

# The origin of split third person marking in Arara (Cariban)

Florian Matter  
University of Bern

## Abstract

Arara, a Cariban language spoken in the state of Pará, Brazil, shows several splits in its third person marking. Intransitive verbs with a third person S and transitive verbs with a 3>3 scenario are marked by either *i-/Ø/t(i)-* or *n(i)-/n(i)-*, depending on the TAM configuration. The forms with *i-/Ø/t(i)-* can be demonstrated to descend from earlier subordinate clause forms, while those with *n(i)-/n(i)-* are original main clause forms. Furthermore, these former subordinate clauses, as well as extant ones, otherwise use main clause person markers in Arara, which is unusual in the Cariban family. I show that Arara has extended main clause marking to (some) subordinate clauses for all persons, except for those with third person subjects. This resulted in a paradigm which contains prefixes from both Proto-Cariban main clauses and Proto-Cariban subordinate clauses, but has the overall organization of Proto-Cariban main clauses. This type of innovative main clause coexists with the clauses showing the original main clause paradigm, leading to the split in third person. I also show that Arara's sister language, Ikpeng has extended this new paradigm to all clause types. I further show that a similar extension has happened in the third member of the Pekodian branch, Bakairi. This is of importance for the reconstruction of Proto-Cariban third person marking. I also show that Arara preserves *n(i)-/n(i)-* in similar semantic contexts as other Cariban languages.

Keywords: Arara, Cariban, historical morphosyntax, person marking, differential argument marking

## 1 Introduction

Arara<sup>1</sup> (ISO 639-3: aap) is a Cariban language spoken in the state of Pará, Brazil. It is spoken by between 300 and 400 people in the villages of Laranjal and Cachoeira Seca (S. D. C. de Souza 2010: 5–7) on the Iriri river, a tributary of the Xingu. Language use is classified as “vigorous” by Simons & Fennig (2019); while there is widespread bilingualism with Portuguese, Arara is the preferred choice for in-group communication (S. D. C. de Souza 2010: 7).

Within the Cariban family, Arara is classified as a member of the Pekodian branch, along with Ikpeng and Bakairi (Figure 1). Earlier classifications usually grouped the latter with Upper Xingu Carib<sup>2</sup> (Girard 1971; Durbin 1977; Kaufman 1994, 2007; Mattéi-Müller 2002; Gildea 2005). When better descriptive data for Ikpeng became available, Meira & Franchetto (2005) established the Pekodian branch as distinct from Upper Xingu Carib, the latter being in a branch on its own. Ikpeng is clearly closely related to Arara, the similarity bordering on co-dialect status (Meira & Franchetto 2005: 130), although there are some differences, one of which will be discussed here (see Ferreira-Alves et al. [2019] for differences in kinship terms).

For quite some time, there was no trace of Arara speakers, and the language was assumed to be extinct (I. C. de Souza 1988: 1). However, their continued existence was discovered in the late 1970s, and contact was established in 1983 (I. C. de Souza 1988: 2). Since then, descriptive work on the language has been carried out, resulting in a notable body of material available today. There are two works on phonology (I. C. de Souza 1988; Alves 2010), a sketch of morphology (S. D. C. de Souza 1993), and contributions on nominal (S. D. C. de Souza 2010) and verbal morphology (Alves 2017). The data used for this

<sup>1</sup>[a'ɾara], autonym [ugɔɾɔŋ'mɔ] (*ugoro-ŋmo* '1+2PRO-PL') (S. D. C. de Souza 2010: 3–4).

<sup>2</sup>Also known as Kuikuro-Kalapalo, a group of closely related varieties spoken in the Upper Xingu area (Franchetto 2010: 123).

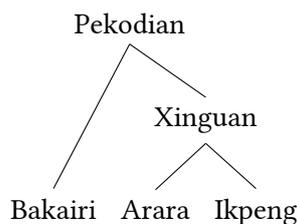


Figure 1: The Pekodian branch, a top-level member of the Cariban family.

paper primarily come from Alves (2017) and S. D. C. de Souza (1993), the two works dealing with verbal morphology.

This paper is concerned with the diachrony of a specific corner of Arara verbal morphosyntax, namely third person marking, and is structured as follows: In Section 2, I introduce the fairly standard Cariban person marking system of Arara, and present the wide array of third person morphs that are found in it. In Section 3, I describe the morphosyntactically conditioned third person split between  $n(i)/n(i)$ - and  $i/\emptyset/t(i)$ - found both in intransitive and in transitive 3>3 scenarios. I demonstrate that the clauses with  $i/\emptyset/t(i)$ - originate from subordinate clauses that have been reanalyzed as main clauses. In Section 4, I further demonstrate that the otherwise identical person marking in main and subordinate clauses stems from a partial extension of main clause inflection to subordinate clauses in Arara. I also discuss the development in Ikpeng, which has undergone further changes, and I show that while Bakairi did not undergo the exact same partial extension, it underwent a similar one. In Section 5, I discuss the consequences of these suggested changes in the Pekodian languages for the reconstruction of Proto-Cariban verbal person marking. I also show that some of the contexts in Arara where  $n(i)/n(i)$ - is preserved have surprisingly similar semantics to the contexts where reflexes of the third person marker \* $n$ - are found in languages of the Pemongan group. In Section 6, I briefly summarize my findings.

## 2 Third person marking in Arara

Arara has a verbal person marking system common for Cariban languages. In most languages in the family, the person markers found in transitive scenarios can either be characterized as showing a hierarchical organization,<sup>3</sup> meaning that only the higher-ranking participant on a hierarchy  $1/2 > 3$  is expressed, or as being portmanteau prefixes marking both participants simultaneously.<sup>4</sup> The Arara instantiation of this system is shown in Table 1. It can be described as follows, using terminology from the hierarchically-based analysis: In transitive verbs, there is one series of P-marking prefixes, which occurs in 3>1, 3>2 and 3>1+2,<sup>5</sup> i.e. INVERSE scenarios. Another series is A-marking and occurs in 1>3, 2>3 and 1+2>3, i.e. DIRECT scenarios. 1>2 and 2>1, i.e. LOCAL scenarios, are expressed with a dedicated portmanteau morpheme and the same marker as 3>1+2, respectively. 3>3, i.e. NONLOCAL scenarios, are the only place where purely third person markers are featured. Intransitive verbs show a lexically and morphologically conditioned split between  $S_A$  and  $S_P$  verbs, in which, for  $SAP$  arguments,  $S_A$  verbs show prefixes similar to the A-marking ones in transitive verbs, and  $S_P$  show the P-marking prefixes. A notable exception is first person marking, where  $1S_A$  marking is entirely distinct from 1>3 marker, and shows an additional, also lexically conditioned, split.

Of interest for the present paper are the third person prefixes, of which there are quite many. These markers are phonologically, lexically, and morphosyntactically conditioned, as shown in Figure 2. Usually, C-initial verbs show V-final prefixes, with V-initial verbs lacking this vowel; the exception are  $S_A$

<sup>3</sup>For the hierarchical analysis, see for example Hawkins (for Waiwai, 1998: 27), Meira (for Tiriyó, 1999: 286), Tavares (Wayana, 2005: 209), Cáceres (for Ye'kwana, 2011: 210), and T. E. Payne & D. L. Payne (for Panare, 2013: 197).

<sup>4</sup>For the portmanteaux analysis, see for example Derbyshire (for Hixkaryana, 1985: 32), E. Koehn & S. Koehn (for Apalaí, 1986: 107), and Hoff (for Kari'ña, 1995: 347)

<sup>5</sup>The presence of the person value 1+2 implies the presence of 1+3. In Cariban languages, this value is expressed with a free pronoun in combination with third person morphology, and will not be discussed here.

Table 1: The Arara person marking system

(a) Transitive					(b) Intransitive		
A/P	1	2	1+2	3	S <sub>A</sub>	S <sub>P</sub>	
1		<i>ko-</i>		<i>j-/in(i)-</i>	1	<i>w-/k-</i>	<i>j-/i-</i>
2	<i>ugu-</i>			<i>m(i)-</i>	2	<i>m-</i>	<i>o-</i>
1+2				<i>kut(i)-</i>	1+2	<i>kut-</i>	<i>ugu-</i>
3	<i>j-/i-</i>	<i>o-</i>	<i>ugu-</i>	<i>i-/Ø/t(i)- or n(i)-/n(i)-</i>	3	<i>Ø or n-</i>	<i>Ø/i-/t- or n(i)-</i>

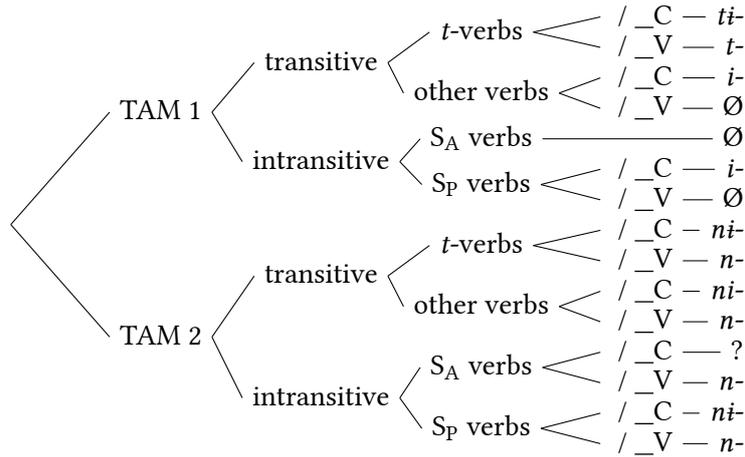


Figure 2: The distribution of Arara third person markers

verbs, where the only C-initial S<sub>A</sub> verb, *ke* ‘to say’, is zero-marked, rather than taking *i-*. There are two lexical splits, one found in transitive, the other in intransitive verbs. A group of transitive verbs which I call *t*-verbs takes the markers *t(i)-* and *n(i)-*, while other transitive verbs take *i-/Ø* and *n(i)-*. Intransitive verbs show a split between S<sub>A</sub> (*Ø* and *n-*) and S<sub>P</sub> (*Ø/i-* and *n-*) verbs, which is purely morphological/lexical, and found throughout Cariban (Meira 2000).<sup>6</sup> The last split, which is the main focus of this paper, is the one between *i-/Ø/t(i)-* on the one hand, and *n(i)-/n(i)-* on the other. It is conditioned by the TAM configuration of the clause.

Before moving on to the discussion of this morphosyntactically conditioned split in Section 3, I will briefly illustrate the lexical split between *t*-verbs, with *t(i)-* and *n(i)-*, and other transitive verbs, with *Ø/i-* and *n(i)-*. (1) demonstrates the lexical split in 3>3 scenarios with the regular transitive verbs *eneŋ* ‘to see’ and *moni* ‘to moor’ and the *t*-verbs *oŋoŋ* ‘to bite’ and *wo* ‘to kill’.

- (1) a. *Ø-eneŋ-ri*  
 3>3-see-REC  
 ‘He saw it/him/her.’ (Alves 2017: 61)
- b. *i-moni-ni*  
 3>3-moor-REC  
 ‘He moored it.’ (Alves 2017: 49)
- c. *t-oŋoŋ-takpi-li*  
 3>3-bite-RESUM-REC  
 ‘S/he/it stopped biting her/him/it.’ (Alves 2017: 123)

<sup>6</sup>The only C-initial S<sub>A</sub> verb, *ke* ‘to say’ is not attested with a *n(i)-/n(i)-* prefix; I suspect that it either shows *n-*, or alternatively a V-initial allomorph similar to Ikpeng *an-*, see Section 4.2.

- d. *ti-wo-li*      *papa*  
 3>3-kill-REC    dad  
 ‘Dad killed it.’ (S. D. C. de Souza 1993: 24)

The  $\_V$  allomorphs  $\emptyset$  and *t-* are shown in (1a) and (1c), the  $\_C$  allomorphs *i-* and *ti-* in (1b) and (1d). The same lexical split is found within the *n(i)-/n(i)-* group of third person markers, in (2) also illustrated with 3>3 scenarios.

- (2) a. *n-eneŋ-a*  
 3>3-see-PERM  
 ‘Let him see (it).’ (S. D. C. de Souza 1993: 75)
- b. *ni-momi-a*  
 3>3-moor-PERM  
 ‘Let him moor (it).’ (Alves 2017: 49)
- c. *nok ni-wo*  
 who 3>3-kill  
 ‘Who killed him?’ (S. D. C. de Souza 1993: 19)

The  $\_V$  allomorph of the *n(i)-/n(i)-* set is always *n-* (2a). Verbs taking *i-/∅* in the *i-/∅/t(i)-* set take *ni-* in the *n(i)-/n(i)-* set, like *momi* ‘to moor’ (1b, 2b). Verbs taking *t(i)-* in the *i-/∅/t(i)-* set take *ni-* in the *n(i)-/n(i)-* set, like *wo* ‘to kill’ (1d, 2c).

### 3 The morphosyntactically conditioned third person split

The main split under discussion here is at the root of the tree in Figure 2, and was already partially illustrated in (1) and (2), where the verb forms in (1) show the *i-/∅/t(i)-* prefixes, but those in (2) the *n(i)-/n(i)-* prefixes. The distribution of *i-/∅/t(i)-* and *n(i)-/n(i)-* can be described as conditioned by the TAM configuration of the clause. S. D. C. de Souza (1993: 12) calls the configurations with *i-/∅/t(i)-* REAL, and those with *n(i)-/n(i)-* VIRTUAL, but does not discuss this difference in detail. *i-/∅/t(i)-* are listed as occurring in interrogatives, permissives, and negated clauses. However, the sole example for a negation with *n-* (3a) is likely a misinterpreted instance of the negative prefix *on-*. This prefix regularly occurs on negated intransitive verbs, replacing all person markers (Alves 2017: 79), as illustrated in (3b)

- (3) a. *n-aŋde pira*  
 3-fall NEG  
 ‘He didn’t fall.’ (S. D. C. de Souza 1993: 17)
- b. *on-orik-pira*      *piu*  
 NEG-dance-NEG P.  
 ‘Piu didn’t dance.’ (Alves 2017: 79)

This only leaves permissive and interrogative constructions as members of S. D. C. de Souza’s (1993) “virtual” TAM group. The permissive, already seen in (2a–b), can be translated with ‘let him/her/them V’, and is formed with a suffix *-a*. Due to its restriction to third person S/A, the permissive only shows a small subset of available person markers, namely the inverse markers in transitive scenarios (4a–b), and *n(i)-/n(i)-* in 3S and 3>3 scenarios (4c) (Alves 2017: 165).

- (4) a. *j-eneŋ-a*  
3>1-see-PERM  
'Let him see me.' (Alves 2017: 166)
- b. *ug-eneŋ-tom-a*  
3>1+2-see-PL-PERM  
'Let them see us.' (Alves 2017: 166)
- c. *n-eneŋ-a*  
3>3-see-PERM  
'Let him see (it).'
- (S. D. C. de Souza 1993: 75)

The other context triggering *n(i)/n(i̇)-* listed by S. D. C. de Souza (1993) are interrogatives, which are not readily definable via their TAM ending. They are overtly marked either by the presence of the verbal suffix *-taŋ*, the question particle *ka*, a question word, or some combination thereof. The reasons for choosing between *-taŋ*, *ka*, or both, are not entirely clear, but might have something to do with past and non-past readings (Alves 2017: 131). In any case, *n(i)/n(i̇)-* is present regardless of what (combination of) free or bound morpheme(s) is used, as seen in (5).

- (5) a. *ni-wo ka*  
3>3-kill Q  
'Did s/he kill him/her?' (S. D. C. de Souza 1993: 19)
- b. *nok n-enep*  
who 3>3-bring  
'Who brought him?' (S. D. C. de Souza 1993: 7)
- c. *warite n-i-taŋ od-anpiri-tpot*  
how 3-COP-INTER DETRZ-start-NMLZ  
'How will it be in the beginning?' (Alves 2017: 91)
- d. *nok n-ipton-dan*  
who 3-go.down-INTER  
'Who will go down?' (S. D. C. de Souza 1993: 73)

(5a) only contains the question marker *ka*, the verb bearing *ni-*. In (5b), a question word *nok* 'who' marks the sentence as interrogative, again with *n-* appearing on the verb. In (5c–d), we see a combination of a question word (*warite* 'how', *nok* 'who') and an overt interrogative suffix *-taŋ* on the verb, both with *n-*.

To these contexts, I add two other TAM configurations not described by S. D. C. de Souza (1993). They trigger third person prefixes that do not have the form *n(i)/n(i̇)-*, but (at least diachronically) contain *n-*, with additional material before *n-*. These are the remote past and the intermediate past, which have the prefixes *ku(-)n-*, shown in (6a), and *mo(-)n-*, shown in (6b).<sup>7</sup> When third person is not marked on the verb, the remote past is expressed by *-n(e)* (6c), and the intermediate past by *-t(e)* (6d).

- (6) a. *modo ipe kun-iŋfi-tak*  
WORM EXIST 3.REM-COP-PST.IPFV  
'There were worms.' (Alves 2017: 194)
- b. *mon-ipi-taŋ-tom*  
3INT-bathe-ASP-PL  
'They were bathing.' (Alves 2017: 176)

<sup>7</sup>Cognates of Arara *ku-* are found together with past tense suffixes in Apalaí, Tiriyo, Ye'kwana, and Werikyana, and without suffixes in Wayana; in Kari'ña both variants occur, with an as of yet not conclusively analyzed modality value (Yamada 2011; Hoff et al. 2017). The diachronic path suggested by Gildea (1998: 97–99) is the addition of an element *\*ki-* to the third person marker *\*n-* in remote past, with subsequent reanalysis of *\*kin-* as a third-person *cum* tense marker, and loss of the tense suffixes. Likely cognates of Arara *mo(-)n-* occur in Wayana (Tavares 2005: 209–210) and Hixkaryana (Derbyshire 1985: 7).

- c. *m-omomi-ti-n*  
 2S<sub>A</sub>-enter-PL-REM  
 ‘You entered.’ (Alves 2017: 177)
- d. *mobu j-anumi-taη-te i-mobu-n lon*  
 canoe 1>3-raise-ASP-INT 1-canoe-PERT FOC  
 ‘I raised my own canoe.’ (Alves 2017: 112)

The *i-/Ø/t(i)-* prefixes are the ones that are represented in the person marking overview by Alves (2017: 158), and such forms prominently occur in her examples. The more frequent TAM suffixes triggering *i-/Ø/t(i)-* are the past *-li*<sup>8</sup> and the imperfective *-naηri*. Alves (2017) calls *-li* the ‘immediate past’, S. D. C. de Souza (1993) the ‘recent past’; I will employ the latter label. While this gloss suggests a contrast of the recent past with the intermediate and remote past discussed above, *-li* also occurs in cases where the event lies further in the past, indicating that there are maybe factors other than recency at play as well (Alves 2017: 110). 3>3 scenarios were already illustrated in (1); other person markers in the recent past are illustrated in (7).

(7) Some transitive verbal markers (Alves 2017: 174, 144–146)

- a. *kut-monogou-ni-ηmo*  
 1+2>3-wait.for-REC-PL  
 ‘We waited for them.’
- b. *o-daη-ri*  
 3>2-hear-REC  
 ‘S/he heard you.’
- c. *uk-taη-ri*  
 2>1-hear-REC  
 ‘You heard me.’
- d. *ko-daη-ri*  
 1>2-hear-REC  
 ‘I heard you.’

The other frequently occurring TAM suffix with *i-/Ø/t(i)-* prefixes is the imperfective, formed with *-naηri*. This is an atemporal form, occurring with past (8a), present (8b), and future (8c) interpretations, as discussed by Alves (2017: 117–119).

- (8) a. *kok ʃimna i-beη-naηri*  
 at.night 1+3PRO 3S<sub>P</sub>-run-IPFV  
 ‘We (used to) run at night.’ (Alves 2017: 119)
- b. *Ø-abu-naηri-ηmo*  
 3>3-dig-IPFV-PL  
 ‘They’re digging it.’ (Alves 2017: 174)
- c. *koglon w-ido-naηri*  
 tomorrow 1S<sub>A</sub>-go-IPFV  
 ‘I will go tomorrow.’ (Alves 2017: 119)

A less frequent third main clause form with *i-/Ø/t(i)-* is the future, formed with *-tome/-tpom/-npom* (Alves 2017: 113–114), shown in (9).

<sup>8</sup> *-li* has an allomorph *-ri* after *η* (Alves 2017: 61); the Iriri dialect shows free variation between *-li* and *-ni* (Alves 2017: 60).

Table 2: Forms with  $n(i)/n(i)-$  and  $i-/Ø/t(i)-$ 

$n(i)/n(i)-$	$i-/Ø/t(i)-$
PERM <i>-a</i>	REC <i>-lí</i>
INTER <i>-taŋ</i>	IPFV <i>-naŋri</i>
REM <i>kun- / -n(e)</i>	FUT <i>-tome/-tpom</i>
INT <i>mon- / -t(e)</i>	

Table 3: Distribution of plural markers

Prefix	TAM	SAP.PL	3PL
$n(i)/n(i)-$	REM	<i>-ti</i>	<i>-tom</i>
	INT	<i>-ti</i>	<i>-tom</i>
	PERM	–	<i>-tom</i>
	INTER	<i>-tom</i>	<i>-tom</i>
$i-/Ø/t(i)-$	FUT	<i>-ti</i>	?
	REC	<i>-ŋmo</i>	<i>-ŋmo</i>
	IPFV	<i>-ŋmo</i>	<i>-ŋmo</i>

- (9) a. *j-enebi-tome mobu koglon*  
 1>3-bring-FUT canoe tomorrow  
 ‘I’ll bring the canoe tomorrow.’ (Alves 2017: 113)
- b. *Ø-eneŋ-ŋpom talem mure-n*  
 3>3-see-FUT bird.sp son-PERT  
 ‘He will then see the bird’s son.’ (Alves 2017: 67)

The main clause forms triggering  $n(i)/n(i)-$  and  $i-/Ø/t(i)-$ , respectively, are summed up in Table 2. The question is, how did this distribution come about? My suggestion is that the forms with  $i-/Ø/t(i)-$  originate from earlier subordinate clauses, which are usually nominalized or adverbialized forms in Cariban languages (Gildea 1998). There are three main facts pointing towards that scenario. One is that Arara nominalizations have  $i-/Ø/t(i)-$  as third person markers, not  $n(i)/n(i)-$ . Accordingly, ‘writer of it’ in (10) does not take *ni-*, but *i-*.

- (10) Arara (Alves 2017: 97)  
*i-deke-tke-ni*  
 3>3-write-ITER-NMLZ  
 ‘The one who was writing (it).’

Another reason are the plural markers used in clauses with  $i-/Ø/t(i)-$ . There is a fairly elaborate interaction between plural marking, person, and TAM, as evidenced by Alves (2017: 137–180) devoting a whole chapter to it. She concludes (2017: 179) that the permissive, the interrogative, and the remote and intermediate past take a third person plural marker *-tom*, whereas the recent past and the imperfective take *-ŋmo*. While she lists the future *-tome/-tpom/-ŋpom* as triggering a plural marker *-ti*, there are no examples of plural third person futures in the available Arara data, only non-third. Since plural marking shows a split between SAP and third person, it is an open question whether third person plural future forms actually take *-ti*, or whether they take *-ŋmo*, like the other forms with  $i-/Ø/t(i)-$ . An overview of the different plural markers is given in Table 3.

The origins of 3PL *-tom* are not entirely clear. An apparent cognate *tomo* is used as a plural marker on nouns in e.g. Werikyana (Gildea 1998: 117), Apalaí (E. Koehn & S. Koehn 1986: 88), Ye’kwana (Cáceres

2011: 82), Tiriyó (Meira 1999: 128–129), and Wayana (Tavares 2005: 151–156). However, these morphemes seem to be purely nominal markers, whereas Arara *-tom* is purely verbal. In contrast, the origins of 3PL *-ŋmo* are very clear; it stems from the Proto-Cariban plural marker *\*komo*, which is widely used for nouns (Gildea 1998: 116–118). Thus, it is exactly the plural marker one would expect in nominalizations, my proposed source of main clause forms with *i-/Ø/t(i)-*. Although I have found no examples illustrating subordinate clauses with plural marking, evidence for the originally nominal character of *-ŋmo* abounds in Arara: Along with its allomorphs *-kom/-gom*, *-ŋmo* is the only marker used for plural marking on nouns (S. D. C. de Souza 2010: 64):

- |   |                                       |
|---|---------------------------------------|
| (11) a. <i>karāŋu-ŋmo</i><br>spoon-PL<br>'spoons' | b. <i>ugon-gom</i><br>man-PL<br>'men' |
|---|---------------------------------------|

As for the recruitment of these originally deverbal forms as main clause forms, I point to the fact that syntactic reanalysis leading to distinct different types of main clause verb forms is common in the family, which is the main topic of the landmark contribution in comparative Cariban morphosyntax, *On reconstructing grammar* (Gildea 1998). While some details of the reconstructed forms have since changed, the scenario described by Gildea (1998) is still valid: There are innovative forms that began their life as deverbalizations (potentially combined with postpositions or attributivizers), combined with an auxiliary copula, other main verb, or demonstrative (Gildea 1998: 153–236). Constructions like 'he is going' or 'this (is) my going' can lose their copular verb or demonstrative, resulting in a reanalysis of the originally deverbal forms as independent main clause forms.

Cognates of the Arara forms presented as innovative here are found elsewhere in the family. An example main clause from Upper Xingu Carib is shown in (12).

- (12) Upper Xingu Carib (dos Santos 2007: 60)  
*e-ingi-li-ko-ha eye-i kayaiha heke*  
 2P-see-PNCT-PL-AFF DEM-COP non.Indian ERG  
 'The white people/person saw you.'

Among other things, the Kuikuro verb form in (12) is recognizable as innovative by the presence of the plural marker *-ko*, another reflex of the nominal plural marker *\*komo*. Thus, the innovated Kuikuro tense-number ending *-li-ko* in (12) is cognate with Arara *-li-ŋmo*. Both stem from earlier nominalizations (Proto-Cariban *\*-ri*, either 'action nominalizer' or 'pertensive'/'possessed' [Gildea 1998: 120]), which is combined with a reflex of the plural marker *\*komo*. Other languages show similar innovative main clause forms with reflexes of *\*-ri* (Gildea 1998: 163–168).

As mentioned, while the *i-/Ø/t(i)-* third person markers in the Arara future suggest a subordinate origin, the attested plural forms do not feature *-ŋmo*. However, cognate forms can be found throughout the family, unambiguously showing that the future originated as a subordinate form, too. The suffix *-tome/-tpom/-npom* is a reflex of the Proto-Cariban nominalizer + attributivizer *\*-topo me* (Gildea 1998: 138–140). This can be illustrated with Wayana, where *-topo-me* still derives a purposive deverbal form (13), as opposed to Arara.

- (13) Wayana (Tavares 2005: 459)  
*Moloinë sisi hjak tīhe ilasilamtohme.*  
*moroinə hihi hja-kə t-iri-he i-rahirami-topo-me*  
 then sun in.sun-into CPL-put-CPL 3-dry-NMLZ-ATTRZ  
 'Then, they placed it into the sun, in order for it to dry.'

Similar to the case of *\*-ri*, there are also other languages which show innovative main clause forms with reflexes of *\*-topo me* (Gildea 1998: 180–181).

To sum up this section, based on a) *i-/Ø/t(i)-* occurring in subordinate clauses; b) 3PL *-ŋmo* being originally nominal; and c) cognate forms functioning as subordinate clauses in other Cariban languages,

Table 4: Proto-Cariban Set I (main clause) person markers

(a) Transitive					(b) Intransitive		
A/P	1	2	1+2	3		S <sub>A</sub>	S <sub>P</sub>
1		*k-		*t(i)-	1	*w-	*uj-
2	*k-			*m(i)-	2	*m-	*yj-
1+2				*kit(i)-	1+2	*kit-	*k-
3	*uj-	*yj-	*k-	*n(i)-	3	*n-	*n(i)-

Table 5: Proto-Cariban Set II person markers

1	*u(j)-
2	*ə(j)-
1+2	*k-
3	*i- (*t-)

I conclude that forms with  $i-/Ø/t(i)-$  are former subordinate clauses turned main clauses, a common innovation in the Cariban family. However, these innovative Arara main clause forms and their cognates in other languages differ in one crucial aspect: They take the hierarchical person marking system shown in Table 1, while innovative main clause forms in other languages show strictly S/P-marking person prefixes, with the A being expressed as an external argument. This is visible in the Kuikuro example in (12), where the prefix  $e-$  simply marks second person, the third person A *kayaiha* ‘non-Indian’ being marked with the postposition *heke*. As we have seen in Section 2, Arara uses the same hierarchically conditioned markers in both innovative and original main clause verb forms, the only difference being in the form of third person markers. This difference between Arara and other Cariban languages, as well as the ultimate origin of the split between main clause  $n(i)-/n(i)-$  and (former) subordinate  $i-/Ø/t(i)-$  is the topic of Section 4.

#### 4 Extension of main clause person marking to subordinate clauses

Proto-Cariban is reconstructed as having the main clause person marking system shown in Table 4, based on Meira et al. (2010: 495) and Gildea & Zúñiga (2016: 497); this paradigm and reflexes thereof are referred to as Set I. It is a hierarchically based system, with transitive verbs showing either the A-oriented or the P-oriented SAP markers in mixed scenarios, third person markers in nonlocal scenarios, and special markers in local scenarios. As discussed in Section 2, Arara has modified or replaced some of the prefixes in this paradigm, but preserves the functionality, with a person hierarchy and A- and P-oriented SAP markers.

Proto-Cariban is also reconstructed as having another series of person markers, referred to as Set II. The reconstructed forms for this set are shown in Table 5, based on Meira et al. (2010: 489). This paradigm was used on nouns and postpositions, and, most importantly, on deverbalized forms. That is, nominalized and adverbialized verbs took Set II markers, not Set I markers. In most cases, the formal possessor of these deverbal forms corresponded to the S/P arguments; a notable exception to this will briefly be discussed in Section 4.3. The Set II paradigm on such deverbalizations differed from the version used on nouns in having an additional third person marker  $*t-$ , occurring on a limited set of transitive verbs – other transitive verbs used  $*i-$ . Various reflexes of the Set II paradigm on possessed nominalizations are illustrated in (14), with reflexes of the two distinct third person markers  $*i-$  and  $*t-$  shown in (14d–e).

- (14) a. Apalaí (E. Koehn & S. Koehn 1986: 93)  
*j-etapa-ri*  
 1-kill-PERT.NMLZ  
 ‘My being killed.’
- b. Hixkaryána (Derbyshire 1985: 231)  
*aj-animi-pito-ri*  
 2-lift.up-NEG.NMLZ-PERT  
 ‘your not being lifted up’
- c. Wayana (Tavares 2005: 154)  
*ku-w-ehi-topo-npiri-komo*  
 1+2-S<sub>A</sub>-be-NMLZ-DEV-PL  
 ‘our ancient people’
- d. Kari’ña (Courtz 2008: 115)  
*i-kupi-nen*  
 3-bathe-AGT.NMLZ  
 ‘his/her bather’
- e. Waiwai (Hawkins 1998: 93)  
*miimo ʃ-i-tofo j-eʃe-tho*  
 house 3-make-NMLZ LK-payment-PST  
 ‘payment for making a house’

Since they originate in deverbalizations, innovative main clause forms usually do not take Set I, but Set II markers, as in Pemón:

- (15) Pemón (Swiggers 2010: 143)  
*i-wə=u-ja*  
 3P-kill=2-ERG  
 ‘You kill him.’

In this 2>3 scenario in an innovative construction, it is simply the third person which is marked, with the Set II prefix *i-*, rather than using a reflex of the Set I 2>3 prefix *\*m(i)-*. This is in contrast to the innovative *i-/Ø/t(i)-* forms in Arara, which show a hierarchically organized paradigm, as usually found in Set I. The extension leading to this difference between Arara and e.g. Pemón is discussed in Section 4.1. In Section 4.2, I show that this extension already happened in the common ancestor of Arara and Ikpeng (Proto-Xinguan), and that Ikpeng underwent further changes. In Section 4.3, I discuss a similar, but formally different innovation in Bakairi, the third member of the Pekodian branch. Finally, Section 4.4 briefly compares subordinate clauses in Xinguan and in Bakairi.

#### 4.1 The extension in Arara

The difference in argument marking in Arara subordinate clauses and subordinate clauses in other Cariban languages is illustrated in (16), the normal Cariban pattern being represented by Tiriyo.

- (16) a. Tiriyo (Meira 1999: 184)  
*ə-w-əturu-to w-ekanipi*  
 2-S<sub>A</sub>-talk-CIRC.NMLZ 1>3-think.PRS.PFV  
 ‘I thought it was in your talk.’
- b. Arara (Alves 2017: 99)  
*m-omomi-tobot in-moŋogu-li*  
 2S<sub>A</sub>-enter-NMLZ 1>3-wait-REC  
 ‘I waited for your entry.’

Table 6: The composition of the Set I/Set II paradigm in Arara

prefix group	example	form from
inverse	3>2 <i>o-</i>	Set I + Set II
direct & local	2>3 <i>m(i)-</i>	Set I
nonlocal	3>3 <i>i-/Ø/t(i)-</i>	Set II

In (16a), the Tiriyo S<sub>A</sub> verb *aturu* ‘to talk’ is nominalized with the circumstantial nominalizer *-to(po)*. It takes the second person Set II prefix *ə-*, which is followed by the obligatory class marker *w-* occurring on deverbal S<sub>A</sub> verbs. In (16b), the Arara S<sub>A</sub> verb *omomi* ‘to enter’ is nominalized with the circumstantial nominalizer *-tobot/-tpot*. It also has a second person argument, which is however not marked with the Set II prefix *o-/w-*, but rather with the same Set I prefix *m-* found in main clause forms. This innovation in subordinate clause person marking has been described as a “wholesale extension of main clause argument structure to subordinate clauses” in the Pekodian branch (Gildea 2012: 465).

Returning to Arara third person marking, it becomes clear that *n(i)-/n(i)-* is a reflex of the Proto-Cariban Set I third person marker *\*n-* (Table 4), while *i-/Ø/t(i)-* originates in the two lexically distributed third person markers *\*i-* and *\*t-* in the Set II paradigm (Table 5). However, the third person markers *i-/Ø/t(i)-* are not in a paradigmatic relationship with other Set II markers, but rather with the Set I markers. Their function in transitive scenarios must accordingly be ‘3>3’, and not ‘3P’. Further, this means that Arara did not actually undergo a complete extension of Set I markers to subordinate clauses. Rather, the extension did not affect the third person; that is, the Set I third person marker *\*n-* did not replace Set II *\*i-* and *\*t-* when Set I was extended. Thus, Arara main clause forms with *n(i)-/n(i)-* are inherited Set I main clause forms, whereas those forms showing *i-/Ø/t(i)-* are main clause forms innovated from subordinate clauses, showing a mixed Set I/Set II paradigm.

What caused this partial extension of Set I prefixes into forms formerly taking Set II prefixes? When considering the forms in Tables 4 and 5, there is an obvious formal overlap between the inverse markers (3>SAP) in the Set I system and the non-third person markers in the Set II system. I suggest that the extension was motivated by the reanalysis of these SAP P-marking prefixes (the first three rows in Table 5) in subordinate clauses as being 3>SAP markers (the first three cells in the bottom row of Table 4a). This then led to an extension of local and direct Set I markers to those parts of the subordinate paradigm which were not identical to the main clause one. However, while the SAP P markers took on the 3>SAP functions, and the SAP>SAP and SAP>3 markers were newly introduced entirely, *\*i-* and *\*t-* simply took on the functions of *\*n-* in Set I, instead of being replaced by it. That is, they underwent a change of function from 3P to 3>3. The composition of this mixed system is shown in Table 6, where the markers in inverse scenarios are ambiguously from Set I or Set II, the direct and nonlocal markers are from Set I, and the nonlocal markers are from Set II. The resulting system is organized like a Set I paradigm, being compatible with either a hierarchical or portmanteaux analysis. It only differs from the original Set I paradigm by the presence of third person *i-/Ø/t(i)-*, and the absence of *n(i)-/n(i)-*.

## 4.2 A further extension in Ikpeng

As mentioned, Ikpeng is rather closely related to Arara. However, none of the available descriptions of Ikpeng morphosyntax (Pachêco 1997; Campetela 1997; Pachêco 2001; Alves Chagas 2013) contains any mention of a third person marker *n-* (or prefixes cognate with Arara *kun-* and *mon-*). That is, the known Ikpeng verbal person marking is essentially identical to the system shown for Arara in Table 1, except for the absence of reflexes of third person *\*n-*. Thus, in contrast to Arara, **only** a mixed Set I/Set II system is attested in the literature on Ikpeng, with no split between reflexes of *\*n-* and *\*i-/\*t-*.

However, given that the facts for the rest of the system are very similar, it is reasonable to assume that the partial extension of Set I to subordinate clauses was already present in the ancestor of Arara and Ikpeng, Proto-Xinguan. This leaves us with three possible reasons for the absence of *n-* in Ikpeng: a) Ikpeng lost all original Set I forms, leaving only innovative forms with the mixed paradigm; b) there

are forms with *n-* in Ikpeng, they just have not been found yet; and c) Ikpeng replaced *n-* in the original Set I forms by analogy with the Set I/Set II third person markers. As I will show, it is in fact a combination of the scenarios b) and c).

Since there is no consensus on the analysis of Ikpeng TAM morphology, I cannot give a complete account; for a contrastive discussion of the analyses by Campetela (1997) and Pachêco (2001) and her own analysis, see Alves Chagas (2013: 135–144). However, certain suffixes are clearly recognizable as cognate with Arara suffixes. In particular, there is a TAM ending *-(t)añ-te*, which is analyzed as a remote past by Campetela (1997: 77) and Pachêco (2001: 80), as a durative remote past by Pachêco (1997: 60), and as a non-immediate past by Alves Chagas (2013: 140–141). Its Arara cognate is the intermediate past *-te*, which usually co-occurs with a suffix *-tañ* whose semantic contribution is unclear (Alves 2017: 112). Examples of these cognate forms are shown in (17).

- (17) a. Arara (Alves 2017: 112)  
*mobu j-anumi-tañ-te i-mobu-n lon*  
 canoe 1>3-raise-ASP-INT 1-canoe-PERT FOC  
 ‘I raised my own canoe.’
- b. Ikpeng (Alves Chagas 2013: 141)  
*munpok Ø-ak-tañte wot*  
 yesterday 3>3-eat-REM fish  
 ‘He ate fish yesterday.’

In contrast to the Arara cognate form, the Ikpeng remote past does not have a third person prefix *mon-*, but shows the usual third person marker, occurring in its zero allomorph in (17b). Also, there is no complementary distribution of third person markers and the *-te* suffix, unlike in Arara (Section 3). Another indication that these forms are cognate is plural marking: both languages show the plural marker *-tom*, as shown in (18).

- (18) a. Arara (Alves 2017: 176)  
*mon-ipi-tañ-tom*  
 3INT-bathe-ASP-PL  
 ‘They were bathing.’
- b. Ikpeng (Campetela 1997: 78)  
*Ø-ajñku-tañ-tom ikpeñ keni.niñkin jatupa gwam*  
 3>3-fish-REM-PL I. all? J. OBL  
 ‘All the Ikpeng fished in the Jatobá (river).’

This strongly suggests that the forms which in Arara preserve third person *n(i)-/n(i)-* have replaced it with *i-/Ø/t(i)-* in Ikpeng, eliminating the third person split present in Proto-Xinguan.

This scenario is additionally supported by two irregular verbs in Ikpeng, which preserve third person *\*n-*. Both relevant third person forms have the same past tense used in (17b) and (18b) above. One of the verbs is the copula, which has the third person form *naki* in that tense, as shown in Table 7. While this prefix *n-* is not identified as such in any of the contributions on Ikpeng, I suggest that it is a reflex of Set I third person *\*n-*, since *naki* is clearly morphologically complex: *a* is a reflex of the Proto-Cariban copula *\*a[p]*, and *-ki* is a reflex of an irregular ending *\*-kə*, which is reconstructible to Proto-Cariban (Gildea 2018: 379)

The other verb is *ke* ‘to say’, which has an irregular third person form *añkanj*. I analyze this form as containing a third person prefix *añ-* (a reflex of third person *\*n-*), the root *ke* ‘to say’, and the tense marker *-(t)añ*, as shown in (19a). This verb shows the usual zero 3S<sub>A</sub> marker in other TAM configurations (19b).

Table 7: Forms of the Ikpeng copula

	REM	REC	NPST
1	<i>Ø-i-taŋte</i>	<i>Ø-eŋfi-li</i>	<i>Ø-eŋfi</i>
2	<i>m-i-taŋte</i>	<i>m-eŋfi-li</i>	<i>m-eŋfi</i>
1+2	<i>kur-am-taŋte</i>	<i>kur-am-li</i>	<i>kur-am-ŋfi</i> or <i>kuŋŋan</i>
3	<i>n-a-ki</i>	<i>imro<sup>a</sup></i>	<i>Ø-eŋfi</i>

<sup>a</sup> This form seems to come from the third person inanimate medial demonstrative \**mərə* (Meira 2002: 268); in Arara, *imro* is still used as a pronoun, see e.g. Alves (2017: 86).

(19) Ikpeng (Pachêco 2001: 279, 136)

- a. *ankang man timuye na*  
*an-k-aŋ man ti-muje na*  
 3-say-REM.CONT AFF COR-wife OBL  
 ‘He said to his wife.’
- b. *eram kun Ø-ke-li*  
 truth EMP 3S<sub>A</sub>-say-REC  
 ‘S/he really said the truth.’

While the expected /\_C allomorph of the prefix would be *ni-*, the absence of an epenthetic *i* with this particular verb is attested in other languages as well, for example in Tiriyó (20a) or in Hixkaryána (20b). I suggest that Ikpeng inherited the idiosyncratic \**nk* cluster in the third person form of ‘to say’, but later added a prothetic vowel.

- (20) a. Tiriyó (Carlin 2004: 537)  
*serə n-ka-n*  
 PROX.DEM.INAN 3-say-UNCERT  
 ‘This is what that one told, he says (it is said).’
- b. Hixkaryána (Derbyshire 1985: 60)  
*onoki oni wjaro n-ka-no*  
 who PROX.DEM.INAN like 3-say-REC  
 ‘Who said this?’

Summing up, the innovations resulting in the third person split in Arara already happened at the Proto-Xinguan stage, before the separation of Arara and Ikpeng. However, Ikpeng has undergone a further innovation, replacing the old third person marker \**n-* with innovative *i-/Ø/t(i)-*. Reflexes of \**n-* were preserved in some forms of the irregular verbs ‘to be’ and ‘to say’.

### 4.3 A similar extension in Bakairi

There is not much reliable descriptive material available on Bakairi, but Meira (2003) gives a nice basic overview of verbal morphology. Bakairi also has a person marking split between original and innovative main clause forms, with the former being found in past tenses, and the latter elsewhere (Meira 2003: 2–5). The person marking paradigm of the original main clause verb forms is shown in Table 8. As in Arara, these forms preserve third person \**n-* from the Proto-Cariban Set I paradigms:

- (21) Bakairi (Meira 2003: 4)  
*n-epagu-də*  
 3-stop-IMM  
 ‘S/he stopped.’

Table 8: Bakairi person marking in Set I main clause forms (Meira 2003: 2–4)

(a) Transitive					(b) Intransitive		
A/P	1	2	1+2	3	S <sub>A</sub>	S <sub>P</sub>	
1		ə-		s-	1	k-/w-	i/j-
2	i/j-			m-	2	m-	ə-
1+2				kid-	1+2	kid-/k-	k-
3	i/j-	ə-	k-	n-	3	n-	

Table 9: Bakairi person marking in innovative main clause forms (Meira 2003: 5–6)

(a) Transitive					(b) Intransitive		
A/P	1	2	1+2	3	S <sub>A</sub>	S <sub>P</sub>	
1		?		k(ə)~-	1	k-/w-	i/j-
2	?			m(ə)~-	2	m-	ə-
1+2				ki~-	1+2	kid-/k-	k-
3	i/j-	ə-	k-	j~-	3	i-/Ø	

The paradigm found in innovative main clause forms is shown in Table 9. Their origins are demonstrated by Meira (2003) with the fact that they can still serve as nominalizations, as shown in (22).

(22) Bakairi (Meira 2003: 8)

*mə-ēgatu-li wəgə mə-ūtu-li*  
 2>3-tell-NMLZ about 2>3-know-GNO  
 ‘You know what you’re telling.’

The nominalized form *məēgatuli* ‘what you’re telling’ is followed by the postposition *wəgə* (Proto-Cariban \**pəkə*), showing its nominal nature, while the form *məūtuli* ‘you know’ serves as the head of the sentence. As in innovative main clause forms in Arara and Ikpeng, inverse scenarios are marked with ambiguous markers which could be either from Set I or Set II, but the other markers are clearly not cognate with the Xinguan ones. Direct scenarios show otherwise unattested prefixes, which can by and large be described as S<sub>A</sub> prefixes (Tables 8b and 9b) combined with nasalization of the following vowel. Besides the nasalization, these prefixes differ from the S<sub>A</sub> markers by a) having *ə*-final allomorphs in the first and second person, written as *Cā-* by Souza (1999), and b) having a *i*-final form in 1+2 scenarios. Nonlocal scenarios show *j*~, where *j*- is likely a reflex of the Set II third person marker \**i*-, another reflex of which is found on intransitive verbs (Table 8b). It is not clear whether local scenarios can be expressed in these verb forms, as there are no examples.

As Meira (2003: 6–10) shows, the nasalization in scenarios with a third person P is a reflex of the Proto-Cariban object nominalizer<sup>9</sup> \**n*-, a distinct morpheme from the third person marker \**n*-. Another reflex of the object nominalizer \**n*- is illustrated in (23) with Ye’kwana.

(23) Ye’kwana (Cáceres 2011: 75)

*ə-n-akətə-ri*  
 2-NMLZ-cut-PERT.NMLZ  
 ‘what you cut’

<sup>9</sup>In his comparative discussion of that prefix, Gildea (1998: 128–129) acknowledges that an interpretation of \**n*- as a nominalizer faces certain difficulties, mainly the fact that nominalizing suffixes also occur. However, he continues to use the term “object nominalizer”, due to the lack of alternative labels, a practice which I follow here; the unclear status of \**n*- is of no relevance to the discussion at hand.

The combination of a second person possessive marker and the object nominalizer *n-* is prefixed to a nominalized form of the verb *akətə* ‘to cut’. The resulting nominalization refers to the object of a past action by the second person; the Set II prefix *ə-* refers to the A, as opposed to the usual S/P pivot found in deverbal forms.

As discussed, the Bakairi forms containing reflexes of this object nominalizer *\*n-* do not use Set II prefixes. Rather, it seems that the Proto-Pekodian Set I  $S_{(A)}$  markers *\*k-* ‘1’, *\*m-* ‘2’, *\*k-* ‘1+2 (?)’, and the Set II marker *\*i-* ‘3’ were combined with *\*n-* nominalizations of transitive verbs, potentially with some additional material between the prefixes and *\*n-*, given the unexpected vowels with the *SAP* prefixes. The fact that intransitive prefixes were recruited for (transitive) >3 scenarios might seem counterintuitive, but it has been suggested that reflexes of the object nominalizer *\*n-* actually derive nominalized intransitive verbs in other Cariban languages (Gildea 1998: 129).

#### 4.4 Two Pekodian innovations?

In Section 4.1, I have demonstrated the partial extension of Set I prefixes to subordinate clauses in Arara, leading to the split between *n(i)-/n(i)-* and *i-/Ø/t(i)-* discussed in Section 3. In Section 4.2, I have shown that Ikpeng underwent a further innovation whereby the original Set I third person marker *\*n-* was replaced with reflexes of the Set II markers *\*i-* and *\*t-*, except in two irregular verbs. In Section 4.3, I have illustrated a similar extension of Set I markers to subordinate clauses in Bakairi, where they were combined with the object nominalizer *\*n-*. While this latter extension has certain similarities with the one found in the Xinguan languages, it is clearly different in terms of the material which was used. Thus, a first conclusion must be that there were at least two distinct extensions of Set I prefixes into subordinate clauses in Pekodian – some of which were then reanalyzed as main clauses.

As for extant subordinate clauses, they also appear to be different in the Xinguan languages and Bakairi, although the available data is not as extensive as for main clauses. Forms with a reflex of the object nominalizer *\*n-* also occur in Ikpeng, used for relative clauses, as shown in (24).

- (24) Ikpeng (Pachêco 2001: 54)  
*petkom o-n-enen-pin Ø-ero-li*  
 woman 2-NMLZ-see-PST 3-go-REC  
 ‘The woman you saw left.’

Unlike other nominalizations in Ikpeng (and unlike the Bakairi forms with reflexes of the object nominalizer *\*n-*), these forms show regular Set II prefixes; otherwise the A-oriented second person marker *m-* would be found in the form in (24). While I have no examples of the Arara reflex of *\*n-* ‘NMLZ’, it does exist (Alves 2017: 75), presumably working similarly to its Ikpeng cognate. In addition to constituting a marked difference between Xinguan and Bakairi, this also suggests that the partial extension of Set I in Xinguan was in fact even less wholesale, in that it did not affect forms with the object nominalizer *\*n-*. Further, the Bakairi nominalizations discussed by Meira (2003: 8–10) show reflexes of both the object nominalizer *\*n-* and the nominalizer *\*-topo*, as shown in (25a). This combination is not found elsewhere in the family (Meira 2003: 9), including Ikpeng, which shows the usual mixed Set I/Set II prefixes on nominalizations with the reflex of *\*-topo* (25b).

- (25) a. Bakairi (Meira 2003: 9)  
*k-ãtə-ho*  
 1>3-cut-NMLZ  
 ‘(something) for my cutting’
- b. Ikpeng (Pachêco 2001: 121)  
*ugun petkom Ø-anon-li emangatkuri biscoito t-eru-tpot anpi ina*  
 DIST.DEM.ANIM woman 3>3-ask-REC girl biscuit 3>3-give-NMLZ boy OBL  
 ‘That woman asked the girl to give the biscuit to the boy.’

Summing up, argument marking in Xinguan and in Bakairi subordinate clauses (and the main clauses that evolved from them) employs clearly different material, suggesting that two different innovations happened. However, only little material is available on Bakairi subordinate clauses, as well as on Arara forms with reflexes of the object nominalizer *\*n-*. More data on these forms is needed for a conclusive answer about the diachronic details of the extension of Set I prefixes into Pekodian subordinate clauses.

## 5 The split in a family-wide perspective

From the *n(i)-/n(i)-* vs *i-/Ø/t(i)-* split in Arara discussed here, two main points for a family-wide perspective follow. The first point is the fact that the absence of the third person marker *\*n-* in some Pekodian main clause verb forms, which can be explained by their innovative nature, is of importance for the reconstruction of Proto-Cariban (Section 5.1). The second point pertains to the semantics of surviving Set I forms elsewhere in the family, which shows parallels to the distribution of *n(i)-/n(i)-* vs *i-/Ø/t(i)-* in Arara (Section 5.2).

### 5.1 Importance for reconstructing third person *\*n-*

The absence of reflexes of the third person prefix *\*n-* in some forms in the Pekodian branch (and its apparent total absence in Ikpeng) has led to doubts about its reconstructability to the Proto-Cariban Set I paradigm (Meira et al. 2010: 497):

The Pekodian languages, which Gildea (1998) largely did not consider, all present multiple conjugations with third-person *Ø-/i-* instead of *n(i)-*, which suggests that at least *\*i-* ‘3[Sp]’ must be reconstructed to Proto-Carib. The question that must be addressed in future research is whether the additional *\*n-* ‘3’ component was already present in Proto-Carib (and lost in some environments in some languages) or whether it represents an innovation after Proto-Carib, which was not equally productive in all branches of the family.

The diachronic developments presented here clearly show that the first scenario, third person *\*n-* being present in Proto-Cariban and subsequently being lost, is the accurate answer. In Arara (and thus Proto-Xinguan), the extension of Set I markers to subordinate clauses did not take place in the third person (Section 4.1), resulting in the split between forms with *n(i)-/n(i)-* and those with *i-/Ø/t(i)-* in Arara (Section 3). Ikpeng further extended *i-/Ø/t(i)-*, replacing almost all reflexes of third person *\*n-* (Section 4.2). Finally, the innovative paradigm attested for Bakairi is not cognate, but the innovation resulted in a similar split, with original main clause forms conserving third person *n-*, but innovative forms showing *j~*. However, a remaining problem for the reconstruction of *\*n-* to Proto-Cariban are *\*n-*-less inflections scattered throughout the family: Tiriyó, Kari’ña, Apalaí, and Werikyana have individual verb forms that otherwise take Set I prefixes, but show no reflex of *\*n-* in their third person forms (Meira et al. 2010: 497).

### 5.2 Functional distribution of *n(i)-/n(i)-* vs *i-/Ø/t(i)-*

It is suggestive that S. D. C. de Souza (1993) coined labels for Arara forms with *n(i)-/n(i)-* and those with *i-/Ø/t(i)-*. She used the term “virtual” for the forms with *n(i)-/n(i)-*: permissives and interrogatives, and – in her analysis – negations; these can be subsumed under irrealis moods. She used “real” for the form with *i-/Ø/t(i)-*, which are all used in declarative statements, and can be considered to show realis mood. While S. D. C. de Souza (1993) missed the clearly realis intermediate and remote past forms, with *mon-* and *kun-*, functional parallels between the Arara “virtual” forms, and the distribution of remaining Set I forms in other Cariban languages can be drawn.

Gildea (1998: 161) suggests the following reoccurring pattern for the gradual replacement of Set I main clause forms with innovative Set II forms:

1. loss of Set I forms with future semantics

2. loss of Set I forms with nonpast semantics
3. loss of aspectual distinctions in Set I past forms
4. loss of remoteness distinction in Set I past forms
5. complete loss of Set I forms

The list of the Arara TAM configurations where mixed Set I/Set II forms are used (Table 2) largely agrees with this pattern: the future is covered by a *i-/Ø/t(i)-* form, which is predicted by stage 1. While there is no Set I nonpast form, that function is now by the imperfective *-naŋri*, which can also bear other temporal meanings (Section 3). This is similar to innovative Set II forms in other languages, summarized by Gildea (1998: 163) as expressing “nonpast”, “universal”, or “nonspecific” tenses. Finally, the innovative “recent” past can be seen as the first step of stage 4. As for the full Set I forms, they include the intermediate and remote past tenses, which is largely compatible with Gildea’s (1998) scheme, except for the continuing presence of aspectual distinctions in combination with degrees of remoteness (stages 3 and 4)

This leaves us with two more remaining Set I forms, the permissive and the interrogative, which do not fit into Gildea’s (1998) tense-aspect base scheme. Rather, they share the property of having irrealis mood values. Interestingly, very similar patterns are found in other Cariban languages, for example in those of the Pemongan dialect continuum, spoken around the tri-border area between Brazil, Venezuela, and Guyana. Since Migliazza (1985), the Pemongan languages, with their respective dialects, are grouped as follows: Macushi, Pemón (Taulipang, Arekuna, Kamarakoto), and Kapón (Akawaio, Ingarikó, Patamona). All languages in this group have innovative main clause verb forms with Set II person prefixes, as illustrated for Macushi in (26).

(26) Macushi

- a. *i-kana?ma=u-ja*  
3-scratch=1-ERG  
‘I scratch it.’ (Carson 1982: 131)
- b. *i-koneka-?pi=i-ja*  
3-make-PST=3-ERG  
‘He made it.’ (Abbott 1991: 24)

There are also some Set I forms used in main clauses, but innovative Set II forms are clearly predominant in these languages (Gildea 1998: 161). These remaining Set I forms show semantic or pragmatic restrictions, differing in strength depending on the language.

Macushi shows the most restricted distribution of Set I forms: Second person forms with *m-* remain in a kind of polite/future imperative (27a), while old 1+2>3 forms (from Proto-Cariban *\*kit(i)-*) take up a prohibitive function (27b). Third person forms show permissive semantics (27c); there are no first person forms.

(27) Macushi (Abbott 1991: 50, 53, 53)

- a. *miikiri*            *j-arakkiri m-es-urima-i*  
MED.DEM.ANIM LK-with    2-DETRZ-tell-IMP  
‘Talk with him!’
- b. *imakui?pi ki?-ku?-ti*  
bad            NEG-do-IMP.PL  
‘Don’t do bad.’
- c. *tiwi*    *n-ati-i*  
PERM 3-go-IMP  
‘Let him go.’

Taulipang (Pemón) shows similar patterns in the forms that are cognate to the Macushi ones in (27), although they are not as restrictive. Second person forms can either be used as a question, or as a polite request (28a). Third person forms can also function as a question, or as a permissive (28b). Former 1+2 forms function as prohibitives (28c), while first person forms refer to the immediate past (28d).

## (28) Pemón

- a. *kareta m-aʔnɪpi-i*  
paper 2-burn-IND  
'Did you burn the book?' or 'Could you burn the book?' (Álvarez 1997: 82)
- b. *paruru n-aimuku-i*  
banana 3-pick.up-IND  
'Did he pick up the banana?' or 'Let him pick up the banana.' (Álvarez 1997: 82)
- c. *piasan pəʔ taure-ʔpə rato=ja komi? ke tuna tiise kiʔi-ni-i*  
shaman with say-PST rat=ERG cold with water also PROH-drink-IND  
'The rat said to the shaman: "Don't drink cold water."' (Álvarez 2000: 100)
- d. *uj-un s-uʔnatə-i*  
1-father 1>3-bury-IMM  
'I buried my father.' (Álvarez 1997: 81)

There is also another irrealis Set I form in Pemón, namely the intentional (29a), the only realis Set I form being a distant past tense (29b).

## (29) Pemón (Álvarez 1997: 84)

- a. *maʔnon a-ʔma se-me-ina*  
girl 2POSS-food 1>3-taste-INTEN  
'Girl, I'm going to taste your food.'
- b. *tuʔke konoʔ s-entaka-tai karaiwa jami? pijau*  
many rain 1>3-spend-REM White.Brazilian PL among  
'I spent many years among the Brazilians.'

For the Kapón varieties Akawaio and Ingarikó, both Caesar-Fox (2003: 107) and Cruz (2005: 435) note that the semantics of the Set I TAM suffixes are not easily establishable; forms with Set I are rather peripheral, Set II forms being dominant in both varieties (Caesar-Fox 2003: 249; Cruz 2005: 434–435). However, both authors agree in that there is a present form with a certain/uncertain distinction ((-ja)-iʔ/(-ja)-tə-iʔ vs -ja-n/-ja-tə-u), a recent past (-i/-təu), a remote past (Akawaio -tai/-tantəu, Ingarikó -tai/-tatəi), and a vetative -nin (Caesar-Fox 2003: 107, 122; Cruz 2005: 266–269, 288–289). As in Macushi and Pemón, second and third person forms of the certain past can also be used for polite imperatives and permissives in Akawaio (Caesar-Fox 2003: 117).

Summing up, varieties in the Pemongan continuum show semantically and/or pragmatically restricted distributions of Set I forms. The degree of the restriction varies, ranging from only permissives and imperatives to interrogatives, to intentionals, and to indicatives (with past semantics). As briefly discussed by Gildea (1998: 78), many of the Pemongan forms discussed above originated in an immediate past \*-i, with subsequent semantic shifts into different irrealis meanings, depending on the person. While I have no detailed story for the irrealis Set I Arara forms, the semantic similarities to the surviving Set I forms in Pemongan are obvious. In addition, like Pemongan \*-i, Arara -(t)an shows different meanings, depending on the person. So far I have only shown its interrogative meaning; an identical form can function as an admonitive, with SAP.P and 2SA prefixes (Alves 2017: 164):

- (30) Arara (Alves 2017: 165)  
*w-aŋko-taŋ*  
 3>2-cut-ADM  
 ‘Watch out, lest it cut you!’

This either suggests a parallel development in the semantic shifts that remaining Set I forms can undergo, or that they are preserved in irrealis contexts for longer, adding another dimension to the progression pattern established by Gildea (1998). As for the third person split of *n(i)-/n(i)-* vs *i-/Ø/t(i)-* in Arara, I have shown it to match the distribution of innovative main clause forms in Cariban languages, further bolstering the hypothesis that *i-/Ø/t(i)-* in main clauses is an innovation.

## 6 Conclusion

I introduced the third person marking split between *i-/Ø/t(i)-* and *n(i)-/n(i)-* in Arara, conditioned by different TAM configurations. I then showed that the forms with *i-/Ø/t(i)-* originate from earlier nominalized subordinate clauses, while the ones with *n(i)-/n(i)-* are old main clause forms. The existence of a split in only the third person was explained as follows: *SAP.P* markers were reanalyzed as *3>SAP* markers, leading to the extension of direct and *S<sub>A</sub>* Set I markers into subordinate clauses. However, in *3S* and *3>3* scenarios, the existing Set II markers took on the function of the third person Set I markers. The subsequent recruitment of some subordinate forms into main clauses resulted in the main clause third person marking split found in Arara.

I also showed that Arara’s sister language Ikpeng underwent a further innovation, extending reflexes of the Set II third person markers *\*i-* and *\*t-* to forms which previously showed a reflex of Set I *\*n-*. The latter is only preserved in the irregular verbs ‘to be’ and ‘to say’. Further, Bakairi shows a similar split, which also developed from the extension of main clause person marking to subordinate clauses with subsequent recruitment into main clauses. However, the material underlying these innovative forms is not cognate with the Xinguan prefixes, suggesting that two independent innovations happened.

From a comparative perspective, the reconstructability of *\*n-* to Proto-Pekodian removes doubts about the presence of third person *\*n-* in the Proto-Cariban Set I paradigm, which were largely based on its partial absence in Pekodian (Meira et al. 2010). However, further descriptive work on subordinate forms in Pekodian is needed in order to work towards a detailed reconstruction of Proto-Pekodian verbal morphosyntax. Further, the semantic contexts in which Arara preserves forms with third person *n(i)-/n(i)-* show conspicuous similarities to the distribution of Set I forms in languages of the Pemongan group. Specifically, it is not only featured in past TAM configurations, but also appears in irrealis contexts. This adds another dimension to the tense-aspect based innovation scheme suggested by Gildea (1998: 161), and further supports the scenario whereby main clause verb forms with *i-/Ø/t(i)-* are innovative.

## 7 Abbreviations

The following abbreviations were used in this paper: 1 first person, 1+2 first and second person, 1+3 first and third person, 2 second person, 3 third person, A agentive transitive argument, ADM admonitive, AFF affirmative, AGT agent, ANIM animate, ASP aspect, ATTRZ attributivizer, CIRC circumstantial, CONT continuative, COP copula, COR coreference, CPL completive, DEM demonstrative, DETRZ detransitivizer, DEV devaluative, DIST distal, EMP emphatic, ERG ergative, EXIST existential, FOC focus, FUT future, GNO gnomic, IMM immediate past, IMP imperative, INAN inanimate, IND indicative, INT intermediate past, INTEN intentional, INTER interrogative, IPFV imperfective, ITER iterative, LK linker, MED medial, NEG negation, NMLZ nominalizer, NPST non-past, OBL oblique, P patientive transitive argument, PERM permissive, PERT perturbative, PFV perfective, PL plural, PNCT punctual, POSS possessive, PRO pronoun, PROH prohibitive, PROX proximal, PRS present, PST past, Q question particle/marker, REC recent past, REM remote past, RESUM resumptive, S intransitive argument, *S<sub>A</sub>* S marked like A, *S<sub>P</sub>* S marked like P, *SAP* speech act participant, UNCERT uncertainty.

## 8 Acknowledgments

I am very grateful to Spike Gildea, Jean Rohleder, Rik van Gijn, and Fernando Zúñiga, who read different versions of this paper and gave valuable feedback.

## References

- Abbott, Miriam (1991). “Macushi”. In: *Handbook of Amazonian Languages*. Desmond C. Derbyshire & Geoffrey K. Pullum (eds.). Vol. 3. Berlin/New York: Mouton de Gruyter: 23–160.
- Álvarez, José (1997). “Split Ergativity and Complementary Distribution of NP’s and Pronominal Affixes in Pemón (Cariban).” In: *Opción* 13.24: 69–94.
- (2000). “Construcciones progresivas en pemón y otras lenguas caribes”. In: *Opción* 32: 96–130.
- Alves, Ana Carolina Ferreira (2010). “Phonological Aspects of Arara (Carib, Brazil)”. MA thesis. Radboud Universiteit.
- (2017). “Morfofonologia, morfossintaxe e o sistema de tempo, aspecto e modo em Arara (Karib)”. Ph.D. dissertation. Universidade de São Paulo.
- Alves Chagas, Angela Fabíola (2013). “O Verbo Ikpeng: estudo morfossintático e semântico-lexical”. Ph.D. dissertation. Universidade Estadual de Campinas.
- Cáceres, Natalia (2011). “Grammaire fonctionnelle-typologique du ye’kwana, langue caribe du Venezuela”. Ph.D. dissertation. Université Lumière Lyon 2.
- Caesar-Fox, Desrey Clementine (2003). “Zauro’nödok Agawayo Yau: Variants of Akawaio spoken at Waramadong”. Ph.D. dissertation. Rice University.
- Competela, Cilene (1997). “Análise do sistema de marcação de caso nas orações independentes da língua Ikpeng”. MA thesis. Universidade Estadual de Campinas.
- Carlin, Eithne B. (2004). *A Grammar of Trio: A Cariban Language of Suriname*. Vol. 55. Duisburger Arbeiten zur Sprach- und Kulturwissenschaft. Frankfurt am Main: Peter Lang.
- Carson, Neusa Martins (1982). “Phonology and Morphosyntax of Macuxi (Carib) (Brazil)”. Ph.D. dissertation. The University of Kansas.
- Courtz, Henk (2008). *A Carib grammar and dictionary*. Toronto: Magoria Books.
- Cruz, Maria Odileiz Sousa (2005). “Fonologia e gramática Ingarikó: Kaʔpon-Brasil”. Ph.D. dissertation. Vrije Universiteit Amsterdam.
- Derbyshire, Desmond C. (1985). *Hixkaryana and Linguistic Typology*. Dallas: Summer Institute of Linguistics.
- de Souza, Isaac Costa (1988). “Contribuição para a fonologia da língua Arara (Karib)”. MA thesis. Universidade Estadual de Campinas.
- de Souza, Shirley Dias Cardoso (1993). “Alguns aspectos morfológicos da língua Arara (Karib)”. MA thesis. Universidade de Brasília.
- (2010). “The morphology of nouns in the Ugoronmo language (Arara of Pará)”. MA thesis. University of North Dakota.
- dos Santos, Gélsama Mara Ferreira (2007). “Morfologia Kuikuro: Gerando nomes e verbos”. Ph.D. dissertation. Universidade Federal do Rio de Janeiro.
- Durbin, Marshall (1977). “The Carib language family”. In: *Carib-Speaking Indians: Culture, Society, and Language*. Ellen B. Basso (ed.). Tucson, AZ: University of Arizona Press: 23–38.
- Ferreira-Alves, Ana Carolina, Angela Fabíola Alves Chagas & Leonard Jéferson Grala Barbosa (2019). “Termos de parentesco: primeiras reconstruções em Proto-Arara-Ikpeng”. In: *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 14: 101–119. DOI: 10.1590/1981.81222019000100007.
- Franchetto, Bruna (2010). “The ergativity effect in Kuikuro (Southern Carib, Brazil)”. In: *Ergativity in Amazonia*. Spike Gildea & Francesc Queixalós (eds.). Amsterdam: John Benjamins: 121–158.
- Gildea, Spike (1998). *On reconstructing grammar: Comparative Cariban morphosyntax*. Oxford: Oxford University Press.
- (2005). “Carib and Cariban Languages”. In: *Encyclopedia of linguistics*. Philipp Strazny (ed.). Vol. 1. New York: Fitzroy Dearborn: 173–176.

- (2012). “Linguistic studies in the Cariban family”. In: *The Indigenous Languages of South America: A Comprehensive Guide*. Lyle Campbell & Verónica Grondona (eds.). Vol. 2. The World of Linguistics. Berlin/New York: Mouton de Gruyter: 441–494.
- (2018). “Reconstructing the copula and nonverbal predicate constructions in Cariban”. In: *Nonverbal predication in Amazonian languages*. Simon E. Overall, Rosa Vallejos & Spike Gildea (eds.). Amsterdam: John Benjamins: 365–402.
- Gildea, Spike & Fernando Zúñiga (2016). “Referential hierarchies: A new look at some historical and typological patterns”. In: *Linguistics* 54.3: 483–529.
- Girard, Victor James (1971). “Proto-Carib phonology”. Ph.D. dissertation. University of California.
- Hawkins, Robert E. (1998). “Wai Wai”. In: *Handbook of Amazonian Languages*. Desmond C. Derbyshire & Geoffrey K. Pullum (eds.). Vol. 4. Berlin/New York: Mouton de Gruyter: 25–224.
- Hoff, Berend J. (1995). “Configurationality and nonconfigurationality in the Carib language of Surinam”. In: *International Journal of American Linguistics* 61.4: 347–377.
- Hoff, Berend J., Spike Gildea & Racquel-María Sapién (2017). “Between Evidentiality, Immediacy, and Epistemic Certainty in Kari’ña”. Talk held at SLE 2017. Zürich.
- Kaufman, Terrence (1994). “The native languages of South America”. In: *Atlas of the world’s languages*. Christopher Moseley & R.E. Asher (eds.). London: Routledge: 46–76.
- (2007). “The native languages of South America”. In: *Atlas of the world’s languages*. Christopher Moseley & R.E. Asher (eds.). London: Routledge: 59–94.
- Koehn, Edward & Sally Koehn (1986). “Apalai”. In: *Handbook of Amazonian Languages*. Desmond C. Derbyshire & Geoffrey K. Pullum (eds.). Vol. 1. Berlin/New York: Mouton de Gruyter: 33–127.
- Mattéi-Müller, Marie-Claude (2002). “En busca de criterios clasificatorios para la Familia Caribe”. In: *Línguas indígenas Brasileiras: fonologia, gramática e história*. Ana Suely Arruda Câmara Cabral & Aryon D. Rodrigues (eds.). 2 vols. Atas do I Encontro Internacional do Grupo de Trabalho sobre Línguas Indígenas da ANPOLL. Belém: EDUFPA: 37–53.
- Meira, Sérgio (1999). “A Grammar of Tiriyo’”. Ph.D. dissertation. Rice University.
- (2000). “The accidental intransitive split in the Cariban family”. In: *Reconstructing Grammar: Comparative Linguistics and Grammaticalization*. Spike Gildea (ed.). Vol. 43. Typological Studies in Language. Amsterdam: John Benjamins: 201–230.
- (2002). “A first comparison of pronominal and demonstrative systems in the Cariban language family”. In: *Current Studies on South American Languages: Selected papers from the 50th International Congress of Americanists in Warsaw and the Spinoza Workshop on Amerindian Languages in Leiden*. Mily Crevels et al. (eds.). Leiden: Research institution of Asian, African, and Amerindian Studies (CNWS): 255–275.
- (2003). “A marcação de pessoa nos verbos em Bakairi (Karíb)”. In: *Ergatividade na Amazônia*. Vol. 2. Francisc Queixalós (ed.). Brasília, Paris: Laboratório de Línguas Indígenas, Centre d’études des langues indigènes d’Amérique: 155–163.
- Meira, Sérgio & Bruna Franchetto (2005). “The Southern Cariban languages and the Cariban family”. In: *International Journal of American Linguistics* 71.2: 127–192.
- Meira, Sérgio, Spike Gildea & Berend J. Hoff (2010). “On the Origin of Ablaut in the Cariban Family”. In: *International Journal of American Linguistics* 76.4: 477–515.
- Migliazza, Ernesto C. (1985). “Languages of the Orinoco-Amazon Region: Current Status”. In: *South American Indian Languages: Retrospect and Prospect*. Harriet E. Manelis Klein & Louisa Stark (eds.). Austin: Texas University Press: 17–139.
- Pachêco, Frantomé B. (1997). “Aspectos da gramática Ikpeng (Karíb)”. MA thesis. Universidade Estadual de Campinas.
- (2001). “Morfossintaxe do verbo Ikpeng (Karíb)”. Ph.D. dissertation. Universidade Estadual de Campinas.
- Payne, Thomas E. & Doris L. Payne (2013). *A Typological Grammar of Panare: A Cariban Language of Venezuela*. Brill’s Studies in the Indigenous Languages of the Americas. Leiden: Brill.
- Simons, Gary F. & Charles D. Fennig, eds. (2019). “Ethnologue: Languages of the World”. Online: <http://www.ethnologue.com>.

- 
- Souza, Tânia Conceição Clemente de (1999). *Discurso e oralidade: um estudo em língua indígena*. Niterói: Publicações do Mestrado em Comunicação, Imagem e Informação, UFF.
- Swiggers, Gisela (2010). “Gramática de Referencia del Pemón”. Ph.D. dissertation. Universidad del Zulia.
- Tavares, Petronila da Silva (2005). “A grammar of Wayana”. Ph.D. dissertation. University of Houston.
- Yamada, Racquel-María (2011). “A New Approach to *ky-* and *-ng* in Kari’ña: Evidentiality or Something Else?” In: *International Journal of American Linguistics* 77.1: 59–89.