha \text{ is head-final, } \beta \text{ can be head-final or head-initial. (Holmberg 2000: 124)}\]

This ban on the configuration *[^\footnote] [[ \footnote X ZP ] X ] was known in earlier literature as the Final-over-Final Constraint and has more recently been rechristened as the Final-over-Final Condition (Sheehan et al. 2017, Roberts 2019: chapter two). As Tuparí embedded clauses consist of a head-final NP immediately dominating a head-initial CP (see the tree in 24), they would appear to violate FOFC.
This section explores that apparent violation in greater detail. I will argue that Tuparí in fact complies with more recent versions of FOFC, which restrict the Condition’s domain of application in important ways (§6.2). By situating the Tuparí facts within the broader typological landscape, we will arrive at a clearer understanding of the distribution of categorial features in the Tuparí clause (§6.3).

6.1 The Final-over-Final Condition (FOFC)

Greenberg (1963) put forth linguistic universals as implicational statements: ‘if a language has property \( X \), it { will / is highly likely to } also have property \( Y \).’ Subsequent research has revealed that such statements are vulnerable to areal and genealogical biases. For instance, while prenominal relative clauses imply object-verb VPs, the inverse implication holds in Eurasia only: on other continents, languages with head-final VPs possess postnominal relatives as often as, or more often than, prenominal ones (Dryer 1992). Greenbergian implicational universals have also been argued to reflect the outcome of probable diachronic changes without bearing on what is synchronically possible in human language (Aristar 1991; Newmeyer 2005; Harris 2008; Whitman 2008; Djamouri & Paul 2019, among others).

Formal syntacticians have sought to abstract away from historical, geographic and genealogical biases and to instead explain typological variation in constituent order through restrictions on the well-formedness of phrase structure. While earlier approaches attempted to account for crosslinguistic variation via a single head-directionality parameter, more recent theorizing has instead proposed to restrict how head-complement and complement-head phrase structure may interact. On the assumption that a head may either precede or follow its complement, we obtain the four configurations in (48):

\[ (48) \quad \text{The four possible phrase structural configurations} \]

\[ \text{a. Head-complement XP embeds head-complement YP:} \]

\[
\begin{array}{c}
\text{XP} \\
\text{X} \\
\text{Y} \\
\text{YP} \\
\text{ZP} \\
\end{array}
\]

\[ \text{b. Complement-head XP embeds complement-head YP:} \]

\[
\begin{array}{c}
\text{XP} \\
\text{YP} \\
\text{X} \\
\text{ZP} \\
\text{Y} \\
\end{array}
\]

\[ \text{c. Head-complement XP embeds complement-head YP:} \]

\[
\begin{array}{c}
\text{XP} \\
\text{X} \\
\text{YP} \\
\text{ZP} \\
\text{Y} \\
\end{array}
\]

\[ \text{d. Complement-head XP embeds head-complement YP:} \]

\[
\begin{array}{c}
\text{XP} \\
\text{YP} \\
\text{X} \\
\text{Y} \\
\text{ZP} \\
\end{array}
\]
The harmonic configurations (48a and 48b) are widely attested, as is the disharmonic configuration in (48c). Indeed, §2 showed that Tupari matrix clauses exhibit the structure in (c): the highest region of the clause is head-initial, but the lower region is head-final. This is also the kind of clause structure assumed in much work on the Germanic languages, which tend to be head-final in the VP but head-initial at the top of the clause. It is that high head-initiality, coupled with Head Movement of the finite verb, that yields Verb Second (Holmberg 2015). The odd tree out in (48) is (d), where head-final XP embeds head-initial YP. This configuration is crosslinguistically much rarer than one would expect a priori, an observation which led Holmberg (2000) to propose FOFC.

The Tupari facts are of relevance to this discussion because the language’s finite embedded clauses are right-headed nominalizations whose complement includes high head-initial projections, realized as second position particles (see the tree in 24).

(49) A finite embedded clause that itself contains a second position particle (particle highlighted)

\[ \text{okiot kut tenon wât’omnâ hê ma’ã} \]
\[ \text{[ okio-t = kut te-nô-n wât’om-nê-a ] hê ma’ã} \]
‘talk about the man who poisoned his friend (NON-WITNESSED)’

the title of a text written by Raul Pat’awre Tupari

In that they involve a head-initial CP and/or head-initial TP inside of a head-final NP, such finite embedded clauses violate the original (and strictest) version of FOFC. See the Appendix for a step-by-step derivation of (49).

6.2 Specifying the domain of application for FOFC

Exceptions to Holmberg’s original ban on \[ \text{[XP [YP Y ZP ] X]} \] have accumulated since it was first proposed, in 2000, so researchers have sought to find principled explanations for those exceptions without abandoning the underlying typology. A productive line of research has searched for the exact domains within which FOFC applies (the working assumption being that any surface counterexamples will cease to be problematic once we identify the right domain of application).

The proposal by Biberauer et al. (2014) takes FOFC to apply within Extended Projections in the sense of Grimshaw (2000, 2005). On this proposal, the configuration \[ \text{[XP [YP Y ZP ] X]} \] is banned only when the heads X and Y bear an identical categorial feature, either \[ +\text{verbal} \] or \[ +\text{nominal} \]. To bear an identical categorial feature X and Y must belong to the same, uninterrupted portion of a tree, consisting of a single lexical item and a shell of functional heads immediately above it.

This revision to FOFC enjoys much empirical support; for reasons of space I discuss only one example of that support here. The Germanic languages’ \[ +\text{v} \] functional projections – v, Aspect, Tense, C – consistently obey FOFC, yet FOFC does not apply when a transitive verb selects a nominal complement. So in a head-final (which is to say, object-verb) VP, the object can be internally head-initial.\(^{14}\)

(50) German examples from Biberauer et al. (2014: 197-198; my highlighting)

a. Johann hat \[ \text{[vp [wp einen Mann ] gesehen]} \].
   Johann has a man seen.
   ‘Johann has seen a man.’

b. ... dass Johann niemals \[ \text{[wp den Verdacht [cp dass er eigentlich ein}}
   that Johann never the suspicion that he actually an
   angenommene Kind sei ] besprochen hat.
   adopted child be.SUBJ discussed has
   ‘... that Johann has never discussed the suspicion that he is actually an
   adopted child.’

\(^{14}\) I have modified example (50c) from Biberauer et al. (2014: 198) so as to show the base position, within the preverbal DP, from which the CP is extraposed. In all other respects the glosses in (50) are unmodified.
In (50a) the DP *einen Mann* ‘a man’ precedes the transitive verb *geschen* ‘seen’; together they form a head-final VP whose complement is head-initial. It is also possible for the preverbal object to itself contain an embedded clause with a head-initial complementizer; this is shown by (50b), where *den Verdacht dass er eigentlich ein angenommenes Kind sei* ‘the suspicion that he is actually an adopted child’ precedes *besprochen* ‘discussed’. (Biberauer et al. note that extraposition of the CP, as in 50c, is preferred in the spoken language, but they report both 50b and 50c as grammatical.) If FOFC applied cross-categorially then the utterances in (50) should be impossible. But if FOFC applies within Extended Projections alone, then there is no violation here: the verb is the base of a [+v] Extended Projection while its object constitutes to a separate, [+N] one.

An alternative approach to restricting the domain of FOFC is provided by Erlewine (2017), who uses data from Mandarin Chinese to argue that the Condition applies within PHASES. Phases were originally proposed by Chomsky (2001) as a means to impose locality conditions on derivations in the Minimalist framework. In this sense they are conceptually similar to (though formally distinct from) Extended Projections, as well as various other locality-defining mechanisms proposed in different theories – for example, the elementary trees of Tree Adjoining Grammar (Frank 2002, 2006). Minimalist research has converged on the idea that three specific functional heads serve to demarcate phase edges: C, v, and D. Erlewine’s proposal is that FOFC applies within – but, crucially, not across – the Spell-Out domains that such heads define. Head-final phrase structure is thus predicted to grammatically embed head-initial phrase structure only at phase boundaries. Using this theory, Erlewine is able to account for the distribution of those morphemes which have been referred to in the literature on Chinese as SENTENCE-FINAL PARTICLES (SFPs) and which appear, on the surface, to violate FOFC. His proposal for the syntactic structure of Mandarin posits two different layers of head-finality in what is otherwise a head-initial clause; these head-final layers correspond to phasal boundaries (Figure 3).

This paper does not apply a phase-based theory to the Tuparí data for two reasons. First, I lack independently-motivated diagnostics to determine which functional heads, if any, serve to demarcate phasal boundaries in the Tuparí clause. This is a thorny problem given the possibility that languages may differ not only in the structural phenomena that diagnose phasehood but also in the functional heads that define phases. (For instance, Erlewine 2017 proposes that it is not the vP but rather a slightly higher functional projection, SFP, that is phasal in Mandarin.) Second, Erlewine’s empirical claims concerning the syntax of SFPs in Mandarin have been challenged (Pan 2018, 2020); I am not, however, in a position to adjudicate between the different accounts of the Mandarin facts. Given the lack of known phasehood diagnostics for Tuparí, the next subsection examines the Tuparí facts only in light of the proposal to restrict
FOFC to Extended Projections. This examination will end up uncovering some interesting facts concerning the distribution of ategorical features in the Tuparí clause.\footnote{A starkly different approach has been pursued in the functionalist literature. Hawkins (2013, 2014) takes the surface violations of FOFC to constitute refutations of the Condition’s typological validity. On his interpretation FOFC is not an exceptionless universal that derives from principles of Universal Grammar, but instead results epiphenomenally from pressures of processing and efficiency. For lack of space I must defer discussion of Hawkins’s approach to future work; but see Sheehan (2013) for some critiques.}

6.3 Reconsidering the Tuparí facts in light of FOFC

§3 provided extensive evidence that Tuparí finite embedded clauses behave in terms of their external syntax like any run-of-the-mill NP: they can serve as subjects, as objects, as the complement of postpositions, as possessors, as predicates, etc. Morphologically, too, finite embedded clauses behave like non-derived nominals: the nominalizer hè can take the full range of case suffixes and can also bear nominal-only suffixes such as negative/privative -om and the collective plural -eat. It seems clear, then, that the ategorical feature of the nominalizing particle hè is [+NOMINAL] (as is the feature of the pronoun hè, from which the nominalizer descends). But what is the ategorical feature of the clausal complement of hè, which contains the same functional projections found in finite matrix contexts?

All evidence indicates that the Tuparí TP, EvidP and AuxPs are unabashedly [+VERBAL].\footnote{Equivalent functional projections in other South American languages may in fact be [+N], as shown by recent scholarship on the phenomenon of NOMINAL TENSE. See Tonhauser (2007) and Thomas (2014) for debate on Guaraní.}

As we have described in prior work (Singerman 2018a, Singerman 2018b: chapter three), Tuparí makes a strict division between the verbal and nominal domains. Lexical roots are strictly classified as either [+v] or [+N]; few are bi- or acategorical. Functional morphology is similarly split, with little to no overlap between nominal morphemes (case suffixes; negative/privative -om; plural -eat) and verbal ones (resultative and evidential suffixes; tense suffixes, particles, and auxiliaries). Now, for a verbal root to combine with nominal functional morphology, an overt process of deverbal nominalization must first apply. We saw this process in §3.1 for the actor nominalizer suffix -at/an and the object nominalizer prefix iy/y. And for a nominal root to combine with verbal functional morphology, it must first undergo an overt process of denominal verbalization. The suffix -nê is what carries out this process in (16c) (§3.2): -nê makes it possible to combine the nominal root puop ‘smart, knowledgeable’ with the aspectual auxiliary aka [ AUXVNDST-PL], the plural evidential suffix -psira and the tense particle =õpot ‘DISTANT.PST’. This same suffix is also at work in the pair of utterances in (4) (§2), where it converts the nominals o’apay ‘my paternal aunt’ and ototo ‘my grandfather/male ancestor’ into predicates capable of combining with tense, aspectual and evidential morphology. The obligatory presence of verbalizing morphology in examples like (4) and (16c) demonstrates that tense, evidentiality and aspect in Tuparí are [+VERBAL] categories. So just as surely as the nominalizer hè is [+N], it selects for a complement that contains [+v] functional material. In this sense finite embedded clauses in Tuparí are the ategorical inverse of the head-final German VPs illustrated in (50), where a [+v] transitive verb selects for a [+N] object. As there is a switch in ategorical feature between the head and the complement in both the Tuparí finite embedded clauses and the German VPs, FOFC is not violated in either case.

So Tuparí finite embedded clauses comply with the revision of the FOFC that Biberauer et al. (2014) advance – despite the superficial FOFC violation that occurs whenever one of those embedded clauses contains a second position tense and/or clause-typing particle (as in 12b, 16c, 25, 27a, 29a, 30, 31 and 49). In fact, reconsidering the Tuparí data in light of this revision to FOFC reveals an empirical contrast undiscovered in prior work: the Tuparí CP exhibits different ategorical behavior than do the TP, EvidP and AuxPs. As discussed above, a nominal root must undergo overt verbalization in order to combine with tense, aspectual and evidential morphology – all of which is [+v]. But the same is not true for the second position clause-typing particles: these freely occur with non-verbalized nominal predicates. (51) provides representative examples, with the particles and nominal predicates highlighted.
Clause-typing particles freely combine with non-verbalized nominal predicates

a. CONTEXT: A speaker tries to ascertain the identity of his interlocutor over a static-filled radio transmission.

\[
\text{Eder } \text{né } '\text{en} \\
\text{Eder } = \text{né } = '\text{en} \\
\text{Eder } = \text{YES}/\text{NO } = \text{2SG} \\
'\text{Are you Eder}?' \\
\text{conversation: 2016-12-06}
\]

b. CONTEXT: A speaker laments that her pet parakeets have flown away.

\[
\text{Kurup'i'om } \text{ta } '\text{on}. \\
kurup'i'-'om = \text{ta}'a = '\text{on} \\
\text{parakeet-PRIV } = \text{ASSERTIVE. } \text{♀ } = \text{1SG} \\
'\text{I am indeed parakeet-less.}' \\
\text{conversation: 2017-08-21}
\]

c. CONTEXT: A speaker and I joke about whether we are related; she hedges her bets.

\[
\text{Hè } \text{nàkòp } '\text{en}. \\
\text{hè } = \text{nàkòp } = '\text{en} \\
\text{PRON } = \text{MAYBE } = \text{2SG} \\
'\text{You may be that thing } [= \text{my relative}].' \\
\text{conversation: 2015-11-11}
\]

The predicate in (51a) is a proper name; in (b), it is \text{kurup'i'om} ‘parakeet-less, without a parakeet’ (built with negative/privative -'om); in (c), it is the pronoun \text{hè} (the diachronic ancestor of the clausal nominalizer). These nominals would need to take overt verbalizing morphology to combine with tense, aspectual, or evidential morphology, just as the nominals \text{o'apay} ‘my paternal aunt’ and \text{ototo} ‘my grandfather/male ancestor’ do in (4). But verbalization is not necessary – nor, to the best of my knowledge, even possible – when a nominal predicate combines directly with a clause-typing particle. So there exists a striking distinction between tense/aspectual/ evidential morphology, on the one hand, and the clause-typing particles, on the other: TAME morphology combines only with [ + verbal] predicates (thereby requiring nominal predicates to undergo verbalization) whereas the clause-typing particles are indifferent to the predicate’s categorial feature. From this I conclude that even though the TP – the projection immediately beneath CP in the Tuparí clause – is explicitly [ + v], C itself is unspecified for category. (52) presents the tree that was first given in (24), now annotated with categorial features:

(52) The structure of finite embedded clauses in Tuparí, now annotated for the categorial features [ + nominal] and [ + verbal]; the CP’s acategoriality is indicated with empty brackets.
The acategoriality of the Tuparí CP is in keeping with broader observations concerning the crosslinguistic behavior of particles. Biberauer (2017) discusses how particles may fail to conform to the predictions of FOFC, noting that they may not occupy a fixed position in the clause, often appear to lack a dedicated categorial feature, and are frequently invariant in form. The Tuparí clause-typing particles do have a fixed position in the clausal spine, in contrast to many of the particles that Biberauer surveys. But aside from the speaker-oriented gender indexicality of = pa’a ‘ASSERTIVE.♂’ and = ta’a ‘ASSERTIVE.♀’, the clause-typing particles are all formally invariant. In this sense they could not be more different from the post-verbal auxiliaries or from the resultative and evidential suffixes, all of which agree with the subject in number. The resultative suffix and certain auxiliaries even agree with singular subjects in terms of physical posture.17 The highlighted morphology in (53) illustrates. Note that both the resultative suffix and the lower auxiliary in (a) reflect the horizontal posture of the singular subject, whereas all postural information is neutralized when the subject is plural, as in (b).

(53) \[ [+\text{VERBAL}] \text{functional morphology agrees with the subject} \]
\[ a. \quad \text{Wapsikatsâ} \quad \text{oýâ} \quad \text{o’apteka.} \]
\[ \text{w-apsikat-sê-a} \quad \text{o-yê-a} \quad \text{o-’apteka} \]
\[ 1\text{SG-think-RSLT} \text{SG-TH} \quad 1\text{SG-AUX} \text{SG-TH} \quad 1\text{SG-AUX} \text{PRES.HABITUAL} \text{SG} \]
\[ ‘I am regularly thinking about it, sitting down.’ \]
\[ \text{conversation: 2018-08-09} \]
\[ b. \quad \text{Oteapsikatsira} \quad \text{otea} \quad \text{oteapteka.} \]
\[ \text{ote-apsikat-sîra-a} \quad \text{ote-a-a} \quad \text{ote-’apteka} \]
\[ 1\text{PL.EXCL-think-RSLT.PL-TH} \quad 1\text{PL.EXCL-AUX.PL-TH} \quad 1\text{PL.EXCL-AUX} \text{PRES.HABITUAL} \text{PL} \]
\[ ‘We-EXCL are regularly thinking about it.’ \]
\[ \text{elicitation: 2017-08-30} \]

While most of the predicate-final tense suffixes and tense particles do not agree with the subject, the polite future does: = ke is used with the second and third persons and = ko, with the first person singular and first person plural exclusive. And the inclusive first person consists of special portmanteaus that conflate the tense morpheme with the nominative enclitic: = kit ‘POLITE.FUT + 1DUAL.INCL’, = kitwat ‘POLITE.FUT + 1PL.INCL’ (Singerman 2020: 459–460). The clause-typing particles, however, never agree with the subject. In this respect, too, we see a clear distinction between \([+V]\) tense, aspectual, and evidential morphology and the acategorial C projection.

6.4 Summary

This section has discussed the syntax of Tuparí finite embedded clauses in light of recent literature that seeks to identify the structural domains within which the Final-over-Final Condition applies. If FOFC is restricted to apply only within Extended Projections, per Biberauer et al. (2014), then the surface FOFC violations instantiated by the language’s finite embedded clauses cease to be theoretically problematic: the head-final projection headed by hé is \([+\text{NOMINAL}]\) and does not belong to the same Extended Projection as the \([+\text{VERBAL}]\) functional material that hé embeds. This discussion has also led to novel observations about the categorial nature of the functional projections that make up the Tuparí clause. In particular, whereas the language’s auxiliary projections, EvidP, and TP are unambiguously \([+\text{VERBAL}]\), its C projection – realized as second position clause-typing particles – is in fact unspecified for category.

7 Conclusion

In this paper I have endeavored to provide both a synchronic analysis and a diachronic explanation for Tuparí finite embedded clauses, a historical innovation unique to this language within the Tuparí branch of the Tupían family. These embedded clauses resemble matrix clauses in many respects: for example, they retain the full set of evidentiality and tense contrasts and

17 I use the term ‘agree’ here in a broad way to include suppletion, which is extensive in Tuparí verbal and auxiliary roots.
they also maintain matrix clause’s second position effects. The myriad resemblances between embedded and matrix clauses make sense given the particular grammaticalization pathway through which the language developed finite embedding: the nominalizer hè grammaticalized from a third person pronoun that was deaccented following a prior clause. This proposal is supported by non-elicited data where, putting aside prosodic differences, hè can be parsed either as the historically conservative pronoun or as the historically innovative nominalizer. But there are also cases where only a hypotactic parse is possible. Such cases highlight the degree to which finite embedding has taken hold in the competence of contemporary Tuparí speakers.

Turning to broader typological considerations, we have seen that Tuparí embedded clauses instantiate the crosslinguistically rare syntactic configuration of [\text{vp} \ [\text{yp} \ Y \ ZP \ ] \ X \ ] (where \(X = N\) and \(Y = T/C\)). This configuration violates Holmberg’s (2000) original formulation of the Final-over-Final Constraint but is unproblematic under more recent revisions to FOFC, in particular the restriction of the Condition to Extended Projections in the sense of Grimshaw. Of course, why FOFC should apply only within Extended Projections (or, if Erlewine 2017 is on the right track, only at phase boundaries) remains open to debate. The analysis offered by Biberauer et al. (2014) utilizes the LINEAR CORRESPONDENCE AXIOM of Kayne (1994). This axiom posits a universal underlying structure of Specifier-Head-Complement, with the result that any surface head-finality must be derived from movement from a head-initial base. Using the LCA to derive FOFC is not without problems, however; for instance, Zeijlstra (2016) argues that Biberauer and colleagues’ Kaynian analysis requires contradictory assumptions concerning the number of specifiers that a projection may have. More broadly, Abels & Neeleman (2012) argue that adopting the LCA forces the syntactician to loosen restrictions on movement – in which case the advantages of the Kaynian approach are cancelled out. My own view agrees with Abels and Neeleman: while in some cases surface head-finality does seem to derive from underlying head-initiality, it is difficult to maintain this position in all circumstances. (See also Takita 2009, who uncovers empirical distinctions between derived head-finality in Mandarin and underlying head-finality in Japanese, and Halm 2021, who provides evidence that the base order of the Hungarian VP must be head-final.) In light of these and other controversies surrounding the LCA, this paper has not sought to provide a Kaynian formalization for FOFC but has instead asked where Tuparí fits into the overall typological landscape.

Many languages with internally headed relative clauses use subordinators that are homophonous with light nouns (‘thing’, ‘stuff’, ‘one’) or third person pronouns (Hiraiwa 2017). Hanink (2021) shows that this is the case for the North American isolate Washo, in which the morpheme responsible for subordinating clauses is simply an unstressed version of the third person pronoun. She proposes that this resemblance is not an accident of diachrony but instead the consequence of how indices behave as formal objects in synchronic syntax. In line with Hanink’s approach to Washo, one reviewer suggests that there is no nominalizing hè (as I have assumed throughout this paper) but that hè is in all instances just the third person pronoun; this pronoun in turn could then select for a finite clause which is nominalized by a null functional head. That analysis is attractive in that it would reduce the two versions of hè to just one, though it would require positing a null nominalizing head for which other evidence is lacking. Furthermore, to my knowledge pronouns in Tuparí cannot take complements, nominalized or otherwise. The reviewer points out that a possible way to distinguish between the two analyses concerns the availability of first and second person internal heads inside of the clausal nominalizations built with hè. If hè is in all cases just the third person pronoun, first and second persons ought not to be able to function as internal heads. I do not have data that bear on this point, though this may be an accidental limitation of my corpus; future field research will need to investigate this possibility. (For an example of a South American language that does allow speech act participants to serve as heads in internally headed relatives, see Salanova 2011 on Mèbengokre, of the Jê family.)

In sum, it is possible that Hanink’s analysis of Washo could be extended to Tuparí. For now, however, I choose to explain the resemblance between nominalizing hè and pronominal hè not synchronically but diachronically, as a consequence of the particular pathway through which these finite embedded clauses grammaticalized.
A Deriving a surface FOFC violation

This appendix offers a derivation of the surface FOFC violation in (54), which was presented in §6.1 as example number 49. This is a transitive VP headed by *ma'ẽ* ‘speak of, talk about’. The verb’s direct object is an entire finite embedded clause that itself contains the second position tense particle = *kut* ‘ANCIENT.PST’.

(54)  
okiot  kut  tenon  wàt’omnā  hè  ma’ã
      [  okio-t  =kut  te-nō-n  wàt’om-nē-a  ]  hè  ma’ã-a
      [  man-NUC  =ANCIENT.PST  3c-friend-NUC  poison-EV.SG-TH  ]  Hè  talk-about-TH
  ‘talk about the man who poisoned his own friend (NON-WITNESSED)’

the title of a text written by Raul Pat’awre Tupari

Since this example does not include any aspectual morphology or aspectual auxiliaries, the derivation given here does not include the Resultative Phrase, the *aux.go*/*auxpositional* projection, or the *auxhabitual* projection, all of which are included in (3), (24) and (52). The derivation given here also excludes the internal structure of the NP *tenon* ‘his own friend’, which is the object of the embedded verb *wàt’om* ‘poison, give poison to’.

Step #1 of the derivation: Merge the embedded verb *wàt’om* ‘poison, give poison to’ with its complement, the NP *tenon* ‘his own friend’.

$$
\begin{align*}
\text{VP} & \\
\text{NP} & \quad V' \\
\text{tenon} & \quad \text{‘his own friend’} \\
\text{wàt’om} & \quad \text{‘poison’}
\end{align*}
$$

Step #2: Merge the subject NP *okiot* ‘the man’ in Spec,V. (Subjects are typically introduced in Spec,v in contemporary syntactic theorizing; this detail is not crucial here.)

$$
\begin{align*}
\text{VP} & \\
\text{NP} & \quad V' \\
\text{okiot} & \quad \text{‘the man’} \\
\text{tenon} & \quad \text{‘his own friend’} \\
\text{wàt’om} & \quad \text{‘poison’}
\end{align*}
$$

Step #3: Merge the evidential suffix, the head of the EvidP. Because this example does not include aspectual auxiliaries, the EvidP is directly above the VP itself. (I present Evid$^0$ here as realized by -nã, which technically conflates the evidential suffix and the theme vowel -a. As described in Singerman 2018b: 384–388, the theme vowel has the phonological effect of deleting an immediately preceding /e/.)

$$
\begin{align*}
\text{EvidP} & \\
\text{VP} & \quad \text{Evid} \\
\text{NP} & \quad V' \\
\text{okiot} & \quad \text{‘the man’} \\
\text{tenon} & \quad \text{‘his own friend’} \\
\text{wàt’om} & \quad \text{‘poison’}
\end{align*}
$$
**Step #4:** Merge the Tense Phrase. As argued in Singerman (2020), the underlying headedness of the Tuparí TP is obscured by post-syntactic operations and is therefore indeterminate. In principle, the TP could be underlyingly head-final:

```
TP
  /\             /\     
EvidP  T         Evid
  /\     /\       /\     
VP    Evid 'ANCIENT.PST'
  /\               /\   
NP    -nā
  /\       /\     
V'       V
  /\     /\     
okiot 'the man'  tenon 'his own friend'  wā't'om 'poison'
```

Alternatively, it could be underlyingly head-initial:

```
TP
  /\     
EvidP  T
  /\     /\     
VP    Evid ' ANCIENT.PST'
  /\     /\         /\     
NP    V'    -nā
  /\       /\     
okiot 'the man'  tenon 'his own friend'  wā't'om 'poison'
```

In the subsequent steps of the derivation I show the TP as underlyingly head-final, though nothing crucial hinges on this representational choice.

**Step #5:** Merge the CP, which is head-initial. The C head is systematically null in declaratives; however, as argued in Singerman (2020), Head Movement from T₀ to C₀ is what brings the tense particles (including = kut 'ANCIENT.PST') to second position.

```
C'
  /\     
C      TP
  /\             /\     
EvidP  T         Evid
  /\     /\       /\     
VP    Evid 'ANCIENT.PST'
  /\     /\               /\     
NP    V'    -nā
  /\       /\     
okiot 'the man'  tenon 'his own friend'  wā't'om 'poison'
```
Step #6: Move the subject *okiot* from its base position (Spec,V) to its derived position, Spec,C. This completes the construction of the embedded clause.

Step #7: Merge the nominal head *hè*, which takes the entire embedded CP as its complement.

Step #8: Merge the transitive verb *ma’ẽ* ‘speak of, talk about’, which selects the entire finite embedded clause as its object.
The tree in (55) presents the output of Step #8 but with the categorial features [+ NOMINAL] and [+ VERBAL] now annotated, as in (52). Note the acategorial CP, sandwiched in between the embedded clause’s [+ VERBAL] TP and the NP headed by hè. Note also the recursive head-finality on the level of the VP. Inside of the embedded clause, the direct object tenon ‘his own friend’ immediately precedes the transitive verb that selects for it, wàt’om ‘poison’; and in the matrix clause, the entire embedded clause precedes the verb that selects for it, ma’ẽ ‘talk about’.

(55) Arboreal representation of example (54), now annotated with categorial features

```
    V' [+v]
       NP [+N]   V [+v]
       CP [ ]
       NP [+N]   N [+N]  ma'ẽ
c                   C' [ ]
okiot
                   C [ ]
                   TP [+v]
kut
                   EvidP [+v]
                   T [+v]
                   ma'ã
                   Evid [+v]  ka'i
                   VP [+v]
                   ma'ẽ
                   V' [+v]
                   -nã
                   tenon
                   wàt'om
                   NP [+N]   V [+v]
                   VP [+v]
                   VP
                   VP
                   VP
                   VP

The matrix VP can be used felicitously on its own, as an imperative meaning ‘Talk about the man who poisoned his own friend (NON-WITNESSED)!’ Or it can be used in a sentential context like (56a). Structurally, the matrix clause in (56a) is just like (56b), where the direct object of the verb is the pronominal proclitic e- ‘2SG’.

(56) a. Okiot kut tenon wàt'omnã hè
[ [ [ okio-t = kut te-nô-n wât'omnê-a ] hè ]
[ v' i3g [ v man-NUC = ANCIENT.PST 3C-friend-NUC poison-EV.SG-TH ] HÈ ]
ma'ã  ko  'on.
ma'ê-a  ] = ko  = 'on
talk.about-TH ] = POLITE.FUT = 1SG
‘Let me talk about the man who poisoned his own friend (NON-WITNESSED).’

b. Ema'ã ko 'on.
[ v' e-ma'ê-a ] = ko  = 'on
[ v 2SG-talk.about-TH ] = POLITE.FUT = 1SG
‘Let me talk about you.’ [ = ‘I will say hello for you.’]

Abbreviations

3C coreferential/reflexive third person
ADV.FOC adverbal focus suffix
AUX auxiliary
AUXco auxiliary series related to the lexical verb ‘go’
AUXhabit habitual auxiliary
DECL declarative
DEF definite
Ethics and consent

The author’s field research in Rondônia has been approved by the Institutional Review Board at the University of Chicago (submission number IRB13-0009), by the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (process number 000457/2015-97), and by the Fundação Nacional do Índio – FUNAI (process number 08620.019511/2015-02).

Acknowledgements

Dozens and dozens of Tuparí speakers have contributed to the findings presented in this paper. I would like to express a special thank you to the many speakers who recorded interviews with me in summer 2017 and summer 2018:

- Arlene Pairosá Tupari
- Cristiane Tupari
- Daiane Arikapu
- Eliane Osi’uy Tupari
- Elizangela Waiká Tupari
- Elizete Tupari
- Fabiano Amôta Tupari
- Flavio Konbiro Tupari
- Gabriel Akurai Tupari
- Gleicia Arikapu
- Ivan Iritop’a Tupari
- Luciano Tupari
- Marilza Kabatoá Tupari
- Marly Tupari
- Miraci Aguissi Tupari
- Neide Yôrô’êta Tupari
- Nilson Tupari
- Paulina Tomika Tupari
- Paulo Watoire Tupari
- Pedro Mâykware Tupari
- Reginaldo Abo Makurap
I would also like to thank the political leadership from the Terra Indígena Rio Branco – especially Adriano Abai Tupari, Dalton Tupari, Pedro Kup'eoyt Tupari, Samuel Tupari, and Severino Tigi Tupari – as well as multiple indigenous schoolteachers: Arlene Pairosá Tupari, Geovane Tupari, Isaias Tarimã Tupari, Juari Tupari, Mauricio Tupari, Nilson Tupari, and Raul Pat’awre Tupari. Without the support of the local leaders and schoolteachers, this research simply would not have been possible! Thank you also to the residents of the villages of Bom Jesus, Serrinha, Trindade, Nazaré, Colorado, Figueira, Cajuí, and Palhal as well to the Tuparí who live in the city of Alta Floresta D’Oeste.

For feedback on this work as it has taken shape over the past few years I thank my dissertation committee members – Lenore Grenoble, Jason Merchant, and Karlos Arregi – as well as several other faculty members and graduate students at the University of Chicago, including but not limited to Amy Dahlstrom, Andy Murphy, Emily Hanink, Erik Zyman, Jessica Kantarovich, Ming Xiang, and Salikoko Mufwene. I would also like to thank audience members at the 22nd meeting of the Workshop on Structure and Constituency in Languages of the Americas, held at the University of British Columbia; Cambridge Comparative Syntax 6; the Hungarian Academy of Sciences; the second Symposium on Amazonian Languages, held at the University of California, Berkeley; and the Language Variation and Change Workshop and the Morphology and Syntax Workshop, both at the University of Chicago.

For the crucial support that they have provided at all stages of my research, I express my deepest gratitude to my colleagues Ana Vilacy Galucio, Denny Moore, and Hein van der Voort, of the Museu Paraense Emílio Goeldi (the Brazilian institution which has served as the official sponsor for my research in Rondônia). Thanks also go to Evani Viotti and Luciana Storto, of the University of São Paulo.

Finally, a special thanks to Editor Johan Rooryck and the anonymous reviewers whose comments did so much to improve this paper. Any and all remaining errors are my responsibility alone.

**Funding information**

The author’s field research in Rondônia has been supported by a Tinker Field Research Grant from the Center for Latin American Studies at the University of Chicago, by an Individual Research Grant and a Kinkade Grant from the Jacobs Research Funds at the Whatcom Museum, by a Language Legacies Grant from the Endangered Language Fund, and by a Doctoral Dissertation Research Improvement Grant from the National Science Foundation’s Documenting Endangered Languages Program (award number 1563228). The author is at present on a year-long research fellowship awarded by NSF-DEL and funded & administered by the National Endowment for the Humanities.

**Competing interests**

The author has no competing interests to declare.

**Author note**

I would like to dedicate this paper to the memory of my dear friend and teacher Dr. Clémence Jouët-Pastré. Her enthusiasm for the Portuguese language and for Brazilian culture inspired many students, myself included. If not for her friendship and guidance I would never have learned Portuguese, let alone become a Brazilianist. Obrigado por tudo, Clémence. You are very, very missed. Saudades!

**Author affiliation**

Adam Roth Singerman orcid.org/0000-0001-9425-5095
University of Chicago (additionally affiliated with the Museu Paraense Emílio Goeldi), US
